Hearing Research 395 (2020) 108020

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Static and dynamic cocktail party listening in younger and older adults

Hartmut Meister [a, *], Fabian Wenzel [a], Axel Kai Gehlen [a], Josef Kessler [b], Martin Walger [a, c]

[a] *Jean-Uhrmacher-Institute for Clinical ENT-Research, University of Cologne, Geibelstr. 29-31, D-50931, Cologne, Germany*
[b] *Department of Neurology, University Hospital Cologne, Kerpener Str. 62, D-50937, Cologne, Germany*
[c] *Clinic of Otorhinolaryngology, Head and Neck Surgery, University of Cologne, Kerpener Str. 62, D-50937, Cologne, Germany*

A R T I C L E   I N F O

Article history:
Received 5 July 2019
Received in revised form
13 May 2020
Accepted 11 June 2020
Available online 13 July 2020

Keywords:
Speech recognition
Multi-talker
Age
Cocktail party
Working memory
Cognitive flexibility

A B S T R A C T

Verbal communication often takes place in situations with several simultaneous speakers ("cocktail party listening"). These situations can be static (only one listening target) or dynamic (with alternating targets). In particular, dynamic cocktail party listening is believed to generate extra cognitive load and appears to be particularly demanding for older listeners.

Two groups of younger and older listeners with good hearing and normal cognition participated in the present study. Three different, spatially separated talker voices uttering matrix sentences were presented to each listener with varying types and probabilities of target switches. Moreover, several neuropsychological tests were conducted to investigate general cognitive characteristics that may be important for speech understanding in these situations.

In a static condition with a priori knowledge of the target talker, both age groups revealed very high speech recognition performance. In comparison, dynamic conditions caused extra costs associated with the need to monitor different potential sound sources and to refocus attention when the target changed. The amount of costs depended on the probability and type of target talker alterations. Again, no significant age-group differences were found. No significant associations of cognitive characteristics and costs could be shown. However, a more fine-grained analysis based on the calculation of general and specific switch costs showed different mechanisms in older and younger listeners.

This study confirms that dynamic cocktail party listening is associated with costs that depend on the type and probability of target switches. It extends previous research by showing that the effects of switching type and probability are similar for younger and older listeners with good hearing and good cognitive abilities. It further shows that, despite comparable costs of dynamic listening, mechanisms are different for the two age groups, as switching auditory attention may be preserved with aging, but monitoring different sound sources appears to be more difficult for older adults.

1. Introduction

Verbal communication often takes place in situations with several people talking simultaneously. These multi-talker situations, frequently referred to as "cocktail-party" listening (Cherry, 1953, Bronkhorst, 2015), pose high demands on both the auditory and the cognitive processes involved: based on an analysis of the auditory scene, the listener must segregate different auditory streams associated with the characteristics of the particular sound sources (Bregman, 1990). Here, spatial and voice cues are important sound characteristics that help to distinguish different talkers in a complex environment. The vital cognitive process associated with cocktail-party listening is attention. Based on the ability to segregate different streams, attention must be directed to the talker of interest and irrelevant competing sound sources must be ignored (Alain and Arnott, 2000; Shinn-Cunningham and Best, 2008).

Cocktail-party listening is not always *static*, with one talker of interest, but is typically *dynamic*, with talkers changing in a sometimes unpredictable manner. Here, the ability to monitor different potential sources of information and to switch attention is crucial. Compared to static situations, dynamic cocktail-party listening causes extra cognitive load reflected in decreased speech recognition performance or increased reaction times. A study by Brungart and Simpson (2007) investigated the ability to switch attention between simultaneously presented sentences of

* Corresponding author. Head Audiology Research, Jean Uhrmacher Institute for Clinical ENT-Research, University of Cologne, Geibelstraße 29-31, 50931, Cologne, Germany.
E-mail address: hartmut.meister@uni-koeln.de (H. Meister).

https://doi.org/10.1016/j.heares.2020.108020
0378-5955/

the “coordinate response measure” (CRM, Bolia et al., 2000). Brungart and Simpson considered different “transition types” (such as changing the target talker’s voice and/or the location) and different “transition probabilities”. Compared to a static multi-talker situation, they found a considerable decline in recognition performance when transitions were introduced and the listeners had to switch attention. Lin and Carlile (2015) also found costs associated with a task that involved switching attention between competing matrix sentences. Switching not only had a negative effect on recognition performance, but also on comprehension—at least when complex questions about sentence content were asked. The authors hypothesized that costs are associated with different cognitive domains, such as working memory capacity (WMC) and executive functions important for cognitive flexibility and task-switching abilities. Indeed, they reported a significant correlation of switch costs and performance in a common memory span task (Reading Span Test). Moreover, a neuropsychological test of visual attention and task-switching speed (Trailmaking Test B) was significantly correlated with word recognition performance.

Cocktail-party listening is expected to be particularly problematic in older listeners due to a typical decline in hearing as well as in particular cognitive functions (Cohen, 1987; Schneider et al., 2010; Humes, 2013). The need to switch attention in dynamic situations may increase cognitive load and be detrimental to speech understanding. For instance, age effects were shown by Tun and Lachman (2008) in a reaction time task performed by telephone. The task was to react to the word “red” by responding “stop” and to the word “green” by responding “go” in a congruent response rule, and vice versa in an incongruent response rule (“stop and go switch task”). Verbal response times showed a significant slowing in older compared to younger listeners when the task was switched between the two rules, suggesting that executive processes play an important role.

Lawo and Koch (2014) examined the effects of switching auditory attention in younger and older listeners with good hearing. Two competing talkers (one female and one male) uttering single-digit numbers were dichotically presented via headphone and the target talker’s sex was indicated on a computer screen for each trial (cf. Koch et al. (2011)). Participants were asked to categorize the target number as being smaller or larger than five. Trials could either be congruent (both numbers smaller or larger than five) or incongruent (one number smaller, the other one larger than five). While the younger listeners were generally faster than the older subjects, no age-related differences were observed when the target talker was switched from trial to trial**.** Similarly, the “congruency effect” – i.e., more accurate and faster responses to congruent than incongruent stimuli – appeared to be comparable for young and older listeners.

In a follow-up investigation, Oberem et al., (2017) examined whether this finding also applies to a more complex acoustic scenario. The dichotic paradigm used in Lawo and Koch (2014) was extended to a “binaural” paradigm including eight different (simulated) talker locations. In general, the study mirrored the findings of Lawo and Koch (2014)—namely that both a decline in accuracy and increase in reaction times reflected the costs of switching auditory attention, but again age did not reveal a statistically significant effect. However, the authors reported a “trend” toward larger switch costs for the older compared to the younger listeners, with the older listeners showing more than twice the difference in reaction times between switch and non-switch trials. Furthermore, incongruent trials had a larger negative effect in the older listeners than in the younger subjects. It was suggested that this could be a sign of deficits in inhibitory processes, since this congruency effect gives evidence that the non-targets were also processed to a certain degree.

While the Oberem et al. study addressed a relatively complex and realistic multi-talker scenario, the stimuli used (single-digit numbers 1–9) were rather simple and atypical compared to those experienced in everyday communication with running speech. In contrast, Singh et al. (2013) presented whole sentences from four loudspeakers at different locations and examined the time course and costs of misdirecting auditory spatial attention in younger and older adults. The sentences were again taken from the CRM corpus with the target indicated by a particular callsign shown on a computer screen. In each trial, the call sign could change depending on four different pre-defined location certainty cues also displayed on the computer screen. Thus, the participants were required to monitor multiple simultaneous sources based on the degree of uncertainty given by the a priori location cues and to switch their attention based on the callsign. Singh et al. found a significant main effect of target location certainty, but no age group effect or any significant interaction with age. However, in a second experiment with a more complex task, an age-related effect surfaced: older listeners were significantly worse than younger listeners when required not only to switch auditory attention, but also to switch the task by reporting the sentence *opposite* to the indicated target.

Taken together, dynamic cocktail-party listening is of great importance for everyday communication, but has not yet been fully investigated in older people. Some studies exclusively considered young participants, while others included older subjects, but the methods used and the outcomes reported appear to be too heterogeneous to draw firm conclusions on possible age-related differences and the mechanisms involved. The aim of the present study is to provide further knowledge by comparing the speech recognition performance of young and older listeners in both static and dynamic multi-talker situations. In order to give more details about the mechanisms involved, different types of costs were determined. Multi-talker listening was assessed by using a sentence recognition paradigm with three competing talkers, akin to Brungart and Simpson (2007) and Lin and Carlile (2015). A priori information about the target was given in the static situation, whereas the target talker changed in an unpredictable manner in the dynamic situations. In order to gain information about the use of various streaming cues such as voice and spatial location, different types and probabilities of target talker switches were considered. Since both auditory factors and cognitive functions are important for speech recognition in adverse conditions, the individual cognitive abilities of the listeners were taken into account. In particular, executive functions are assumed to play a significant role for multi-talker listening (cf. Meister et al., 2013; Tamati et al., 2013; Lin and Carlile, 2015; Goossens et al., 2017). Thus, besides the ability to focus attention, working memory and cognitive flexibility were determined to represent important core executive functions (Diamond, 2013). In order to rule out gross sensory effects, only listeners without hearing impairment were considered at this stage of the investigation.

We hypothesized that the older listeners would show higher costs than the young listeners in the dynamic situations due to the cognitive load associated with monitoring different potential targets and switching attention. Specifically, we anticipated that older listeners would perform worse when the target talker changes frequently and when they cannot rely on robust cues, such as a consistent target voice, that may help in following the target phrase across switches. Finally, we expected that recognition performance in the dynamic situations would be related to individual cognitive resources of the participants.

2. Methods

2.1. Multi-talker setup

A setup with three competing talkers was used. Sentences were presented via free-field speakers (KRK Rokit 5) at a distance of 1.2 m from the listener's head located at −60° (left), 0° (center), and +60° (right). Participants were instructed to face the loudspeaker ahead all times. The presentation level was 65 dB SPL for each talker. Stimuli were selected from the Oldenburg Sentence Test (OLSA, Wagener et al., 1999), a matrix test comprising sentences such as "Stefan kauft acht nasse Autos" (Stefan buys eight wet cars) or "Doris malt fünf grüne Tassen" (Doris draws five green cups). Each sentence consists of 5 words with 10 alternatives for each word position. Three different talker voices were used. Two of these voices were the male (mean fundamental frequency F0 = 122 Hz) and the female (F0 = 202 Hz) voices of the original OLSA recordings. An additional female voice with a low fundamental frequency (F0 = 143 Hz) was created by modifying the OLSA female talker in terms of the fundamental frequency and the formant frequencies using PRAAT (Boersma and Weening, 2001) and as described in Darwin et al. (2003). Pilot testing revealed that this manipulation resulted in a realistic impression of a female talker with a deep voice. The rationale behind this approach was to provide the listeners with robust stream segregation cues based on both voice characteristics and spatial separation (Kidd et al., 2005; Kitterick et al., 2010). In each trial, three competing sentences from the OLSA were presented simultaneously, one from each of the three talkers. The target sentence was always indicated by the keyword "Stefan", the other two were masking sentences (compare Fig. 1). All sentences had the same length (2.5 s), but differed words at all five word positions. The listeners' task was to repeat back as many words as possible from the target sentence while ignoring the masking sentences.

This setup was used to assess static and dynamic multi-talker listening including both non-switch and switch trials (Fig. 1a and b). In the *static* condition, the target talker and location remained the same across trials, exclusively representing non-switch trials. A priori information about the target was given by indicating that the target voice and location were fixed for a sequence of ten consecutive presentations. This was done for all possible combinations of voice and location (3 x 3), resulting in a total of 90 trials that were grouped into nine test lists of ten trials each. Due to the a priori information, the listeners were able to pre-focus their attention on the upcoming target before the keyword ("Stefan") was presented.

In the *dynamic* condition, no a priori information about the location and the voice of the target talker were provided, requiring the listeners to monitor several potential sources of information and to (re-)focus their attention as soon as the keyword was recognized. Similar to the study by Brungart and Simpson (2007), two different switching types were taken into account: in switching type A (STA), each talker remained at a fixed location and the switch involved the target sentence moving to a different talker. Thus, as in everyday life, both the voice and the location changed with a target switch. In contrast, in switching type B (STB), the target voice remained the same, but changed location. Here, the participants were aware of who the target talker was, but not of where the target phrase was presented. We expected a constant target voice would help with following the target across switches and that this would be especially beneficial for older listeners to reduce the presumably high cognitive load in the dynamic condition.

The two switching types are illustrated in Fig. 2 by showing arbitrary examples of three consecutive trials: for STA in the first trial, the target was presented by the male talker at 0°, and in the next trial by the high female voice at −60° followed by the deep female voice at +60°. In the STB condition, only the target position changed (0°, followed by +60° and −60°, as an example). Assignments of the target locations and voices were predefined using a quasi-random scheme and were not exactly uniformly distributed across a test list. Hence, the listeners could not make any predictions about the upcoming target.

For each switching type, two different switching probabilities were applied, namely 20% (SP20) and 100% (SP100). At SP100, a switch occurred after each trial in a given test list (30 trials), thus purely consisting of switch trials (i.e., "pure" block, see Fig. 1b). At SP20, six switches occurred per list at random intervals of 2−5 trials, and the remaining 24 trails were non-switch presentations, thus constituting a block of "mixed" trials. The sequences remained the same for each participant, but were different for each switching type. In total, four different dynamic conditions (STA_SP20, STA_SP100, STB_SP20, STB_SP100) were used.

2.2. Neuropsychological tests

All participants completed three neuropsychological tests tapping into cognitive domains expected to be associated with multi-talker listening: focused/sustained attention, working memory capacity and cognitive flexibility.

2.2.1. Focused/sustained attention

The test d2-R (Brickenkamp et al., 2010) was used to assess focused and sustained attention. This is a paper and pencil test

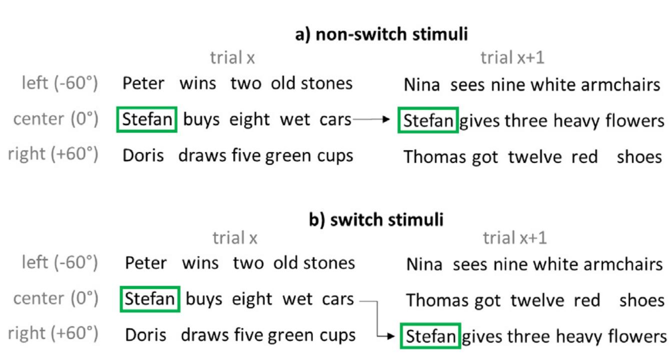

Fig. 1. Illustration of the non-switch and switch trials used with the multi-talker paradigm. Three matrix sentences were always presented simultaneously at −60°, 0°, +60°. The keyword "Stefan" indicated the target phrase.

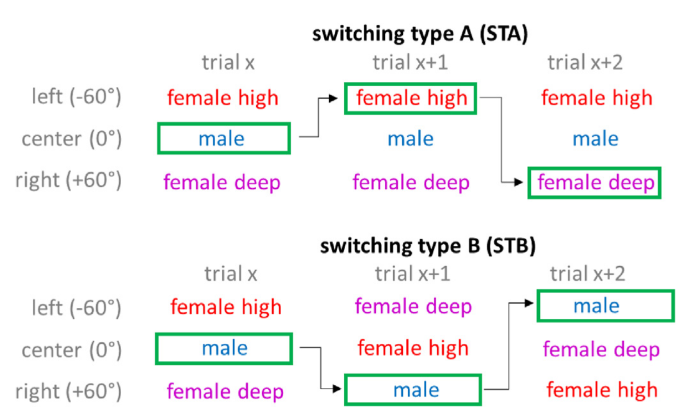

Fig. 2. Illustration of the two switching types. In STA, talkers remain at fixed locations, whereas in STB, talkers change their location but the target voice remains constant. Arbitrary examples are given for switches, with the green box indicating the target.

during which participants are instructed to identify targets (letter “d” with two dashes) located between distractors (letter “p” with zero to four dashes and “d” with zero, one, three and four dashes) in fourteen random sequences. For each sequence, 20 s are given. The number of correctly identified items (i.e., total number of processed items minus the amount of omissions and false alarms) is taken as the outcome score.

2.2.2. Working memory capacity

WMC was assessed using the German version of the Reading Span Test (RST, Carroll et al., 2015). In brief, this test presents sentences in blocks of 2–6 stimuli on a computer screen. The task is to read each sentence aloud and to judge immediately after presentation whether the sentence is meaningful or not. After each block, participants are required to recall either the first or the last word of each sentence. The percentage of correctly recalled words is the outcome measure.

2.2.3. Cognitive flexibility

The Trailmaking Test (“TMT”, Reitan, 1958) is a paper and pencil test that requires connecting numbers (version A) or numbers and letters (version B) shown on a sheet in ascending order. The time required for completing the task is used as the outcome. In version B, participants are required to switch between numbers and letters. A measure of cognitive flexibility was obtained by dividing the outcome of the Trailmaking Test B by the outcome of the Trailmaking Test A, as suggested by Arbuthnott and Frank (2000), and referred to in the following as TMT_B:A. This measure cancels out the pure speed component, which influences the outcome of both test versions and highlights the challenge associated with switching attention (from numbers to letters and vice versa) in version B. Larger values of TMT_B:A are associated with greater difficulty in version B compared to version A and thus reflect lower executive control and cognitive flexibility (Arbuthnott and Frank, 2000; Kortte et al., 2002).

2.3. Procedures

Testing took place in a double-walled sound-attenuating and sound-treated booth (Gravenstein Acoustics, Germany, w: 3 m, l: 4 m, h: 2.4 m) during one single session of about 2.5 h in total. Participants were permitted to take several breaks in order to avoid fatigue. Each test session began by explaining the study’s objectives to the participant and obtaining their written informed consent. The study was approved by the local ethics committee and followed the guidelines of the Helsinki declaration. Thereafter, a pure-tone audiogram and the three neuropsychological tests were conducted, starting with the Trailmaking Test (A and B), followed by the test d2-R and the RST. Each test was introduced in a short familiarization phase as per the test instructions. After that, the multi-talker speech recognition tests were introduced in a more extensive familiarization and training phase. Various sentences from the OLSA were presented, the three voices were introduced and the multi-talker setting was demonstrated. Next, a list of 30 trials was presented in order to demonstrate the different possible conditions (i.e., static condition, dynamic conditions including switching type and switching probability). The familiarization phase was completed once it was certain that the participant had understood all tasks.

During the actual test, the different conditions were examined in quasi-random order to balance conditions and avoid strategies that could emerge due to expectations about the next condition. Each dynamic condition was assessed with one test list of 30 presentations (i.e., 30 target sentences and 60 simultaneous masker sentences per condition). The static condition was assessed using all combinations of voices and locations, as described above. The order of the conditions was the same for each participant, but reversed for half of them for the purpose of compensating possible learning or fatigue effects.

2.4. Participants

Data from 32 listeners were included in the study, 18 older listeners aged 62–78 years (13 women, 5 men, mean 69.7 years) and 14 younger listeners aged 18–33 years (12 women, 2 men, mean 24.6 years). Inclusion criteria were good hearing as assessed by pure-tone audiometry, no signs of attentional disorders or other general health issues, and German as a native language. All listeners had normal or corrected vision and passed dementia screening (DemTect, Kalbe et al., 2004). Better-ear pure-tone thresholds (BEHL) were always at or smaller than 25 dB HL when averaged across 0.5, 1, 2 and 4 kHz and thus did not exhibit hearing impairment according to the WHO definition (Mathers et al., 2000). For convenience, the groups are referred to as young normal-hearing listeners (YNH) and old normal-hearing listeners (ONH). Pure-tone thresholds of the two study groups are given in Fig. 3.

Descriptive statistics based on the Kolmorogov-Smirnov and Shapiro-Wilk tests and visual inspection of quantile-quantile (Q-Q) plots indicated that BEHL was normally distributed in both groups. Independent sample t-tests revealed that BEHL was significantly ($t = 8.22$, $p < 0.001$) higher in the older listeners (mean 13.3 ± 4.87 dB HL) than in the younger listeners (mean 1.7 ± 2.33 dB HL). Thus, despite both groups fulfilling the “normal-hearing” criterion and due to the variance in hearing loss, BEHL was taken into account as a factor in the statistical analysis.

The outcome of the neuropsychological tests are shown in Table 1. Kolmorogov-Smirnov, Shapiro-Wilk tests and visual inspection of Q-Q-plots indicated that the TMT_B:A scores were non-normally distributed for both groups. Thus, possible group differences were analyzed by performing Mann-Whitney U-Tests for TMT_B:A and independent samples t-tests for the outcome of the test d2-R and the RST. Significant group difference were found for the outcome of the test d2-R ($t = 4.15$, $p < 0.001$) and the RST ($t = 2.60$, $p = 0.014$). In contrast, the group difference for TMT_B:A did not reach significance ($U = 101.5$, $p = 0.357$). Closer inspection of the data revealed one extreme outlier for the Trailmaking Test in the older listeners (TMT_B:A = 9.1)

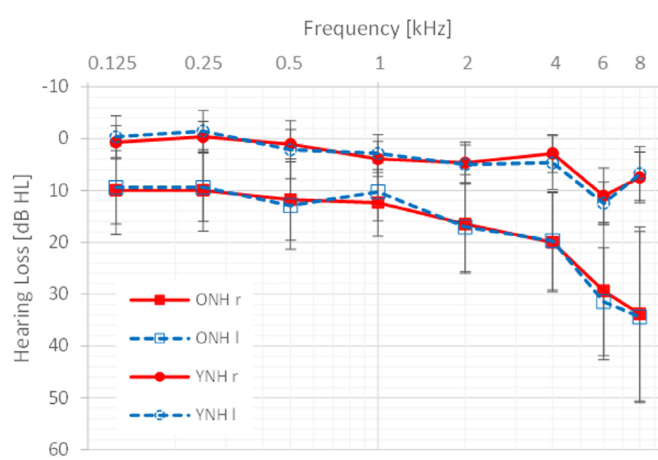

Fig. 3. Pure-tone thresholds of the older (ONH) and younger listeners (YNH) separated for the left (l) and right (r) ear.

Table 1
Outcome of the neuropsychological tests for the older (ONH) and younger listeners (YNH).

	Test d2-R [items]			RST [%]			TMT_B:A		
	mean	median	sd	mean	median	sd	mean	median	sd
ONH	138.6	135.5	27.8	47.3	52.0	12.3	2.61	2.20	1.75
YNH	182.3	175.5	31.8	58.2	57.5	11.2	2.11	1.90	0.83

2.5. Data analysis

2.5.1. Outcome measure for multi-talker listening

Analysis did not consider the redundant keyword ("Stefan"), but only the four other words in each target sentence—i.e., 120 items per test list (30 presentations x 4 target words) in each of the dynamic conditions and a total of 360 items (10 presentations x 3 voices x 3 locations x 4 target words) in the static condition. The percentage of correctly recognized target words was used as the outcome measure.

2.5.2. Costs of dynamic vs. static listening

In order to focus on the load associated with dynamic cocktail-party listening, costs were calculated as the difference between the individual performance in the static and dynamic condition. Importantly, this also considered differences in stimulus presentation between the different conditions, since assignments of the target locations and voices were not uniformly distributed across a test list, as described above. For instance, during 12 items of the test list, the target sentence was presented by the deep female voice at -60°, 10 times by the high female voice at $+60^{\circ}$ and 8 times by the male voice at 0°. The frequency of presentation was then used as a weighting factor, allowing the outcome in the static condition to be calculated as an individual baseline: if the outcome was 95% correct for the deep female voice at -60°, 100% for the high female voice at $+60^{\circ}$ and 90% for the male voice at 0°, the resulting baseline score would be 95.33% (i.e., $(12 \cdot 95\% + 10 \cdot 100\% + 8 \cdot 90\%)/30$). This score was then used as the individual static baseline for calculating the costs in the dynamic conditions.

2.5.3. General and specific switch costs

In order to examine the underlying mechanisms in dynamic multi-talker listening in more detail, general and specific switch costs were determined—two concepts widely used in cognitive psychology for addressing switching abilities (Rogers and Monsell, 1995; Craik and Bialystok, 2006). These concepts assume that trials are presented both in "pure" blocks (i.e., test lists exclusively containing non-switch *or* switch trials) and "mixed" blocks (i.e., test lists containing non-switch *and* switch trials). In the present study, pure blocks were used for the static condition (only non-switch trials) and for the dynamic conditions with a switching probability of 100% (SP100, only switch trials). Mixed blocks were applied for the dynamic conditions with a switching probability of 20% (SP20).

General switch costs, sometimes also referred to as global costs or mixing costs, are calculated by the performance difference in non-switch trials in a pure block and non-switch trials in a mixed block. They are thought to reflect maintaining and scheduling different mental task sets as well as the control processes necessary for addressing stimulus ambiguity (Kray and Lindenberger, 2000; Rubin and Meiran, 2005). In our case, general switch costs were calculated as the difference in performance between the trials in the static situation (pure block) and the non-switch trials in the dynamic situations with a switching probability of 20% (mixed block), and thus reflect the need to monitor different potential sources in the dynamic situation, where no a priori target information was given.

Specific switch costs, sometimes also referred to as local costs, are calculated by the performance difference in non-switch and switch trials in the mixed block and reflect reconfiguration processes associated with changing task or stimulus sets across trials (Rogers and Monsell, 1995). In our case, specific switch costs were assessed in the dynamic situations with a switching probability of 20% and represented the performance drop in the trials where the target sentence changed.

2.5.4. Statistical analysis

Three separate statistical analyses using generalized linear mixed models (GLMMs) were performed separately for the static situation and the costs in the dynamic conditions as well as for the assessment of general and specific switch costs. GLMMs represent a special form of linear mixed models and were used since they allow handling more complex experimental designs with unbalanced groups and data that must not necessarily be normally distributed (McCulloch and Searle, 2001; Baayen et al., 2008; Gordon, 2019). Moreover, GLMMs allow the consideration of random effects to control for individual variability when identifying relationships between predictors and outcomes, and thus give more detailed information on how individual factors contribute to the results observed. By handling the participants as random-effect factors, their heterogeneity was taken into account in all analyses. To judge the goodness of the models, Akaike Information Criterion (AIC, Akaike, 1974) was used.

The variables used for these models are outlined in Table 2. We expected that, despite all subjects being classified as non-impaired listeners, individual hearing thresholds could affect performance in both the static and the dynamic situation. Thus, better ear hearing loss (BEHL) was included as independent variable in all models. Additionally, since we hypothesized that the cognitive factors captured by the neuropsychological tests would be associated with the load in the dynamic situations, these variables are included in model II, addressing the costs of dynamic vs. static listening.

In general, p-level was set at 0.05 and Bonferroni-corrected for multiple comparisons. All analyses were performed using IBM SPSS 23.

3. Results

3.1. Static multi-talker listening

In the static condition—with a priori knowledge of the target talker's voice and location—overall performance was 96% correct for the older listeners and 97% for the younger listeners. Fig. 4 shows word recognition scores for the different voices, locations, and groups. It appears that the center location yields somewhat lower scores than the lateral locations and that the largest voice and group differences occur at 0°. A GLMM with voice, location and group and their two- and three-way interactions as well as BEHL as fixed-effect factors and subjects as random-effect factor (see Table 2) revealed a significant main effect of location ($F(2, 269) = 21.563$, $p < 0.001$). Moreover, significant interactions of voice x location (($F(4, 269) = 7.945$, $p < 0.001$), voice x group (($F(2, 269) = 3.478$, $p = 0.032$) and location x group (($F(2, 269) = 9.778$, $p < 0.001$) were found.

The complete set of parameter estimates of the GLMM is given in Table 3. Reference levels are location = right ($+60^{\circ}$), voice = deep female and group = YNH listeners.

Post hoc tests based on pairwise contrasts showed that the center location revealed significantly lower scores than at -60° ($t = -3.157$, $p < 0.001$) and at $+60^{\circ}$ ($t = -4.263$, $p < 0.0001$). Moreover, performance at -60° was significantly lower than

Table 2
Variables used in the models of static multi-talker listening (model I), costs in the dynamic conditions (model II) and switch costs (model III).

Model I: Static multitalker listening	
dependent variable	percentage correct target words
fixed effects (within subjects)	location (−60°, 0°, +60°), voice (high female, deep female, male)
fixed effects (between subjects)	listener group (YNH, ONH), BEHL
Random effects	subjects
Model II: Costs of dynamic vs. static listening	
dependent variable	costs
fixed effects (within subjects)	switching type (STA, STB), probability (SP20, SP100)
fixed effects (between subjects)	listener group (YNH, ONH), BEHL, TMT_B:A, test d2-R, RST
random effects	subjects
Model III: Switch costs	
dependent variable	switch costs
fixed effects (within subjects)	cost type (general, specific), switching type (STA, STB)
fixed effects (between subjects)	listener group (YNH, ONH), BEHL
random effects	subjects

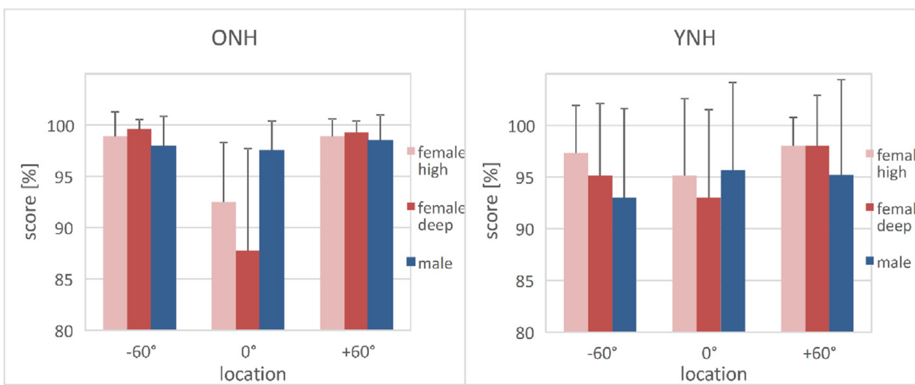

Fig. 4. Correctly recognized target words in the static condition for the older (left panel) and the younger listeners (right panel). Male voice (blue), high female voice (light red), deep female voice (deep red, color online).

Table 3
Parameter estimates of model I (static multitalker listening). Reference levels of the discrete variables are location = right (+60°), voice = deep female, and group = YNH listeners.

Model I				
Term	Estimate	Std.-error	t	p
(Intercept)	4.598	.0033	1398.34	.000
location = left	-.022	.0092	−2.442	.015
location = center	-.072	.0173	−4.178	.000
[voice = male]* [location = left]	-.025	.0117	−2.127	.034
[voice = high female]* [location = left]	.005	.0100	.495	.621
[voice = male]*[location = center]	.062	.0175	3.560	.000
[voice = high female]* [location = center]	.038	.0194	1.938	.054
[voice = male]*[location = right]	-.022	.0068	−3.271	.001
[voice = high female]* [location = right]	-.006	.0041	−1.339	.182
[voice = male]*[group = ONH]	.017	.0087	1.992	.047
[voice = high female]* [group = ONH]	.000	.0061	.061	.951
[voice = deep female]* [group = ONH]	.000	.0044	.089	.929
[location = left]*[group = ONH]	.017	.0086	1.984	.048
[location = center]*[group = ONH]	-.029	.0105	−2.764	.006

at +60° ($t = -2.523$, $p = 0.012$).

The voice × location interaction was mainly due to the fact that significant voice differences were predominantly found for the center position. Here, the deep female voice revealed poorer performance than both the high female voice ($t = -1.977$, $p = 0.049$) and the male voice ($t = -4.318$, $p < 0.001$), and the high female voice revealed poorer performance than the male voice ($t = -2.978$, $p = 0.006$). At +60°, only the difference between the deep female voice and the male voice was significant ($t = -2.755$, $p = 0.019$), and at −60°, only the difference between the high female voice and the male voice ($t = -2.300$, $p = 0.022$) were significant.

Pairwise contrasts for the voice × group interaction failed significance (all $p > 0.071$), whereas for the location × group interaction, the ONH listeners performed significantly worse in the center position ($t = -2.660$, $p = 0.008$), but had a significantly better performance than the YNH listeners at −60° ($t = 1.211$, $p = 0.012$).

3.2. Dynamic multi-talker listening

Fig. 5 shows the costs in the dynamic conditions relative to the static baselines. Costs appear to be higher for 100% switching probability compared to 20% and for switching type A compared to switching type B. A GLMM on costs with switching type, switching probability, group and their two- and three-way interactions as well as TMT_B:A, RST, test d2-R and BEHL as fixed-effect factors was performed. Subjects were considered as a random-effect factor (see Table 2). The model revealed significant effects of switching type ($F(1, 112) = 8.966$, $p = 0.003$) and switching probability ($F(1, 116) = 33.749$, $p < 0.001$). As mentioned in the participants' description above, one extreme outlier was detected with the Trailmaking Test. However, excluding this subject from the analysis did not significantly change the results.

The complete set of parameter estimates of the GLMM is given in Table 4. Reference levels are switching type = STA and switching probability = SP20.

Pairwise contrasts revealed that costs were higher for STA compared to STB ($t = 2.868$, $p = 0.005$) and higher for SP100 compared to SP20 ($t = 5.383$, $p < 0.001$).

3.3. General and specific switch costs

The load due to monitoring several possible sources of information as well as switching attention to a new target was dissociated by determining general and specific switch costs. This was feasible for the conditions considering a 20% probability of a target switch that presented both switch and non-switch trials. For both switching types (STA, STB), specific switch costs appear to be higher than general switch costs (Fig. 6). However, the pattern seems to be dissimilar in younger and older listeners.

A GLMM on switch costs was computed, with cost type, switching type, group and BEHL as fixed-effect factors as well as subject as a random-effect factor (see Table 2). Moreover, two- and three-way interactions of cost type, switching type and group were considered as fixed-effect factors. This analysis revealed a significant main effect of cost type ($F(1, 119) = 20.394$, $p < 0.001$) and a significant cost type × group interaction ($F(1, 119) = 6.715$, $p = 0.011$).

Again, the complete set of parameter estimates of the GLMM is given in Table 5. Reference levels are cost type = specific, switching type = STB.

Pairwise contrasts revealed a significant group difference for general switch costs ($t = 2.206$, $p = 0.029$), but not for specific switch costs ($t = -1.418$, $p = 0.159$).

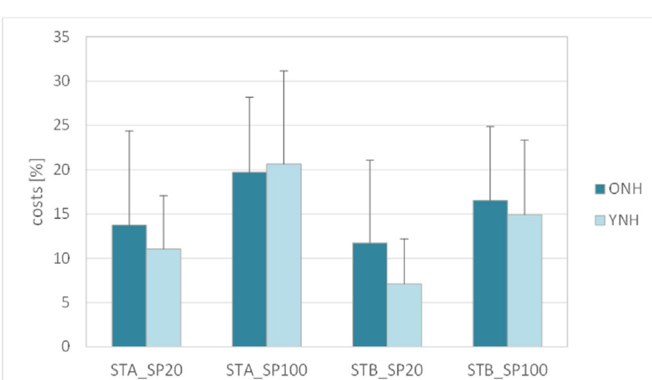

Fig. 5. Costs of dynamic multi-talker listening relative to the static situation given for the two different switching types (STA, STB) and switching probabilities (SP20, SP100).

4. Discussion

This study examined different mechanisms in static and dynamic multi-talker situations with three competing talkers. We were particularly interested in age-related effects and thus compared word recognition performance in younger and older listeners. We hypothesized that our older listeners—despite their good hearing abilities—would have more difficulty especially in dynamic multi-talker conditions due to higher cognitive demands than in the static condition, where a priori information about the target talker was provided. We further anticipated that the performance in the dynamic conditions is associated with general cognitive abilities, particularly core executive functions such as cognitive flexibility and working memory capacity.

4.1. Static multi-talker listening

In the static situation where the target talker was indicated in advance, both listener groups performed at a very high level. Target sentences were always masked by two simultaneous non-target sentences. In multi-talker listening, both energetic masking (EM) and informational masking (IM) play a role. While EM is thought of as a more peripheral effect due to the spectro-temporal overlap of the competing signal energy, IM addresses the additional irrelevant information given by the simultaneously presented masker sentences and thus reflects a higher-level mechanism (cf. Kidd et al., 2008). The amount of IM is affected by target-masker similarity and stimulus uncertainty. Due to the high similarity of the competing sentences' structure, it can be assumed that IM was high, despite the favorable target-to-masker ratio applied. However, this considerable masking was obviously overcome by strong grouping cues (namely location and voice), likely helping the listeners to segregate the competing speech streams and focus attention on the pre-known talker of interest. Despite the good hearing of the older subjects, it could have been expected that their slight high-frequency hearing losses may compromise the use of interaural amplitude differences, and age-related declines in suprathreshold auditory temporal processing may compromise their ability to use inter-aural temporal cues. However, we did not find evidence for such effects since the ability to segregate the talkers obviously was largely preserved in the older listeners.

Statistical analysis revealed a main effect of talker location. When the target talker was presented at 0° azimuth, performance scores were significantly lower than when presented at +60° or at −60°. This finding may be explained by "better ear listening" with lateral locations yielding a more favorable target-to-masker ratio than the center location due to the head shadow effect. Moreover, performance was significantly better at +60° than −60° and may thus indicate a right ear advantage (see Hugdahl and Westerhausen, 2016 for a review). The analysis also revealed a significant location × group interaction. The YNH listeners showed significantly better performance than the ONH for the center location, whereas the opposite held true for −60°. Thus, better ear listening appeared to be more pronounced in the older participants.

Table 4
Parameter estimates of model II (costs of dynamic vs. static listening). Reference levels of the discrete variables are switching type = STA, switching probability = SP20.

Model II				
Term	Estimate	Std.-error	t	p
(Intercept)	2.653	.1116	23.773	.000
switching type = STA	.205	.0701	2.932	.004
switching probability = SP20	-.419	.0732	−5.719	.000

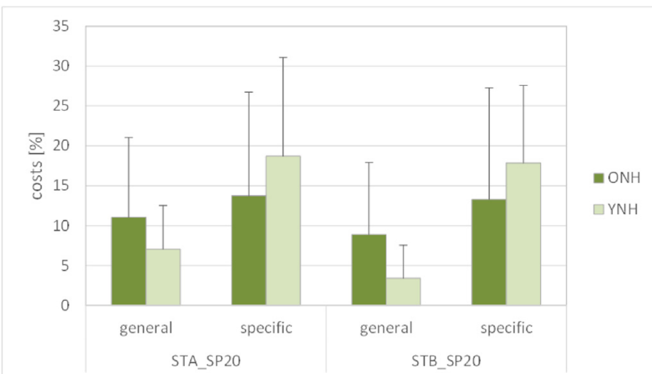

Fig. 6. General and specific switch costs of the two listener groups for the two switching types (STA, STB).

The finding that the young listeners were better at the relatively poorer TMR associated with the center location is in line with previous research showing worse performance for (good-hearing) older listeners when speech is masked by multi-talker babble (cf. Tun and Wingfield, 1999; Gordon-Salant, 2005). However, why there was a significant age effect in the opposite direction at the left talker location is unclear, since there were no differences in instructions or presentation schemes for the two groups. Closer inspection of the data revealed that the YNH displayed larger variability for this condition since two listeners showed a performance slightly below 90%, whereas all others performed at 95% or above. Moreover, a significant location × voice interaction was found. This was mainly due to the fact that voice differences were especially pronounced for the center location. Thus, the availability of richer binaural cues for the target on ±60° may have reduced voice differences at the lateral locations. However, it is important to note that we did not have any specific hypotheses regarding an interplay of voice and location and are not aware of similar findings in the literature. Hence, despite the statistical significance of these interactions, they should be interpreted with caution. Moreover, the small differences and the near-ceiling performance make the evaluation of these effects difficult.

In general, our results from the static multi-talker situation are in line with studies showing that both young and older listeners make use of better ear listening with spatially distributed sound sources (cf. Avivi-Reich et al., 2015). They are also in line with results reported by Gygi and Shafiro (2012, 2014), who showed that young and older listeners basically performed similarly on concurrently presented CRM sentences when they were spatially separated. In contrast, our results appear to be inconsistent with the studies by Besser et al. (2015) and Ezzatian et al. (2015), who showed poorer performance for older listeners when competing speech signals were presented—possibly pointing to age-related difficulties in auditory stream segregation. However, this inconsistency might be due to the fact that these studies adaptively measured speech recognition thresholds associated with 50% intelligibility, whereas our study (as well as Gygi and Shafiro, 2012, 2014) used a relatively favorable fixed target-to-masker ratio that does not necessary reflect difficulties associated with conditions revealing higher masking energy.

4.2. Dynamic multi-talker listening

Cost calculations revealed that performance in the dynamic listening conditions was about 10–20% lower than in the static situation. Costs could even be found in conditions with rare switches (SP20), but they were significantly more pronounced in the condition with switches on every trial (SP100). This seems reasonable since permanently redirecting attention appears to be more demanding than directing attention on a target that remains constant over a number of trials. However, the fact that a performance decline relative to the static situation was also associated with rare switches shows that monitoring different sources due to missing a priori information about the target also causes cognitive load. This is in line with Kitterick et al. (2010), who determined the benefit of knowledge about the voice, location and time of utterance of the target talker in multi-talker situations. Using the coordinate response measure (CRM) in young normal-hearing listeners, they found a SRT improvement of 3.2 dB SNR when a priori voice information was given compared to the condition when the target talker's voice was uncertain. Additional a priori location information improved the SRT by 5.1 dB SNR.

In our experiment, switching type A (STA) caused significantly higher costs than switching type B (STB). This may be due to the fact that the target voice remained fixed in STB, whereas both voice and location could change in STA, and thus none of the cues offered any usable information about the target talker. This result is consistent with a report by Samson and Johnsrude (2016), who found a small but significant performance improvement in a three-talker mixture, when the target voice was constant rather than altered across trials. Furthermore, Lin and Carlile (2019) recently showed the benefit of a constant voice on costs in a non-spatial attention

Table 5
Parameter estimates of model III (switch costs). Reference levels of the discrete variables are cost type = specific, switching type = STB.

Model III				
Term	Estimate	Std.-error	t	p
(Intercept)	.731	.731	7.241	.000
Cost type = general	-.525	.1088	−4.821	.000
[cost type = general]*[group = ONH]	.191	.0865	2.206	.029
[cost type = specific]*[group = ONH]	-.191	.1345	−1.418	.159

switching paradigm. This confirms that voice continuity is a strong cue for guiding attention in dynamic listening situations. Nevertheless, keeping the voice of the target talker fixed in STB did not reduce costs to zero, in line with Best et al. (2008, 2010, cited in Bronkhorst, 2015), who showed that the beneficial effect of a constant voice almost vanished when the target location changed.

We hypothesized that a constant voice would be especially helpful for the ONH group to reduce the detrimental effect of cognitive load in the dynamic situation. This expectation is consistent with the findings of Johnsrude et al. (2013), who showed that older listeners were able to use their familiarity with a speaker's voice to overcome possible effects of sensory and cognitive decline in a competing talker condition. However, since we did not find a significant switching type $\times$ group interaction, it can be concluded that both listener groups used voice information to follow the target phrases across switches.

Despite differences in methods and materials, our study confirms the results by Brungart and Simpson (2007). They also used a three-talker setup and applied different "transition types" (our STA corresponds to their TT I and our STB to their TT II) and different transition probabilities (0, 1/6, 1/3, 2/3, 1) in nine normal-hearing listeners. Unfortunately, the age of their listeners was not reported. However, both studies found increased performance when the voice of the target talker remained constant (in STB and TT II) rather than varying across trials (in STA and TT I). Moreover, in both investigations, performance decreased considerably even when switches were rare (i.e., 20% or 1/6) relative to the static situation (i.e., 0% switches). A direct comparison between the two studies is difficult since the results reported in Brungart and Simpson are either collapsed across the different transition types or the different transition probabilities. An exception was the outcome with a transition probability of 0%, where Brungart and Simpson reported about 90% correct responses and our results approached 96–97% on average, despite the somewhat simpler structure of the CRM (two target words) compared to the OLSA (four target words). The lower values in the former study might be due to different reasons. First, with the CRM, only the male speakers were presented and an additional spatial configuration with narrow talker separation ($-15°$, $0°$, $+15°$) was used, which may have reduced the ability to segregate the auditory streams. Second, it is not clear if the listeners had a priori knowledge of the target talker in this condition. Thus, target uncertainty and the need to monitor different potential sources of information may also have affected the results, an effect addressed in more detail below.

We further hypothesized that the older listeners would perform worse than the younger listeners, especially for frequent target switching (i.e., SP100). However, though a significant main effect of switching probability occurred, no significant interaction with the listener groups was observed. Still, the apparent similarity in performance does not imply that identical mechanisms are at work in both groups—a matter we discuss more specifically in the following.

4.3. General and specific switch costs

General and specific switch costs were determined in order to disassociate effects related to monitoring different potential sources of information (i.e., performance in the non-switch trails in the mixed block vs. non-switch trials in the pure block) and refocusing attention when a switch has occurred (i.e., performance in the switch vs. non-switch trials in the mixed block).

The results in Fig. 6 indicate that, while specific switch costs are basically higher than general switch costs, the pattern is different in younger and older listeners. This is supported by the significant group x cost type interaction: while older listeners experienced comparably large general switch costs, the opposite appears to hold for specific switch costs. Notably, this pattern does not indicate that the YNH listeners show higher absolute switch costs than the ONH, since the performance in the switch trials was very similar in both groups (72.8% correct in older and 73.4% in younger listeners). In fact, the older listeners tended to show worse performance in the non-switch trials (87.2% compared to 91.6% in the younger listeners). Since specific switch costs are calculated as the difference between non-switch and switch trials, they consequently appear to be higher in the younger listeners. However, a follow-up of the interaction revealed a significant group difference for general costs, but not for specific costs. We interpret this pattern as suggesting that missing a priori information about the upcoming target and the corresponding need to monitor different information sources is particularly demanding in older listeners. This interpretation is in line with some previous research. For instance, Getzmann et al. (2017) used a simulated stock-price marketing scenario wherein company names and numbers were presented via four spatially distributed loudspeakers. The task was to indicate whether the number associated with a predefined target company was less than or equal to five or greater than or equal to six. According to a pseudorandomized scheme, the target talker (voice and location) could change. Younger and older participants with good hearing were included. Although not discussed in detail, the older subjects performed similarly to younger subjects in the switch trials, but showed about a 5% decline in performance in the non-switch trials (cf. Fig. 2, Getzmann et al., 2017). This observation strongly resembles our findings. Oberem et al. (2017) also compared groups of younger and older listeners using a number categorization task (see Introduction). They did not find a significant group difference for switching costs, but reported that older listeners were more frequently affected by distractor stimuli than younger subjects when non-switch trials were presented.

Our findings are also in line with a number of studies that examine switch costs in non-auditory tasks in older and younger subjects. For instance, reviews by Craik and Bialystok (2006) and Wasylyshyn et al. (2011) show that age is associated with general but not with specific switch costs. Craik and Bialystok (2006) display such a pattern for a broad age range and link this finding to similar trajectories of age-variant measures of conscious processing and age-invariant measures of automatic processing. Such an equivalence is in line with the outcome of the multi-talker experiment by Singh et al. (2008), who conclude that age-related differences are found for slower, more controlled behavior but not for faster, more automatic behavior such as spatial attention switching.

However, contrasting results were reported by Tun et al. (2008). They found a significant association with age for specific costs, but less reliable age effects for general costs. This may be attributed to the complexity of the switching tasks applied. For instance, in a study examining young normal-hearing listeners, Lin and Carlile (2015) calculated specific switch costs and reported that the costs of switching observed in young normal-hearing listeners depended on the cognitive load associated with the task as a whole: costs were significantly larger when the participants were required to answer complex questions about the content of the sentences as compared to simply repeating back the sentences. This is consistent with the conclusions of Singh et al. (2013), who reported that relatively simple attention-switching is preserved in older listeners, but that age group differences become evident in complex tasks (e.g., performing multiple switches at the same time).

4.4. Cognitive abilities and dynamic multi-talker listening

Previous research established significant relationships between several cognitive abilities and speech recognition in multi-talker listening (Meister et al., 2013; Gygi and Shafiro, 2014; Lin and Carlile, 2015). Given that particularly dynamic multi-talker situations are cognitively demanding, we expected to find associations between the outcome of neuropsychological tests and the costs of dynamic vs. static listening. Specifically, we predicted listeners with higher working memory capacity, cognitive flexibility (i.e., core executive functions, cf. Diamond, 2013) and higher attention to show lower costs. However, we did not find evidence that these domain-general cognitive abilities were significantly associated with dynamic multi-talker listening. One could argue that the chosen set of neuropsychological tests may not be representative of the cognitive processes involved in the specific multi-talker situations. However, we have deliberately selected frequently used, well-established tests that specifically cover the domain of executive control. Moreover, very similar methods were used by Lin and Carlile (2015), and they actually did show a significant correlation between WMC (as assessed by the RST) and switch costs in young normal-hearing listeners. One reason for this discrepancy may be that they used six words per stimulus, possibly provoking more memory load, and thereby strengthening the association with individual WMC. Another reason may be that they assessed switch costs, while we took the difference in word recognition between the dynamic and static conditions as the outcome measure. However, even when applying the same metric (i.e., specific switch costs), we did not find a significant association with the results of our neuropsychological tests.

The fundamental rationale behind expecting an association between general cognitive abilities and speech recognition is that the load in demanding listening situations strains cognitive capacity. Given that our older listeners were relatively atypical for their age group in terms of both hearing and cognitive ability, it could have been that our simulated cocktail-party situations were not strenuous enough to deplete cognitive resources—although they were perceived as being demanding. Another argument put forward by Schneider et al. (2010) is that it is not clear whether executive control extends all the way from the more cognitive processing down to perceptual processing of speech stimuli. Thus, taking speech comprehension rather than speech recognition or word identification into account may show a stronger association with cognitive functions. This is compatible with Lin and Carlile (2015), who report larger switch costs when their listeners were required to comprehend the sentences, though they also showed a significant correlation between WMC and speech recognition.

4.5. Summary and future directions

The present study compared static and dynamic multi-talker listening using presentations of three concurrent talkers. Motivated by the work of Brungart and Simpson (2007), different switching types and switching probabilities were taken into account in order to allow a comprehensive examination of mechanisms in dynamic cocktail-party listening. Despite differences in methods and material, our findings largely confirmed those of Brungart & Simpson. Moreover, our study extended their findings by comparing different age groups, examining the underlying mechanisms of attention switching and taking different cognitive abilities of the listeners into account.

When comparing the two age groups, we found similarly high performance in the static condition. Moreover, no significant age-related differences were found when the costs of dynamic listening were analyzed. All older listeners had good hearing and good cognitive abilities. As expected, the older subjects showed somewhat lower working memory capacity and focused attention. The latter may especially be attributed to the speed component of the paradigm used (see test d2-R). However, in terms of cognitive flexibility, both groups had comparable results. This similarity in cognitive ability probably limited potential group differences in the demanding dynamic conditions. The results therefore give evidence that at least older listeners with good hearing and good cognitive abilities are able to segregate different talkers and to direct attention on a target phrase when robust location and voice cues are given.

Notably, the similar costs of dynamic multi-talker listening in both age groups were unexpected. Nevertheless, when applying the concept of general and specific switch costs, dissimilar patterns occurred, providing evidence for different underlying mechanisms. The fact that older listeners revealed significantly higher general costs but similar specific costs to the young listeners showed that age was negatively associated with the need to monitor different potential sources of information, but not with the requirement to refocus attention when the target talker had changed. In terms of everyday communication, this implies that age-related problems might especially occur when it is uncertain who is the target talker in a crowded conversation. Once the target has been identified, refocusing attention appears to be similar in older and younger people. In daily life, identifying the target talker is facilitated by the fact that linguistic context information—but also visual cues—are usually available.

We further hypothesized that performance in the demanding dynamic situation would be related to the cognitive abilities of the listeners, but found no significant association. We are only aware of one study (Lin and Carlile, 2015) that also considered the outcome of neuropsychological tests with a dynamic listening condition, and our results stand in contrast to their finding. However, this might rely on specific design aspects of the studies and the listener groups assessed.

The present study considered healthy older participants with good hearing and good cognition who are not representative of the general population. Future research will examine multi-talker listening in subjects with hearing loss and/or a more pronounced decline in executive functions. A focus on the interplay of sensory and cognitive abilities in the dynamic situation could also provide clues as to whether such paradigms are suitable for promoting ecological validity in clinical audiology—where typically static situations have been considered to date.

Acknowledgements

This study was supported by a grant of the Deutsche Forschungsgemeinschaft (DFG), reference ME 2751/3-1. Parts of this study were presented at the SpiN2019 workshop in Ghent, Belgium, January 10–11, 2019. We are grateful to the participants for their time and dedication and three anonymous reviewers for their helpful comments on a former version of the manuscript.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automat. Contr. 19 (6), 716–723. https://doi.org/10.1109/TAC.1974.1100705.

Alain, C., Arnott, S.R., 2000. Selectively attending to auditory objects. Front. Biosci. 5, D202–D212. https://doi.org/10.2741/alain. Published 2000 Jan 1.

Arbuthnott, K., Frank, J., 2000. Trail making test, part B as a measure of executive control: validation using a set-switching paradigm. J. Clin. Exp. Neuropsychol. Aug 22 (4), 518–528.

Avivi-Reich, M., Jakubczyk, A., Daneman, M., Schneider, B.A., 2015. How age, linguistic status, and the nature of the auditory scene alter the manner in which listening comprehension is achieved in multitalker conversations. J. Speech Lang. Hear. Res. 58 (5), 1570–1591. https://doi.org/10.1044/2015_JSLHR-H-14-0177.

Baayen, R.H., Davidson, D.J., Bates, D.M., 2008. Mixed-effects modeling with crossed random effects for subjects and items. J. Mem. Lang. 59 (4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005.

Best, V., Ozmeral, E.J., Kopčo, N., Shinn-Cunningham, B.G., 2008. Object continuity enhances selective auditory attention. Proc. Natl. Acad. Sci. Unit. States Am. 105,

13174–13178. https://doi.org/10.1073/pnas.0803718105.
Best, V., Shinn-Cunningham, B.G., Ozmeral, E.J., Kopčo, N., 2010. Exploring the benefit of auditory spatial continuity. J. Acoust. Soc. Am. 127, EL258. https://doi.org/10.1121/1.3431093.
Boersma, P., Weenink, D., 2001. PRAAT, a system for doing phonetics by computer. Glot Int. 5 (9/10), 341–345.
Bolia, R.S., Nelson, W.T., Ericson, M.A., Simpson, B.D., 2000. A speech corpus for multitalker communication research. J. Acoust. Soc. Am. 107, 1065–1066.
Bregman, A.S., 1990. Auditory Scene Analysis. The Perceptual Organization of Sound. MIT Press, Cambridge.
Brickenkamp, R., Schmidt-Atzert, L., Liepmann, D., 2010. d2-R - Test d2 – Revision (Aufmerksamkeits- und Konzentrationstest). Hogrefe, Göttingen.
Bronkhorst, A.W., 2015. The cocktail-party problem revisited: early processing and selection of multi-talker speech. Attent. Percept. Psychophys. Jul 77 (5), 1465–1487. https://doi.org/10.3758/s13414-015-0882-9.
Brungart, D.S., Simpson, B.D., 2007. Cocktail party listening in a dynamic multitalker environment. Percept. Psychophys. Jan 69 (1), 79–91.
Carroll, R., Meis, M., Schulte, M., Vormann, M., Kießling, J., Meister, H., 2015. Development of a German reading span test with dual task design for application in cognitive hearing research. Int. J. Audiol. Feb 54 (2), 136–141.
Cherry, E.C., 1953. Some experiments on the recognition of speech, with one and with two ears. J. Acoust. Soc. Am. 25, 975–979. https://doi.org/10.1121/1.1907229.
Cohen, G., 1987. Speech comprehension in the elderly: the effects of cognitive changes. Br. J. Audiol. 21 (3), 221–226. https://doi.org/10.3109/03005368709076408.
Craik, F.I., Bialystok, E., 2006. Cognition through the lifespan: mechanisms of change. Trends Cognit. Sci. 10 (3), 131–138. https://doi.org/10.1016/j.tics.2006.01.007.
Darwin, C.J., Brungart, D.S., 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 (5), 2913–2922.
Diamond, A., 2013. Executive functions. Annu. Rev. Psychol. 64, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750.
Ezzatian, P., Li, L., Pichora-Fuller, K., Schneider, B.A., 2015. Delayed stream segregation in older adults: more than just informational masking. Ear Hear. 36 (4), 482–484. https://doi.org/10.1097/AUD.0000000000000139.
Getzmann, S., Jasny, J., Falkenstein, M., 2017. Switching of auditory attention in "cocktail-party" listening: ERP evidence of cueing effects in younger and older adults. Brain Cogn. Feb 111, 1–12. https://doi.org/10.1016/j.bandc.2016.09.006.
Goossens, T., Vercammen, C., Wouters, J., van Wieringen, A., 2017. Masked speech perception across the adult lifespan: impact of age and hearing impairment. Hear. Res. Feb 344, 109–124. https://doi.org/10.1016/j.heares.2016.11.004.
Gordon, K.R., 2019. How mixed-effects modeling can advance our understanding of learning and memory and improve clinical and educational practice. J. Speech Lang. Hear. Res. 62 (3), 507–524. https://doi.org/10.1044/2018_JSLHR-L-ASTM-18-0240.
Gordon-Salant, S., 2005. Hearing loss and aging: new research findings and clinical implications. J. Rehabil. Res. Dev. 42 (4 Suppl. 2), 9–24. https://doi.org/10.1682/jrrd.2005.01.0006.
Gygi, B., Shafiro, V., 2012. Spatial and temporal factors in a multitalker dual listening task. Acta Acust. Uni. Ac. 98, 142–157.
Gygi, B., Shafiro, V., 2014. Spatial and temporal modifications of multitalker speech can improve speech perception in older adults. Hear. Res. Apr 310, 76–86. https://doi.org/10.1016/j.heares.2014.01.009.
Hugdahl, K., Westerhausen, R., 2016. Speech processing asymmetry revealed by dichotic listening and functional brain imaging. Neuropsychologia Dec 93 (Pt B), 466–481. https://doi.org/10.1016/j.neuropsychologia.2015.12.011.
Humes, L.E., 2013. Understanding the speech-understanding problems of older adults. Am. J. Audiol. Dec 22 (2), 303–305. https://doi.org/10.1044/1059-0889(2013/12-0066.
Johnsrude, I.S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H.P., Carlyon, R.P., 2013. Swinging at a cocktail party: voice familiarity aids speech perception in the presence of a competing voice. Psychol. Sci. 24 (10), 1995–2004. https://doi.org/10.1177/0956797613482467.
Kalbe, E., Kessler, J., Calabrese, P., Smith, R., Passmore, A.P., Brand, M., Bullock, R., 2004. DemTect: a new, sensitive cognitive screening test to support the diagnosis of mild cognitive impairment and early dementia. Int. J. Geriatr. Psychiatry. Feb 19 (2), 136–143.
Kidd Jr., G., Arbogast, T.L., Mason, C.R., Gallun, F.J., 2005. The advantage of knowing where to listen. J. Acoust. Soc. Am. Dec 118 (6), 3804–3815.
Kidd, G., Mason, C.R., Richards, V.M., Gallun, F.J., Durlach, N.I., 2008. Informational masking. In: Yost, W.A., Popper, A.N., Fay, R.R. (Eds.), Auditory Perception of Sound Sources. Springer Handbook of Auditory Research, vol. 29. Springer, Boston, MA.
Kitterick, P.T., Bailey, P.J., Summerfield, A.Q., 2010. Benefits of knowing who, where, and when in multi-talker listening. J. Acoust. Soc. Am. Apr 127 (4), 2498–2508. https://doi.org/10.1121/1.3327507.
Koch, I., Lawo, V., Fels, J., Vorländer, M., 2011. Switching in the cocktail party: exploring intentional control of auditory selective attention. J. Exp. Psychol. Hum. Percept. Perform. Aug 37 (4), 1140–1147. https://doi.org/10.1037/a0022189.
Kortte, K.B., Horner, M.D., Windham, W.K., 2002. The trail making test, part B: cognitive flexibility or ability to maintain set? Appl. Neuropsychol. 9 (2), 106–109.
Kray, J., Lindenberger, U., 2000. Adult age differences in task switching. Psychol. Aging 15 (1), 126–147. https://doi.org/10.1037//0882-7974.15.1.126.
Lawo, V., Koch, I., 2014. Examining age-related differences in auditory attention control using a task-switching procedure. J. Gerontol. B. Psychol. Sci. Soc. Sci. Mar 69 (2), 237–244. https://doi.org/10.1093/geronb/gbs107.
Lin, G., Carlile, S., 2015. Costs of switching auditory spatial attention in following conversational turn-taking. Front. Neurosci. Apr. 20 9, 124. https://doi.org/10.3389/fnins.2015.00124.
Lin, G., Carlile, S., 2019. The effects of switching non-spatial attention during conversational turn taking. Sci. Rep. 9 (1), 8057. https://doi.org/10.1038/s41598-019-44560-1. May 30.
Mathers, C., Smith, A., Concha, M., 2000. In: Organization, W.H. (Ed.), Global Burden of Hearing Loss in the Year 2000. Retrieved from. http://www.who.int/healthinfo/statistics/bod hearingloss.pdf.
McCulloch, C.E., Searle, S.R., 2001. Generalized, Linear, and Mixed Models. John Wiley & Sons, New York.
Meister, H., Schreitmüller, S., Grugel, L., Ortmann, M., Beutner, D., Walger, M., Meister, I.G., 2013. Cognitive resources related to speech recognition with a competing talker in young and older listeners. Neuroscience 232, 74–82. https://doi.org/10.1016/j.neuroscience.2012.12.006. Mar 1.
Oberem, J., Koch, I., Fels, J., 2017. Intentional switching in auditory selective attention: exploring age-related effects in a spatial setup requiring speech perception. Acta Psychol. (Amst). Jun 177, 36–43. https://doi.org/10.1016/j.actpsy.2017.04.008.
Reitan, R.M., 1958. Validity of the Trail Making test as an indicator of organic brain damage. PerceptMotor Skills 8, 271–276.
Rogers, R.D., Monsell, S., 1995. Costs of a pre-dictable shift between simple cognitive tasks. J. Exp. Psychol. Gen. 124, 207–231.
Rubin, O., Meiran, N., 2005. On the origins of the task mixing cost in the cuing task-switching paradigm. J. Exp. Psychol. Learn. Mem. Cogn. 31 (6), 1477–1491. https://doi.org/10.1037/0278-7393.31.6.1477.
Samson, F., Johnsrude, I.S., 2016. Effects of a consistent target or masker voice on target speech intelligibility in two- and three-talker mixtures. J. Acoust. Soc. Am. Mar 139 (3), 1037–1046.
Schneider, B.A., Pichora-Fuller, M.K., Daneman, M., 2010. The effects of senescent changes in audition and cognition on spoken language comprehension. In: Gordon-Salant, S., Frisina, R.D., Popper, A., Fay, R.R. (Eds.), The Aging Auditory System: Perceptual Characterization and Neural Bases of Presbycusis, Springer Handbook of Auditory Research. Springer, Berlin, pp. 167–210. https://doi.org/10.1007/978-1-4419-0993-0_7.
Shinn-Cunningham, B.G., Best, V., 2008. Selective attention in normal and impaired hearing. Trends Amplif. Dec 12 (4), 283–299. https://doi.org/10.1177/1084713808325306.
Singh, G., Pichora-Fuller, M.K., Schneider, B.A., 2008. The effect of age on auditory spatial attention in conditions of real and simulated spatial separation. J. Acoust. Soc. Am. 124 (2), 1294–1305. https://doi.org/10.1121/1.2949399.
Singh, G., Pichora-Fuller, M.K., Schneider, B.A., 2013. Time course and cost of misdirecting auditory spatial attention in younger and older adults. Ear Hear. 34 (6), 711–721. https://doi.org/10.1097/AUD.0b013e31829bf6ec. Nov-Dec.
Tamati, T.N., Gilbert, J.L., Pisoni, D.B., 2013. Some factors underlying individual differences in speech recognition on PRESTO: a first report. J Am Acad Audiol. Jul-Aug 24 (7), 616–634. https://doi.org/10.3766/jaaa.24.7.10.
Tun, P.A., Wingfield, A., 1999. One voice too many: adult age differences in language processing with different types of distracting sounds. J. Gerontol. B Psychol. Sci. Soc. Sci. 54 (5), P317–P327. https://doi.org/10.1093/geronb/54b.5.p317.
Tun, P.A., Lachman, M.E., 2008. Age differences in reaction time and attention in a national telephone sample of adults: education, sex, and task complexity matter. Dev. Psychol. Sep 44 (5), 1421–1429. https://doi.org/10.1037/a0012845.
Wagener, K., Kühnel, V., Kollmeier, B., 1999. Entwicklung und Evaluation eines Satztests in deutscher Sprache I: design des Oldenburger Satztests [Development and evaluation of a German sentence test – Part I: design of the Oldenburg sentence test]. Z. Audiol. 38 (1), 4–15. German.
Wasylyshyn, C., Verhaeghen, P., Sliwinski, M.J., 2011. Aging and task switching: a meta-analysis. Psychol. Aging. Mar 26 (1), 15–20. https://doi.org/10.1037/a0020912.