Benefit of binaural listening as revealed by speech intelligibility and listening effort

Jan Rennies^{a)} and Gerald Kidd, Jr.

Department of Speech, Language and Hearing Sciences, Boston University, Boston, Massachusetts 02215, USA

(Received 27 April 2018; revised 13 September 2018; accepted 13 September 2018; published online 12 October 2018)

In contrast to the well-known benefits for speech intelligibility, the advantage afforded by binaural stimulus presentation for reducing listening effort has not been thoroughly examined. This study investigated spatial release of listening effort and its relation to binaural speech intelligibility in listeners with normal hearing. Psychometric functions for speech intelligibility of a frontal target talker masked by a stationary speech-shaped noise were estimated for several different noise azimuths, different degrees of reverberation, and by maintaining only interaural level or time differences. For each of these conditions, listening effort was measured using a categorical scaling procedure. The results revealed that listening effort was significantly reduced when target and masker were spatially separated in anechoic conditions. This effect extended well into the range of signal-to-noise ratios (SNRs) in which speech intelligibility was at ceiling, and disappeared only at the highest SNRs. In reverberant conditions, spatial release from listening effort was observed for high, but not low, direct-to-reverberant ratios. The findings suggest that listening effort assessment can be a useful method for revealing the benefits of spatial separation of sources under realistic listening conditions comprising favorable SNRs and low reverberation, which typically are not apparent by other means. © 2018 Acoustical Society of America. <https://doi.org/10.1121/1.5057114>

[AKCL]

Pages: 2147–2159

I. INTRODUCTION

The perception of speech in complex acoustic scenes is influenced by many factors including the particularly adverse effects of environmental noise and reverberation. Listeners with normal hearing often are able to overcome these adverse effects and may recognize and comprehend the speech of a target voice even at very unfavorable signal-to-noise ratios (SNRs) when the interfering sounds originate from different spatial locations than the target talker. The advantage of spatial separation of sound sources in this listening situation, frequently termed the “cocktail party effect” (e.g., recent series of reviews in [Middlebrooks et al., 2017](#)) is commonly attributed to spatial “unmasking” resulting from the exploitation of acoustic head-shadow effects and binaural processing of interaural time and phase differences. The focus of most previous studies of the benefits of spatial separation of sound sources has been on the improvements found for *speech intelligibility*, and a large body of literature and modeling efforts exist that characterize the impact of various contributing factors [see [Bronkhorst \(2000, 2015\)](#) and [Mattys et al. \(2012\)](#), for relevant reviews]. In contrast to speech intelligibility, however, the role of spatial release of *listening effort* has received very little attention so far—despite a generally increasing interest in measuring listening effort as an important tool for evaluating communication scenarios. The goal of this study is to systematically investigate both speech intelligibility and listening effort in a

number of spatial listening conditions to better understand the role of spatial release from listening effort in combatting the adverse effects of noise and reverberation.

An important and intuitive approach to investigating how various factors affect speech perception in adverse conditions is to measure speech intelligibility. The potential benefit of listening with two ears for improving speech intelligibility in adverse conditions has been investigated for many years since the pioneering work of [Cherry \(1953\)](#); see also reviews in [Yost, 1997](#); [Carlile, 2014](#)). The literature is in broad agreement that disparities between targets and maskers with respect to interaural level and time/phase differences (ILDs and ITDs/IPDs) can be exploited by the auditory system to reduce masking and segregate the sources in complex sound fields. For example, relative to the reference condition in which a single noise source is co-located with a speech target in an anechoic environment, spatially separating the sources by changing the relative azimuth of the masker can produce considerable reductions (improvements) in speech reception threshold (SRT_{50}) (i.e., the SNR at which 50% of the presented words can be reported correctly). Several studies have investigated this “binaural release from masking” and have reported SRT shifts that often exceeded 10 dB (e.g., [Beutelmann and Brand, 2006](#); [Beutelmann et al., 2010](#); [Lavandier and Culling, 2008](#); [Bronkhorst, 2000](#)). However, this large benefit of spatially separating speech and noise sources measured in anechoic environments is strongly reduced in highly reverberant rooms (e.g., [Beutelmann et al., 2010](#); [Rennies et al., 2011a](#); [Lavandier and Culling, 2008](#)). This reduction in binaural benefit is mainly due to two effects of the reflections at the room boundaries on the noise: first, acoustic head-shadow is

^{a)}Also at: Fraunhofer IDMT, Project Group Hearing, Speech and Audio Technology, D-26129 Oldenburg, Germany. Electronic mail: jan.rennies@idmt.fraunhofer.de

reduced because reflections arrive at the listener's position from various directions and, second, the interferers are more de-correlated between ears so that ITD/IPD differences are more difficult to exploit. This reduction of binaural unmasking adds to the detrimental effect of reverberation and echoes on the speech signal (e.g., Steeneken and Houtgast, 1980; Arweiler and Buchholz, 2011; Rennies *et al.*, 2011a; Warzybok *et al.*, 2013), which is superimposed with delayed and attenuated versions of itself, resulting in temporal smearing and significantly reduced temporal modulations, which are considered crucial for good speech understanding (Steeneken and Houtgast, 1980).

The studies of binaural advantages for speech intelligibility reviewed above, as well as many others reported over the years, have led to a relatively good understanding of the factors that most affect intelligibility in noise and reverberation, and quantitative models have been developed that can predict intelligibility in many binaural conditions with a very high degree of precision (e.g., Lavandier and Culling, 2010; Lavandier *et al.*, 2012; Leclère *et al.*, 2015; Chabot-Leclerc *et al.*, 2016; Rennies *et al.*, 2014b; Hauth and Brand, 2018). In contrast, binaural effects in listening effort have received very little attention. Listening effort can be defined as “the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a listening task” (Pichora-Fuller *et al.*, 2016) and has been the subject of an increasing number of studies in recent years, although these studies differ somewhat in terminology (e.g., “listening effort,” “listening difficulty,” “ease of listening”). Several studies have employed scaling procedures to directly assess perceived listening effort (e.g., Morimoto *et al.*, 2004; Luts *et al.*, 2010; Sato *et al.*, 2012; Rennies *et al.*, 2014a; Schepker *et al.*, 2016; Krueger *et al.*, 2017a; Krueger *et al.*, 2017b). Other studies used a dual-task approach in which an increase in effort was related to the decrease in performance on a secondary task that occurred when the difficulty of a primary listening task was increased (e.g., Sarampalis *et al.*, 2009; Gosselin and Gagne, 2010; Houben *et al.*, 2013; Johnson *et al.*, 2015; Xia *et al.*, 2015). This decrease in secondary task performance was then interpreted as increased cognitive load due to effortful listening. Another class of studies used (neuro-)physiological responses as a measure of listening effort. These include pupil dilation (e.g., Kramer *et al.*, 1997; Zekveld *et al.*, 2010; Zekveld *et al.*, 2014; Ohlenforst *et al.*, 2017b; Wendt *et al.*, 2017), various physiological indices (e.g., Hicks and Tharpe, 2002; Mackersie and Cones, 2011), and/or electroencephalographic (EEG) responses (e.g., Obleser *et al.*, 2012) (for a review, see Klink *et al.*, 2012a,b).

The main reason for the increasing interest in the assessment of listening effort is that it may add information relevant for speech perception in realistic communication situations beyond that which is measurable with speech intelligibility. For example, it has been shown that listening effort ratings can be used to discriminate among listening conditions in which speech intelligibility is at or close to ceiling (e.g., Houben *et al.*, 2013; Rennies *et al.*, 2014a; Schepker *et al.*, 2016; Krueger *et al.*, 2017b), but which still differ considerably in how challenging it is for the listener to

follow the target talker. This observation is particularly relevant for certain practical applications because many everyday communication situations are characterized by positive or only marginally negative SNRs (Smeds *et al.*, 2015) where speech intelligibility typically is very high (at least for normal-hearing listeners). Viewed in this light, listening effort assessment can be a very useful tool for evaluating realistic listening scenarios.

Despite the clear importance of listening effort measurement and the undisputed role of binaural factors in speech perception, very few studies have addressed the specific contribution of spatial separation of sound sources in listening effort, especially in conditions in which speech intelligibility is at ceiling (most of the above-mentioned studies used diotic or monaural stimuli). Some notable exceptions are found in studies that included listening effort assessments in a battery of tests for evaluating the benefit of bilateral hearing aids and cochlear implants (e.g., van Schoonhoven *et al.*, 2016; Devocht *et al.*, 2017). The general conclusion from these studies was that measuring listening effort can be a useful tool in audiological applications and can provide clinically relevant information beyond that which is gleaned from intelligibility measurements. In addition, three recent studies explicitly tested the role of target and masker separation in listening effort using different experimental paradigms: Zekveld *et al.* (2014) measured SRTs and pupil dilation as a measure of cognitive load for co-located and spatially separated target and interferer voices (interferers fixed at 0° or ±90° azimuth) both for same- and different-gender interferers. They found that both spatial separation and gender differences increased speech intelligibility performance (i.e., reduced SRTs). However, cognitive load was only reduced by gender differences, but not by spatial separation, when both conditions were compared at SNRs corresponding to the respective SRTs. Xia *et al.* (2015) argued that level cues might have affected the results of Zekveld *et al.* (2014) because the SRTs differed considerably across conditions and that cognitive load was evaluated at markedly different SNRs even though speech intelligibility performance was controlled. To address that issue, Xia *et al.* (2015) also varied talker gender and spatial separation, but for conditions with equal-level target and interferer voices. They measured cognitive load in a dual-task paradigm. Speech intelligibility was controlled by adjusting the spatial separation until approximately the same speech intelligibility benefit was achieved as was observed for a gender difference between target and interferers. In contrast to Zekveld *et al.* (2014), Xia *et al.* (2015) found that spatial separation produced a *larger* reduction in cognitive load than did gender difference. In a third study addressing binaural listening effort, Andéol *et al.* (2017) used functional near-infrared spectroscopy to measure cognitive load at three SNRs for interferers co-located with, and spatially separated from, the target talker. They also measured speech intelligibility in the same conditions. Their results indicated that spatial separation reduced listening effort at intermediate SNRs, but not at adverse or favorable SNRs. For all combinations of SNR and masker presentation condition, the observed speech intelligibility was below ceiling (i.e., performance scores were always

worse than about 115 rationalized arcsine units, even in the most favorable condition with the highest SNR and spatially separated interferers).

In summary, previous studies have reported contradictory findings concerning the role of spatial separation of sources in perceived listening effort. These studies focused on conditions in which target speech intelligibility was below ceiling, and typically tested only a very limited set of spatial conditions. To the best of our knowledge, no study has systematically investigated binaural effects in intelligibility and listening effort for a broad range of SNRs and performance levels including binaural conditions in which speech intelligibility is at or above ceiling. Therefore it is uncertain whether binaural hearing provides an advantage in these conditions and, if so, what the magnitude of the benefit (i.e., reduction in effort) may be. Furthermore, it also is unclear how the well-known effects of reverberation on the binaural advantage for speech intelligibility in noise relate to the perception of listening effort. The general goal of this study was to improve our understanding of the role of spatial source separation in listening effort. To accomplish this goal, psychometric functions for both intelligibility and listening effort were measured in the same subjects over a broad range of SNRs for conditions producing varying degrees of spatial unmasking in speech intelligibility. Both anechoic and reverberant listening situations were included to enable us to gauge the potential effects of reverberation on spatial release from masking and from listening effort.

II. METHODS

A. Subjects

Ten subjects (six female, four male) aged between 18 and 30 yrs participated in all parts of this study. All had pure tone thresholds not exceeding 20 dB hearing level at audiometric frequencies between 125 Hz and 8 kHz. Nine subjects had English as their native language. One subject had been born in Poland and moved to the U.S. at the age of 1. This subject considered English as the primary language and self-judged English proficiency as “native-equivalent.” Some subjects had experience with speech intelligibility measurements, but none of them had participated in listening effort assessments before. Subjects were paid on an hourly basis for their participation.

B. Stimuli and equipment

The target speech was always uttered by the same female talker and consisted of sentences taken from the American English matrix sentence test (Kollmeier *et al.*, 2015). The sentences consisted of five words in the fixed structure *name-verb-numeral-adjective-object*. For each word group ten alternatives were available, which could be randomly combined to produce syntactically correct, but semantically unpredictable sentences. The test material consisted of 90 such sentences, which were presented to the subjects in lists of 20 sentences (in the training phase) or 30 sentences (all test conditions). Due to the lack of semantic predictability, each sentence appeared to the subjects as one

of any of the 10^5 possible random combinations, and memorizing any of the 90 sentences was not possible. The sentence lists had been optimized to produce highly homogeneous SRTs (see Kollmeier *et al.*, 2015).

The interfering noise was also taken from the American matrix test and had been generated by randomly and repeatedly superimposing sequences of sentences from the target talker, which resulted in a stationary noise with the same long-term spectrum as the target speech. Speech and noise were convolved with head-related room impulse responses (HRIRs) to produce the desired spatial conditions. For the anechoic conditions, HRIRs were taken from the database of Kayser *et al.* (2009). The HRIR had been recorded with a head-and-torso simulator in an anechoic chamber with a distance of 80 cm between the source and the center of the head. The target was always presented from the front (S_0), while the azimuthal angle of the single noise interferer was varied from to be 0° , 45° , 100° , 135° , or 180° (incident from the front, right, or back). In addition, a condition with one noise source in each hemisphere was included, i.e., two uncorrelated portions of the stationary speech-shaped noise of the matrix test were convolved with HRIRs corresponding to -45° and $+135^\circ$, respectively.

To explore the role of ITDs and ILDs for the stimuli of the present study in more detail, two manipulated HRIRs for the noise were included, in which only one of the two cues was preserved. HRIRs for an azimuthal angle of 100° were used, because the largest binaural benefit was expected for the S_0N_{100} condition. The manipulation was conducted in the same way as described by Ellinger *et al.* (2017): HRIRs with ILDs only were created by backward-shifting the left-ear HRIR by $748 \mu\text{s}$, which corresponded to the interaural delay as determined by the maximum of the interaural cross-correlation function ($S_0N_{100, \text{ILD-only}}$). ITD-only HRIRs were created by copying the right-ear HRIR to the left channel, and then forward-shifting this channel by $748 \mu\text{s}$ ($S_0N_{100, \text{ITD-only}}$).

For the reverberant conditions, HRIRs were taken from the study of Rennie *et al.* (2011a). These HRIRs had been generated using CATT Acoustics v8.0a (Gothenburg, Sweden) for a simulated rectangular room of $10 \text{ m} \times 15 \times 3 \text{ m}$ (width $\times$ length $\times$ height) with a reverberation time of about 2.5 s and different distances between target and listener. For this study, HRIRs for a listener close to the talker [receiver 1 at a distance of 0.5 m in Rennie *et al.* (2011a), see their Fig. 1] with relatively high direct-to-reverberant ratios (DRRs) of about -7 dB (see Table I of Rennie *et al.*, 2011a) were selected. In addition, HRIRs for a listener far away from the talker (receiver 4 at a distance of 13 m in their study) with DRRs lower than -17 dB were included. For each listener position the target was always at the same frontal position, while the noise was presented either from the same position (S_0N_0) or from 90° to the right of the listener (S_0N_{90}).

In all conditions of this study the speech level was fixed at 55 dB sound pressure level (SPL). The noise level was varied to produce the desired SNRs. For the reverberant conditions, the SPLs were calculated including the reverberation energy, i.e., based on the root-mean-square (rms) value of the entire speech and noise signals. For each condition, the right ear served as a reference point, i.e., all levels and SNRs

reported in this study were calculated based on the right-ear signals (in asymmetric conditions this was usually the ear with the worse SNR). Stimuli were generated and controlled using MATLAB. For the intelligibility measurements, the AFC-MATLAB framework of Ewert (2013) was used. Listening effort measurements were conducted using a custom MATLAB framework. The digital output was D/A converted RME HDSP 9632 (ASIO) 24-bit sound card and delivered to the subjects via Sennheiser HD280 pro headphones (Wedemark, Germany) in sound-attenuated booths. The setup was calibrated to SPL using a Brüel and Kjær (B&K, Nærum, Denmark) 4153 artificial ear, a B&K 4947 1/2 in. microphone, a B&K ZC-0032 preamplifier, and a B&K 2250 sound level meter.

C. Procedures

1. Speech intelligibility measurements

Speech intelligibility was measured in a closed-set procedure, i.e., subjects could select the words they had recognized after each sentence on a graphical user interface, which consisted of the entire matrix of 50 words. Subjects confirmed their choices by pressing a button, which triggered the presentation of the next sentence. Each condition was measured with a random list of 30 sentences. The initial SNR was always 0 dB. The SNR of the subsequent stimuli was varied adaptively using the procedure described by Brand and Kollmeier (2002). The 30 sentences were equally split into two interleaved tracks, one of which converged to an SNR corresponding to 80% correctly understood words (SRT_{80}), while the other converged to 20% correctly understood words (SRT_{20}). Each combination of HRIR type (anechoic, ILD-/ITD-only, reverberant) was measured completely before a new condition was started. The order of conditions was randomized for each subject. Three lists of 20 sentences adaptively converging to the SRT_{50} were used to familiarize the subjects with the speech material and task to reduce training effects typical for matrix sentence test (Wagener *et al.*, 1999). Half of the subjects conducted the speech intelligibility measurements first, and then the listening effort measurements. The other subjects conducted the listening effort measurements first.

2. Listening effort measurements

Listening effort was measured by categorical listening effort scaling using a constant stimuli procedure. Each presentation consisted of one target sentence temporally centered in a 1-s longer portion of the masker. The sentence and masker were presented in a loop until the subjects made their assessment of “How much effort do you have to spend to understand the speech?” The response scale was the same as used in previous studies (Rennies *et al.*, 2014a; Schepker *et al.*, 2016; Krueger *et al.*, 2017a; Krueger *et al.*, 2017b) and consisted of 13 listening effort ratings with the seven named categories “no effort” (1 effort scaling categorical unit, ESCU), “very little effort” (3 ESCU), “little effort” (5 ESCU), “moderate effort” (7 ESCU), “considerable effort” (9 ESCU), “very high effort” (11 ESCU), and “extreme effort” (13 ESCU), as well as of six unnamed categories in-

between. In addition, a 14th category named “only noise” was provided for conditions in which subjects could not hear any target speech (and hence an assessment of listening effort would not be meaningful, see Krueger *et al.*, 2017a).

Speech and noise were presented at different fixed SNRs that had been selected in pilot experiments to cover a large range of the response scale for each condition. For the anechoic speech and noise stimuli, the SNRs were -15 , -10 , -5 , 0 , 5 , and 10 dB for the maskers N_0 , N_{180} , and $N_{-45}N_{135}$, and -25 , -20 , -15 , -10 , -5 , 0 , 5 , and 10 dB for the maskers N_{45} , N_{100} , and N_{135} . All 42 combinations of maskers and SNRs were combined in one block with a random order. For each combination a new random target sentence was used, and the next combination was presented immediately after the subjects made their choice. For the reverberant speech and noise stimuli in the high-DRR condition, SNRs ranged from -12 to 9 dB in steps of 3 dB for the N_0 masker, and from -18 to 9 dB (same step size) for the N_{90} masker. Both noise azimuths and all SNRs for the high-DRR condition were combined in one block in a random order. For the low-DRR condition, SNRs ranged from -9 to 9 dB for both noise azimuths, and SNR and noise azimuths were again combined in one block in random order. ILD- and ITD-only conditions were measured as separate blocks, each with SNRs from -25 to $+10$ dB in steps of 5 dB. Each block was repeated four times resulting in four listening effort ratings of each data point from each subject. All blocks were conducted by the subjects in a random order. The median of these four ratings was calculated as the estimated listening effort for each data point and subject (for the present conditions and subjects, the average standard deviation across these four repetitions turned out to be about 1.3 ESCU). The listening effort assessment included no training phase.

D. Data analyses

The first step of the analyses of the speech intelligibility data consisted of fitting a sigmoidal psychometric function (Brand and Kollmeier, 2002) to the obtained SRT_{20} and SRT_{80} data for each subject and condition. The chance performance level of the model function was set to 0% because listeners were not forced to respond. The different conditions were then analyzed by deriving SNRs at a fixed performance level (e.g., 50% recognized words) from the psychometric functions, and comparing these SNRs across conditions.

The same approach was employed for the listening effort data, where model functions were fitted to the median values across the four repetitions of each data point and subject to estimate psychometric functions. Ratings of only noise were excluded from the fitting. The same model function as used by Krueger *et al.* (2017a) and Krueger *et al.* (2017b) was used, which consists of two straight lines that intersect at 7 ESCU (i.e., the mid-point of the listening effort scale). The transition between the two lines is smoothed by a Bézier function. The model function has three degrees of freedom: the slopes of the two lines and the SNR of the intersection point. Despite the relatively broad range of SNRs employed in each condition, the resulting listening effort ratings did not cover a large range of the scale for all subjects and conditions.

Specifically, in some cases no data points at or beyond 10 ESCU in the upper part of the scale, or at or below 4 ESCU in the lower part of the scale were available. In these cases, a first order polynomial was used as a simpler model function. Furthermore, the influence of floor effects on the fitting procedure was limited by investigating the ratings at the highest SNRs in each condition. Some subjects tended to rate listening effort consistently lower than others such that several of the highest SNRs in a given condition were rated as no effort (1 ESCU). In these cases, only the lowest SNR for which this rating was obtained was kept for fitting the model function. Note that this is equivalent to the data collection procedure in the adaptive version of the categorical listening effort scaling (Krueger *et al.*, 2017a), where SNRs at which the subject rated no effort are not repeated or exceeded on an individual basis. For the present data, this procedure of fitting psychometric functions resulted in reasonable fitting accuracy as expressed by the rms error between data and model functions. The average rms error across all 120 combinations of subject and condition was 0.9 ESCU (standard deviation 0.4 ESCU). The largest rms error was 2.5 ESCU for a single combination, and below 1.5 ESCU for more than 95% of the combinations (115 out of 120). Furthermore, the rms error did not differ for different parts of the model functions [estimated by considering low (1–5 ESCU), medium (5–9 ESCU), and high (9–13 ESCU) listening effort ratings separately]. As for the speech intelligibility data, SNRs at a fixed listening effort rating (e.g., “medium effort”) were derived from the psychometric functions and then compared across conditions.

For both speech intelligibility and listening effort, the distributions of individual data for each data point included in the analyses were tested for normal distribution. If normality could be assumed, repeated-measures analyses of variance (ANOVAs) were conducted with a significance level of 0.05. Degrees of freedom were Greenhouse-Geisser corrected. Paired comparisons were conducted as *post hoc* tests if appropriate, and Bonferroni corrections were applied for multiple comparisons. If normality could not be assumed, Friedman tests followed by Wilcoxon rank sum comparisons were conducted.

III. RESULTS

A. Anechoic conditions

Speech intelligibility data for anechoic conditions are shown in the top panel of Fig. 1. Symbols represent interindividual mean values of the SRT_{20} and SRT_{80} , errors indicate standard deviations. Lines represent psychometric functions obtained as parametric averages across individual functions. Different gray scales and line styles/symbols show the different masking conditions. Speech intelligibility was worse in symmetric masking conditions (dashed lines), i.e., higher SNRs were required to obtain a given speech intelligibility performance. Intelligibility functions for S_0N_0 and $S_0N_{-45}N_{135}$ were virtually identical, while intelligibility was slightly better for a masker from behind (corresponding to an SNR shift of about 2.5 dB). For all conditions with a single lateral noise source, a considerable binaural advantage was observed. The shift of the psychometric function compared to a frontal masker was largest for N_{100} (about 12.5 dB) and

FIG. 1. Speech intelligibility (top) and listening effort (bottom) as a function of SNR in anechoic conditions. Different gray scales and symbols/line styles represent the different masking conditions.

somewhat smaller for N_{45} and N_{135} (about 10 dB). These observations were supported by four separate Friedman tests using SNR required to achieve 20%, 50%, 80%, and 95% speech intelligibility, respectively, as independent variable, and noise azimuth (six levels) as within-subject factor. These values were selected to represent low, medium, and high intelligibility as well as ceiling performance (95%). At each intelligibility value the factor noise azimuth was significant at the corrected significance level of 0.05/4. *Post hoc* Wilcoxon comparisons of N_0 with all other noise azimuths showed that this was always due to significant differences to N_{45} , N_{100} , and N_{135} at a corrected significance level of 0.05/5. N_0 also differed significantly from N_{180} for 20%, 50%, and 80% speech intelligibility, but not for 95%.¹

The corresponding listening effort data are shown in the bottom panel of Fig. 1. Symbols and error bars indicate interindividual median values and interquartile ranges. Lines represent psychometric functions obtained by the fitting model function to the median data across subjects. For all masking

conditions, listening effort decreased with increasing SNR. There was no difference in listening effort in the conditions with (rather) symmetric maskers at the two ears (dashed lines). In these conditions, listening effort varied from extreme effort (13 ESCU) to no effort (1 ESCU) across the SNR range from about -15 to $+14$ dB. In conditions with a lateral masker, listening effort was considerably lower at low SNRs. At the highest effort rating of 13 ESCU, the difference in SNR compared to symmetric masking conditions was about 10 dB. This spatial advantage decreased with increasing SNR, and was absent at very favorable SNRs.

This spatial advantage is further illustrated in Fig. 2. Solid lines in different gray scales show SNRs corresponding to fixed categorical listening effort values from 13 ESCU (bottom) to 1 ESCU (top). This representation resulted in a bowl-shaped pattern, which was most pronounced at high effort values, and became increasingly flatter for decreasing listening effort. To address the question if spatial unmasking for listening effort was significant, separate Friedman tests were conducted for each ESCU from 13 to 1 with independent variable SNR and within-subject factor noise azimuth. The significance level was corrected to 0.05/13 for multiple

tests. At this corrected significance level, the main effect of noise azimuth was significant for all listening effort categories except 1 ESCU. Wilcoxon comparisons were employed for effort categories between 2 and 13 ESCU to test if the listening effort in the reference condition (S_0N_0) differed from the other noise azimuths (five comparisons per effort category, significance level set to 0.05/5). This *post hoc* analysis indicated that the binaural advantage was significant for the N_{100} and N_{135} masker for all effort categories from 13 to 2 ESCU, while the difference between N_0 and N_{45} was not significant for listening effort below 5 ESCU. There was no difference between the frontal masker and the N_{180} or $N_{-45}N_{135}$ masker for any effort category.

The representation in Fig. 2 can also be used to compare listening effort and speech intelligibility in binaural anechoic listening conditions. Symbols connected by dashed lines indicate constant speech intelligibility of 50% (asterisks) and 95% (crosses) derived from the average psychometric function (Fig. 1, top). These iso-intelligibility curves followed a similar bowl-shaped pattern as the listening effort data. The comparison to the listening effort data showed that the SRT_{50} corresponded to about 9 ESCU (considerable effort) for all masking conditions. The difference between the two curves did not exceed 2 dB. Similarly, speech intelligibility reached ceiling performance (95%) at SNRs corresponding to about 7 ESCU (moderate effort) for all masking conditions (differences ≤ 3 dB).

B. ILD- and ITD-only conditions

The data for the N_{100} condition with ILD- and ITD-cues only are shown in Fig. 3 as gray lines and symbols. For comparison psychometric functions obtained with full HRIRs in the S_0N_0 - and S_0N_{100} -condition are replotted in black. Both intelligibility (top) and listening effort data (bottom) were consistent in that data obtained with ITD cues only were very similar to data obtained with the full HRIRs in the S_0N_0 condition. In other words, maintaining ITDs, but not ILDs of the N_{100} masker, did not produce a binaural advantage. In contrast, data obtained with the $N_{100,ILD\text{-}only}$ HRIRs were very similar to data for the full HRIR in the S_0N_{100} condition, indicating that a considerable binaural advantage was obtained by the ILD cues alone. This advantage was about 2 dB lower than for the full HRIR.

C. Reverberant conditions

The data measured in reverberant conditions are shown in Fig. 4. As for anechoic conditions, speech intelligibility (top panel) always increased with increasing SNR. Reverberation had an influence on speech intelligibility: data for the low-DRR condition (gray lines and symbols) were considerably shifted to the right compared to low-DRR data (black lines) indicating that a higher SNR was required to obtain the same intelligibility performance. For the N_0 masker this difference was about 5 dB at the SRT_{50} and somewhat smaller (larger) below (above) this point due to the shallower slope in the low-DRR condition. Furthermore, no difference was observed between the S_0N_{90} - and S_0N_0 -condition in a highly reverberant condition. In contrast, the mean SRT_{20} and SRT_{80} were about

FIG. 2. Iso-listening-effort contours in anechoic binaural listening conditions extracted from the fitted psychometric functions. Solid lines in different gray scales represent different effort categories between 1 and 13 ESCU as indicated at the corresponding curves. Dashed lines and symbols show constant speech intelligibility of 50% (asterisks) and 95% (crosses). The right panel indicates the corresponding data in the condition with two maskers. The gray dotted line (left panel) and cross (right panel) at the bottom indicate the better-ear SNR-advantage relative to the S_0N_0 reference condition.

FIG. 3. Speech intelligibility (top) and listening effort (bottom) as a function of SNR for conditions with ILD-cues (dark gray line) and ITD-cues only (light gray line). Data for full HRIRs are replotted from Fig. 1 (black line).

5 and 4 dB lower for separated than for co-located maskers in the high-DRR condition (black lines and symbols), indicating that a considerable binaural advantage was present only for the high-DRR condition. These observations were supported by two-way ANOVAs for repeated measures conducted separately for intelligibility values of 20%, 50%, 80%, and 95%. The factors were noise azimuth (0° and 90°) and receiver position (R1: high DRR, R4: low DRR), and the SNR required to obtain the respective performance was the independent variable. At 95% intelligibility the main effect noise azimuth did not reach significance at the corrected significance level of 0.05/4 (see top part of Table I). All other main effects and the interaction between the two factors were significant at all intelligibility levels. For each ANOVA four paired comparisons were conducted as *post hoc* tests to test if spatial separation produced significant differences for a given receiver position and if the receiver position produced significant differences for a given spatial separation. These tests showed that spatial separation did not significantly modify the required SNR for

FIG. 4. Speech intelligibility (top) and listening effort (bottom) as a function of SNR in reverberant conditions. Black and gray curves and symbols represent the high-DRR [receiver (R) 1] and low-DRR condition (R4), respectively.

any of the four intelligibility values in the low-DRR condition ($p \geq 0.209$). The remaining three comparisons showed significant differences for each intelligibility value ($p \leq 0.002$), i.e., there was significant spatial release from masking in the high-DRR condition, and a significantly better speech intelligibility in the high-DRR condition for both the N_0 and N_{100} masker.

The corresponding listening effort data are shown in the bottom panel of Fig. 4. As for anechoic conditions listening effort always decreased with increasing SNR. In analogy to the speech intelligibility data, there was no apparent difference between the N_0 and N_{100} masker in the low-DRR condition. In contrast, spatial separation reduced listening effort in the high-DRR condition. This spatial release from listening effort depended on SNR: At the upper end of the listening effort scale (13 ESCU) the difference between spatially separated and co-located noise was about 6 dB. This difference decreased with increasing SNR (about 5 dB at 7 ESCU, about 3.5 dB at 5 ESCU, and about 2 dB at 3 ESCU) and was absent for 1 ESCU. To statistically explore these observations

TABLE I. ANOVA summaries of speech intelligibility data (top part) and listening effort data (LE, bottom part) in reverberant conditions. The factors were receiver position (pos.) and noise azimuth (azim.). The corrected significance level for multiple testing was $0.05/4 = 0.0125$ for intelligibility, and $0.05/3 = 0.017$ for effort. p -values significant at these levels are printed in bold face.

	Factor	Deg. of freedom	F	p value
Speech intelligibility 95%	pos.	1, 9	401.024	< 0.001
	azim.	1, 9	8.631	0.017
	rec. * azim.	1, 9	26.040	0.001
Speech intelligibility 80%	pos.	1, 9	454.409	< 0.001
	azim.	1, 9	25.089	0.001
	pos. * azim.	1, 9	65.886	< 0.001
Speech intelligibility 50%	pos.	1, 9	545.382	< 0.001
	azim.	1, 9	40.469	< 0.001
	pos. * azim.	1, 9	82.326	< 0.001
Speech intelligibility 20%	pos.	1, 9	122.642	< 0.001
	azim.	1, 9	21.096	0.001
	pos. * azim.	1, 9	61.058	< 0.001
LE 13 ESCU	pos.	1, 9	7.248	0.025
	azim.	1, 9	0.258	0.624
	pos. * azim.	1, 9	7.050	0.026
LE 7 ESCU	pos.	1, 9	63.061	< 0.001
	azim.	1, 9	18.924	0.002
	pos. * azim.	1, 9	39.658	< 0.001
LE 1 ESCU	pos.	1, 9	14.774	0.004
	azim.	1, 9	4.475	0.063
	pos. * azim.	1, 9	3.501	0.094

separate two-way ANOVAs were run for three different listening effort categories (1, 7, and 13 ESCU) with SNR as the independent variable. As above the factors were noise azimuth and receiver position. The analyses are summarized in the bottom part of Table I. At medium effort (7 ESCU), the main effects of receiver position and noise azimuth as well as the interaction were significant ($p \leq 0.002$) at the corrected significance level of $0.05/3$, indicating that DRR affected listening effort, and that the spatial advantage depended on receiver position. The same four *post hoc* paired comparisons as for the speech intelligibility data were conducted indicating that spatial separation had a significant effect in the high-DRR condition, and that listening effort was significantly increased by increased reverberation for both noise azimuths (all $p \leq 0.002$). As for speech intelligibility, spatial separation did not affect listening effort in the low-DRR condition. At 13 ESCU (extreme effort) the main effect of noise azimuth was not significant. The observed trends regarding the main effect of receiver position ($p = 0.025$) and the interaction between noise azimuth and receiver position ($p = 0.026$) failed to reach significance due to the correction of the significance level. At the listening effort category no effort (1 ESCU), the main effect of receiver position was significant ($p = 0.004$), while the main effect of noise azimuth and the interaction were not.

IV. DISCUSSION

A. Comparison to previous data

Some of the conditions measured in the present study were a repetition of previous studies employing the same or similar acoustic conditions and matrix sentences as target

speech. One main difference to the present experiments is that the previous studies employed the German matrix test (Wagener *et al.*, 1999) and native German listeners rather than the American English matrix test. The German and American English matrix tests share many aspects regarding, e.g., their design criteria and the fact that the long-term spectrum of the masking noises matches that of the talker. However, the tests differ in their reference SRT_{50} for normal-hearing listeners, which is about 3 dB lower for the American English test (-10.0 dB SNR) than for the German test (-7.1 dB SNR, see Kollmeier *et al.*, 2015). Thus, one can expect generally lower SNRs when comparing the present data at a fixed performance level to previous studies.

Despite this difference many effects observed in the present study for anechoic speech were very similar to those reported in previous studies. For example, Beutelmann and Brand (2006) and Warzybok *et al.* (2013) measured binaural speech intelligibility using the same setup of a frontal target and co-located or spatially separated noise maskers. Beutelmann and Brand (2006) found a maximum spatial release from masking of about 11 dB for a noise azimuth of 100° , and a slightly reduced release for 45° (about 9 dB). Similarly, Warzybok *et al.* (2013) reported a masking release of about 8 dB for a noise azimuth of 135° . These values are close to those found in the present study (12.5, 10, and 9 dB, for separations of 100° , 45° , and 135° , respectively). Similar spatial release from masking for lateral maskers also was reported in other studies (e.g., Platte and vom Hövel, 1980; Plomp and Mimpen, 1981; Bronkhorst and Plomp, 1988). For a masker presented from behind the listener, both the present data and the data of Beutelmann and Brand (2006) indicated a small advantage for an N_{180} masker compared to a frontal masker. Although the effect (1 dB) reported by Beutelmann and Brand (2006) was not significant, the effect found here (about 2 dB) was. Peissig and Kollmeier (1997) investigated conditions with one fixed lateral noise source and a second noise source presented from a variety of azimuths. Their general observation was that the spatial release from masking diminished to less than 2 dB when the second noise source was located in the hemisphere opposite the first noise source. The same effect was observed in the present study for the $N_{-45}N_{135}$ masker, confirming the finding that there is little or no spatial release from masking for stationary noise sources arranged on both sides of the target talker.

The reverberant HRIRs of the present study were taken from Rennies *et al.* (2011a). The present data again were in close agreement with that previous study in that reverberation increased the SRT_{50} for a co-located masker (about 5 dB in the present study vs about 4 dB in the previous study) and that binaural release from masking was only observed for the high-DRR condition (about 4.5 dB vs about 6 dB). In summary, the speech intelligibility data obtained in the present study followed the expected trends and broadly confirmed previous findings reported in the literature.

The listening effort results for a frontal (co-located) masker may be compared to the findings from previous studies that used diotic stimuli. Rennies *et al.* (2014a) reported listening effort ratings that decreased from 13 to 4 ESCU over the range from about -10 to $+6$ dB SNR. The results of

the present study for the S_0N_0 condition indicated that 13 and 4 ESCU corresponded to SNRs of about -15 and $+2$ dB, respectively, i.e., the psychometric function was shifted leftwards along the SNR abscissa by 4–5 dB compared to the psychometric function fit to the data of Rennie *et al.* (2014a), while the slope of the listening effort curve was similar. Comparing listening effort ratings at a fixed SNR, the difference corresponded to about 3 to 4 ESCU. The shift along the SNR scale is similar in magnitude to the 3 dB difference in reference SRT_{50} between the German and American English matrix test, suggesting that this general intelligibility difference roughly translates to listening effort ratings. Another possible reason for the shift is the different group subjects, which might generally have assigned lower listening effort ratings in the present study irrespective of the particular test. However, Rennie *et al.* (2014a) reported that their data were within about 1 ESCU of the results from other studies that had used the same German speech material (Schulte *et al.*, 2008; Luts *et al.*, 2010; Schepker *et al.*, 2011). This difference is considerably smaller than the observed difference of 3–4 ESCU, which supports the conclusion that the difficulty in speech intelligibility for a given set of speech materials affects listening effort ratings even at higher SNRs where speech intelligibility is at ceiling.

B. Binaural release from listening effort

In contrast to previous studies reviewed in Sec. I (e.g., Zekveld *et al.*, 2014; Xia *et al.*, 2015; Andol *et al.*, 2017), the present study also investigated binaural conditions in which speech intelligibility was at ceiling, which may arguably be the SNR region where the evaluating listening effort is most valuable. The data showed a significant spatial release from listening effort across a wide SNR range, i.e., perceived listening effort continues to vary with SNR even after intelligibility has reached asymptote. In other words, the same SNR led to a significantly lower listening effort when target speech and noise masker were spatially separated compared to co-located/symmetric conditions. This effect extended well into the SNR domain in which speech intelligibility was at ceiling (cf. Fig. 3). Depending on noise azimuth, speech intelligibility approached ceiling (95%) at SNRs between about -6 dB (N_0 and $N_{-45}N_{135}$) and -16 dB (N_{100} , N_{135}). At higher SNRs, differences in listening effort between noise azimuths were significant up to 2 ESCU, which corresponded to SNRs between about $+8$ dB (N_0) and $+5$ dB (N_{100}). In other words, spatial separation of target and masker provided a reduction of listening effort at SNRs more than 14 dB above the SNRs where intelligibility was estimated to reach ceiling. This indicated that this observation does not critically depend on the exact estimate of ceiling performance in speech intelligibility. At even higher SNRs listening effort did not differ significantly across noise azimuths due to floor effects of the listening effort scale. This seems intuitive because no binaural advantage can be expected in conditions of no effort already in the co-located case. Above these floor effects, however, evaluating listening effort considerably extends the range of SNRs in which spatial advantage can be measured, which may be a valuable

tool for future studies measuring speech perception in realistic (i.e., binaural) acoustic scenes with not too unfavorable SNRs (which often dominate everyday listening). It remains to be investigated if binaural hearing also provides an advantage in listening effort at high SNRs in conditions with no long-term better-ear (such as with symmetrically arranged interfering talkers as used in the other studies). It should be noted that the SNRs corresponding to specific listening effort ratings were extrapolated from the psychometric functions and should thus be considered with caution. However, the general finding that listening effort decreased significantly with spatial separation even when intelligibility was at ceiling could also be derived from the raw data, i.e., without combining different data points to fit a psychometric function. For example, in the anechoic room the highest SNR at which meaningful listening effort ratings were measured for all ten subjects (i.e., before floor effects occurred for individual subjects as discussed in Sec. II D) was -5 dB. At this SNR, speech intelligibility was at ceiling for all spatial conditions (see Fig. 1), but listening effort ratings were significantly lower in the S_0N_{100} -condition than in the S_0N_0 condition according to a Wilcoxon signed rank test ($Z = -2.521$, $p = 0.012$). Nevertheless, this important finding should be further investigated and validated in future studies (see also Sec. IV E).

The SNR region in which intelligibility is below ceiling is also worth looking at in more detail. The present data showed that speech intelligibility and listening effort were highly correlated in sub-ceiling intelligibility regions for all conditions, i.e., a fixed intelligibility performance was associated with a fixed listening effort rating irrespective of the binaural configuration or the degree of reverberation (Fig. 3). This is in line with conclusions of Zekveld *et al.* (2014), who measured listening effort at SNRs adaptively converging to the fixed performance level of 50%. They found that spatial masker separation increased intelligibility (in that SRTs were lower), but did not reduce listening effort (as measured in terms of pupil dilation). This was in contrast to a change in gender of the interfering talkers, which both improved intelligibility and reduced listening effort. They argued that spatial separation “did not affect cognitive processing load during listening” and that one possible interpretation could be “that spatial separation eases speech understanding at a more peripheral level of processing, perhaps subcortical, whereas voice cues have to be dealt with at the cortical level by using top-down processing.” The data of the present study support this notion, in that conditions with spatially separated maskers were rated with the same or a similar categorical effort as conditions with co-located maskers when compared at the same level of speech intelligibility. In other words, spatial separation *per se* did not reduce listening effort at a controlled level of intelligibility.

C. Role of interaural time and level differences

The conditions with ILD- and ITD-only cues of the HRIR corresponding to the N_{100} masker showed that a large portion of the observed speech intelligibility improvement and listening effort reduction could be explained by ILDs,

while ITDs alone did not provide a significant advantage. While it may be expected for the stationary noise masker employed here that better-ear listening plays a major role, the present data are not fully in line with data of previous studies. For example, Bronkhorst and Plomp (1988) measured SRT_{50} for frontal speech and speech-shaped noise maskers from different azimuths, and with both ITD- and ILD-cues, or with only one of the cues. For a noise azimuth of 90° they found a spatial advantage of about 10 dB when all cues were provided. This advantage reduced by about 2 dB when only ILD cues were provided, which is very similar to the present data for a noise azimuth of 100° (see Fig. 3). However, Bronkhorst and Plomp (1988) also found that providing ITD-cues only still provided a considerable advantage over the co-located condition (about 5 dB). Similarly, Culling *et al.* (2004) found that both ILD and ITD cues in isolation produced about equally large spatial unmasking (3–4 dB) for speech-shaped noise maskers from 90° and a frontal target compared to a co-located masking condition. In contrast, Rennie *et al.* (2011b) reported SRTs that were dominated by better-ear listening in that monaurally presented speech (best ear only), diotically presented speech (best ear copied to the other side), and binaurally presented speech (including ILD and ITD cues) produced the same SRTs for different degrees of reverberation and a fixed spatial separated setup of target and noise masker. In the present N_{100} data, the possible advantage due to isolated ITD cues was a trend at best (about 1 dB) and was not significant, which is in line with the data of Rennie *et al.* (2011b), but not with the data of Bronkhorst and Plomp (1988) and Culling *et al.* (2004). The reason for the lack of a larger advantage due to ITD-cues alone in the present study is unclear. It is worth noting though that this observation was consistent for both speech intelligibility and listening effort. To further explore the role of better-ear listening in the present study, the anechoic HRIRs were evaluated with respect to their head-shadow effects. To this end, the noise maskers for all azimuths were set to a fixed digital level at the right ear (as used for calibrating the experimental setup) and filtered in third-octaves from 160 Hz to 8 kHz. The resulting band levels were subtracted from the band levels of the N_0 masker (which had the same spectrum as the target speech). These band level differences were then weighted with band-importance functions for speech in noise from the Speech Intelligibility Index (ANSI, 1997) and summed to obtain a speech-weighted SNR for each noise azimuth. These SNRs can be interpreted as speech-weighted SNR advantages due to better-ear listening when the noise is combined with a frontal target. The dotted line at the bottom of Fig. 2 shows the resulting speech-weighted SNR advantage as a function of noise azimuth, which was aligned vertically to coincide with the data point for 13 ESCU in the N_0 condition. The corresponding value point for the $N_{-45}N_{135}$ masker is indicated as a cross in the right panel of Fig. 2. It is evident that the better-ear SNR advantage and the iso-listening effort data for 13 ESCU essentially overlap (differences < 1 dB), i.e., the spatial unmasking in difficult listening conditions can be fully explained by better-ear listening. Note that the amount of spatial unmasking as observed by listening effort

(i.e., the depth of the “bowls” in Fig. 2) is similar for high to medium effort ratings. It is also very similar to the spatial unmasking observed for speech intelligibility (see dashed lines in Fig. 2), i.e., in the SNR region in which speech intelligibility is below ceiling and in which listening effort is between about 7 and 13 ESCU, better-ear effects can quantitatively account for the spatial unmasking observed in the present study. This supports the observation that ILDs played a major role, possibly leaving little additional room for further improvements due to ITD processing. It should be noted that better-ear listening cannot explain the reduction of spatial unmasking at lower listening effort ratings (i.e., the flattening of the curves), because the better-ear effects are independent of overall level/SNR. At high SNRs, better-ear effects seem to play a smaller role and eventually have no more impact on perceived listening effort.

D. Relation between speech intelligibility and listening effort

For anechoic conditions iso-intelligibility and iso-effort curves showed a very similar dependence on noise azimuth (Fig. 2) as discussed above. To explore the relation between effort and intelligibility in more detail, all group psychometric functions were sampled between -28 and 15 dB SNR in steps of 1 dB, i.e., speech intelligibility and listening effort values were extracted for the same SNRs and conditions over the entire range from no intelligibility/maximum effort to high intelligibility/minimum effort. Figure 5 shows the resulting speech intelligibility values plotted against listening effort. Each symbol represents one combination of intelligibility and effort at the same SNR. Different symbols and gray scales represent different measurement conditions. This representation illustrates that both quantities are clearly related when they are both below ceiling, i.e., lower intelligibility corresponds to a higher listening effort in the range of listening effort ratings between about 6 and 13 ESCU. At lower effort ratings (i.e., toward higher SNR), intelligibility is

FIG. 5. Speech intelligibility data plotted against listening effort data. Lines represent a sigmoidal fit to the data (solid line) and the same fit shifted horizontally by ± 1 ESCU.

at ceiling while effort continues to decrease. This relation is common to all conditions in this study as illustrated by lines in Fig. 5. These lines represent a sigmoid function fitted to the scatter plot (solid line) and the fit shifted horizontally by plus and minus 1 ESCU (dashed line). Most data points are within the area of 1 ESCU around the fitted function, indicating that listening effort can be reasonably well predicted from sub-ceiling speech intelligibility. This seems to be independent from the available spatial cues or the degree of reverberation, suggesting that the relation between speech intelligibility and categorical listening effort can be generalized over a broad range of factors affecting speech perception.

The fact that listening effort is sensitive to changes in spatial separation at higher SNRs, while speech intelligibility is not, may be relevant for evaluating the benefit of hearing devices. For example, adaptive noise suppression algorithms or beamformers operate differently at different SNRs and levels of reverberation (e.g., Doire *et al.*, 2017) and some previous studies reported that an algorithm benefit was not found in terms of SRT, but in terms of (monaural) listening effort (e.g., Luts *et al.*, 2010; Marzinzik and Kollmeier, 2002). Similar results have been found for the benefit of hearing aids (e.g., Hällgren *et al.*, 2005), where dynamic range compressors behave differently depending on whether the target or the masker controls the applied gain (see also Smeds *et al.*, 2015). In this light, (binaural) listening effort may be a relevant measure for evaluating the benefit of signal enhancement algorithms and hearing devices at favorable SNRs.

E. Limitations and outlook

The present study employed one particular method for assessing listening effort—categorical scaling as used in several previous studies (e.g., Krueger *et al.*, 2017a; Krueger *et al.*, 2017b; Luts *et al.*, 2010; Rennie *et al.*, 2014a; Schepker *et al.*, 2016)—and focused on one specific set of stimuli and conditions. Although this approach revealed some novel insights into binaural speech perception, the most complete understanding of this topic would be based on additional work providing converging evidence for the present findings and extending them to other types of stimuli and listening situations. For example, a variety of other methods typically used to study listening effort could be adapted to investigate the role of spatial separation of sound sources in a manner that parallels the work described in this study. These efforts may include neurophysiological measurements such as pupilometry and EEG, which have been shown to provide useful and relatively objective measures (i.e., not depending on the subjective judgment of the listener) of cognitive load in various types of listening tasks. A comparison between subjective and objective methods seems particularly relevant because some studies have found that the different approaches do not necessarily produce congruent results (e.g., Zekveld *et al.*, 2011). Furthermore, (re)examining the present (or similar) conditions may be necessary/beneficial for achieving a comprehensive understanding of all aspects relevant to characterizing listening effort. For example, Pichora-Fuller *et al.* (2016) included the

listener's motivation to engage in the listening task as a critical component in their proposed framework for understanding effortful listening. Gauging such factors—which clearly are relevant in realistic listening situations—is difficult to accomplish by relying solely on the type of subjective methods employed in this study (or, presumably, by any study design relying on artificial and not particularly rewarding acoustic settings in laboratory experiments).

Future work should also resolve any potential inconsistencies between the findings from different measurement approaches resulting from the specific and sometimes idiosyncratic terminology associated with specific approaches (e.g., for the scaling procedure terms include listening effort, listening difficulty, ease of listening, etc.). These differences in terminology were hypothesized to have contributed to the inconsistencies and incompatible conclusions sometimes found between different studies investigating the impact of hearing impairment and hearing-loss compensation on listening effort (Ohlenforst *et al.*, 2017a). Although the present study employed a repeated-measures design, which tends to minimize the effects of large differences in performance between subjects that inevitably occur when comparing across studies employing different groups of subjects (e.g., as in Ohlenforst *et al.*, 2017a), it would be interesting to compare our findings with those from other subjective assessment methods. One such approach would be to employ multiple-alternative forced-choice procedures to match the listening effort of test conditions—perhaps varying the comparison stimulus according to an adaptive procedure—to that of a reference condition. One way to implement this could be to adaptively adjust the SNR (which is known to be monotonically related to listening effort over a large range) and directly assess the SNR differences required for equal listening effort (which were derived indirectly from the psychometric functions in this study). Such matching procedures have been applied extensively in the past (e.g., in loudness measurements) and often are considered more sensitive to small changes between conditions than categorical scaling procedures (e.g., Anweiler and Verhey, 2006; Verhey and Heeren, 2015), although some studies found that both methods produced statistically equivalent results (e.g., Hots *et al.*, 2014; Rennie *et al.*, 2013). Since this is the first study to report a spatial benefit in terms of subjective listening effort in conditions with optimal speech intelligibility, future studies using alternative listening effort assessment methods are required to validate this finding.

In addition, the role of spatial hearing in listening effort should be evaluated over a broader range and, to the extent possible, more ecologically relevant listening conditions. For example, speech-on-speech masking should be considered, especially because it has been shown to induce large binaural unmasking effects even in conditions where acoustic better-ear effects are minimized by the symmetrical placement of multiple maskers (e.g., Marrone *et al.*, 2008). Extending the current work to include speech-on-speech listening conditions also would allow investigating the interaction of binaural processing and other unmasking cues (such as gender differences or masker intelligibility), which are not applicable in speech-in-noise conditions.

V. CONCLUSIONS

- (1) Spatial separation of a single noise source from a frontal target is beneficial for reducing listening effort over a large range of SNRs. Psychophysical scaling of perceived listening effort is a suitable tool for measuring these effects, revealing spatial release of listening effort well into the range of positive SNRs, at which speech intelligibility is at ceiling.
- (2) This spatial release from listening effort disappears in strongly reverberant conditions, similar to what is known from binaural speech intelligibility studies.
- (3) At a fixed performance level of speech intelligibility (below ceiling), speech intelligibility and listening effort are strongly correlated for all spatial and reverberation conditions, indicating that spatial separation *per se* does not reduce listening effort at a controlled level of intelligibility.
- (4) The spatial advantage measured in the conditions of this study seemed to be dominated by better-ear listening, while ITDs played a minor role.

ACKNOWLEDGMENTS

This study was supported by the German Research Foundation (DFG) Grant No. RE 4160/1-1, and the National Institutes of Health–National Institute on Deafness and other Communication Disorders Grant No. R01 DC04545.

¹Note that the intelligibility value of 95% was outside the range of values that were actually measured in the experiment, i.e., it was extrapolated from the psychometric functions. Conclusions based on this value should therefore be considered with caution.

Andéol, G., Suied, C., Scanella, S., and Dehais, F. (2017). “The spatial release of cognitive load in cocktail party is determined by the relative levels of the talkers,” *J. Assoc. Res. Otolaryngol.* **18**, 457–464.

#ANSI (1997). American National Standards S3.5-1997. “Methods for the calculation of the Speech Intelligibility Index” (Acoustical Society of America, New York).

Anweiler, A. K., and Verhey, J. L. (2006). “Spectral loudness summation for short and long signals as a function of level,” *J. Acoust. Soc. Am.* **119**, 2919–2928.

Arweiler, I., and Buchholz, J. M. (2011). “The influence of spectral characteristics of early reflections on speech intelligibility,” *J. Acoust. Soc. Am.* **130**, 996–1005.

Beutelmann, R., and Brand, T. (2006). “Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners,” *J. Acoust. Soc. Am.* **120**, 331–342.

Beutelmann, R., Brand, T., and Kollmeier, B. (2010). “Revision, extension, and evaluation of a binaural speech intelligibility model,” *J. Acoust. Soc. Am.* **127**, 2479–2497.

Brand, T., and Kollmeier, B. (2002). “Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests,” *J. Acoust. Soc. Am.* **111**, 2801–2810.

Bronkhorst, A. W. (2000). “The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions,” *Acustica* **86**, 117–128.

Bronkhorst, A. W. (2015). “The cocktail-party problem revisited: Early processing and selection of multi-talker speech,” *Atten. Percept. Psychophys.* **77**, 1465–1487.

Bronkhorst, A. W., and Plomp, R. (1988). “The effect of head-induced interaural time and level differences on speech intelligibility in noise,” *J. Acoust. Soc. Am.* **83**, 1508–1516.

Carlile, S. (2014). “Active listening: Speech intelligibility in noisy environments,” *Acoustics Australia* **42**, 98–104.

Chabot-Leclerc, A., MacDonald, E. N., and Dau, T. (2016). “Predicting binaural speech intelligibility using the signal-to-noise ratio in the envelope power spectrum domain,” *J. Acoust. Soc. Am.* **140**, 192–205.

Cherry, E. C. (1953). “Some experiments on the recognition of speech, with one and with two ears,” *J. Acoust. Soc. Am.* **25**, 975–979.

Culling, J., Hawley, M. L., and Litovsky, R. Y. (2004). “The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources,” *J. Acoust. Soc. Am.* **116**, 1057–1065.

Devocht, E. M., Janssen, A. M., Chalupper, J., Stokroos, R. J., and George, E. L. (2017). “The benefits of bimodal aiding on extended dimensions of speech perception: Intelligibility, listening effort, and sound quality,” *Trends Hear.* **21**, 1–20.

Doire, C. S. J., Brookes, M., Naylor, P. A., Hicks, C. M., Betts, D., Dmour, M. A., and Jensen, S. H. (2017). “Single-channel online enhancement of speech corrupted by reverberation and noise,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.* **25**, 572–587.

Ellinger, R. L., Jakine, K. M., and Gallun, F. J. (2017). “The role of interaural differences on speech intelligibility in complex multi-talker environments,” *J. Acoust. Soc. Am.* **141**, EL170–EL176.

Ewert, S. D. (2013). “AFC—A modular framework for running psychoacoustic experiments and computational perception models,” in *Proceedings of the Conference on Acoustics, AIA-DAGA 2013*, pp. 1326–1329.

Gosselin, P. A., and Gagné, J. P. (2010). “Use of a dual-task paradigm to measure listening effort,” *Can. J. Speech Lang. Pathol. Audiol.* **34**, 43–51.

Hällgren, M., Larsby, B., Lyxell, B., and Arlinger, S. (2005). “Speech understanding in quiet and noise, with and without hearing aids,” *Int. J. Aud.* **44**, 574–583.

Hauth, C. F., and Brand, T. (2018). “Modeling sluggishness in binaural unmasking of speech for maskers with time-varying interaural phase differences,” *Trends Hear.* **22**, 1–10.

Hicks, C. B., and Tharpe, A. M. (2002). “Listening effort and fatigue in school-age children with and without hearing loss,” *J. Speech Lang. Hear. Res.* **45**, 573–584.

Hots, J., Rennie, J., and Verhey, J. L. (2014). “Loudness of subcritical sounds as a function of bandwidth, center frequency, and level,” *J. Acoust. Soc. Am.* **135**, 1313–1320.

Houben, R., van Doorn-Bierman, M., and Dreschler, W. (2013). “Using response time as a measure of listening effort,” *Int. J. Audiol.* **52**, 753–761.

Johnson, J., Xu, J., Cox, R., and Pendergraft, P. (2015). “A comparison of two methods for measuring listening effort as part of an audiologic test battery,” *Am. J. Audiol.* **24**, 419–431.

Kayser, H., Ewert, S. D., Anemüller, J., Rohdenburg, T., Hohmann, V., and Kollmeier, B. (2009). “Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses,” *EURASIP J. Adv. Signal Process.* **2009**(1), 298605.

Klink, K., Schulte, M., and Meis, M. (2012a). “Measuring listening effort in the field of audiology—A literature review of methods (part 1),” *Z. Audiol.* **51**, 60–67.

Klink, K., Schulte, M., and Meis, M. (2012b). “Measuring listening effort in the field of audiology—A literature review of methods (part 2),” *Z. Audiol.* **51**, 96–105.

Kollmeier, B., Warzybok, A., Hochmuth, S., Zokoll, M. A., Usilar, V., Brand, T., and Wagener, K. C. (2015). “The multilingual matrix test: Principles, applications, and comparison across languages: A review,” *Int. J. Audiol.* **54**, 3–16.

Kramer, S., Kapteyn, T., Festen, J., and Kuik, D. (1997). “Assessing aspects of auditory handicap by means of pupil dilatation,” *Int. J. Audiol.* **36**, 155–164.

Krueger, M., Schulte, M., Brand, T., and Holube, I. (2017a). “Development of an adaptive scaling method for subjective listening effort,” *J. Acoust. Soc. Am.* **141**, 4680–4693.

Krueger, M., Schulte, M., Zokoll, M. A., Wagener, K. C., Meis, M., Brand, T., and Holube, I. (2017b). “Relation between listening effort and speech intelligibility in noise,” *Am. J. Audiol.* **26**, 378–392.

Lavandier, M., and Culling, J. F. (2008). “Speech segregation in rooms: Monaural, binaural, and interacting effects of reverberation on target and interferer,” *J. Acoust. Soc. Am.* **123**, 2237–2248.

Lavandier, M., and Culling, J. F. (2010). “Prediction of binaural speech intelligibility against noise in rooms,” *J. Acoust. Soc. Am.* **127**, 387–399.

- Lavandier, M., Jelfs, S., Culling, J. F., Watkins, A. J., Raimond, A. P., and Makin, S. J. (2012). "Binaural prediction of speech intelligibility in reverberant rooms with multiple noise sources," *J. Acoust. Soc. Am.* **131**, 218–231.
- Leclère, T., Lavandier, M., and Culling, J. F. (2015). "Speech intelligibility prediction in reverberation: Towards an integrated model for speech transmission, spatial unmasking, and binaural de-reverberation," *J. Acoust. Soc. Am.* **137**, 3335–3345.
- Luts, H., Eneman, K., Wouters, J., Schulte, M., Vormann, M., Buechler, M., Dillier, N., Houben, R., Dreschler, W. A., Froehlich, M., Puder, H., Grimm, G., Hohmann, V., Leijon, A., Lombard, A., Mauler, D., and Spriet, A. (2010). "Multicenter evaluation of signal enhancement algorithms for hearing aids," *J. Acoust. Soc. Am.* **127**, 1491–1505.
- Mackersie, C. L., and Cones, H. (2011). "Subjective and psychophysiological indices of listening effort in a competing-talker task," *J. Am. Acad. Audiol.* **22**, 113–122.
- Marrone, N. L., Mason, C. R., and Kidd, G., Jr. (2008). "Tuning in the spatial dimension: Evidence from a masked speech identification task," *J. Acoust. Soc. Am.* **124**, 1146–1158.
- Marzinzik, M., and Kollmeier, B. (2002). "Noise reduction schemes for digital hearing aids: I. Listening effort and speech intelligibility," *Z. Audiol.* **41**, 4–21.
- Mattys, S. L., Davis, M. H., Bradlow, A. R., and Scott, S. K. (2012). "Speech recognition in adverse conditions: A review," *Lang. Cogn. Process.* **27**, 953–978.
- Middlebrooks, J., Simon, J. Z., Popper, A. N., and Fay, R. R. (eds.) (2017). "The auditory system at the cocktail party," in *Springer Handbook of Auditory Research* (Springer International Publishing, Cham, Switzerland).
- Morimoto, M., Sato, H., and Kobayashi, M. (2004). "Listening difficulty as a subjective measure for evaluation of speech transmission performance in public spaces," *J. Acoust. Soc. Am.* **116**, 1607–1613.
- Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., and Maess, B. (2012). "Listening conditions and memory load drive a common alpha oscillatory network," *J. Neurosci.* **32**, 12376–12383.
- Ohlenforst, B., Zekveld, A. A., Jansma, E. P., Wang, Y., Naylor, G., Lorens, A., Lunner, T., and Kramer, S. E. (2017a). "Effects of hearing impairment and hearing aid amplification on listening effort: A systematic review," *Ear Hear.* **38**, 267–281.
- Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., Versfeld, N. J., and Kramer, S. E. (2017b). "Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation," *Hear. Res.* **351**, 68–79.
- Peissig, J., and Kollmeier, B. (1997). "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," *J. Acoust. Soc. Am.* **101**, 1660–1670.
- Pichora-Fuller, K. M., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M., Tremblay, K. L., and Wingfield, A. (2016). "Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL)," *Ear Hear.* **37**, 5S–27S.
- Platte, H.-J., and vom Hövel, H. (1980). "Zur Deutung der Ergebnisse von Sprachverständlichkeitsmessungen mit Störschall im Freifeld" ("On the interpretation of speech intelligibility measurements with noise in free field"), *Acustica* **45**, 139–150.
- Plomp, R., and Mimpen, A. M. (1981). "Effect of the orientation of the speaker's head and the azimuth of a noise source on the speech-reception threshold for sentences," *Acustica* **48**, 325–328.
- Rennies, J., Brand, T., and Kollmeier, B. (2011a). "Prediction of the influence of reverberation on binaural speech intelligibility in noise and in quiet," *J. Acoust. Soc. Am.* **130**, 2999–3012.
- Rennies, J., Brand, T., and Kollmeier, B. (2011b). "Prediction of binaural intelligibility level differences in reverberation," in *Proceedings of the 12th Annual Conference of the International Speech Communication Association—Interspeech*, Florence, Italy, pp. 37–40.
- Rennies, J., Holube, I., and Verhey, J. L. (2013). "Loudness of speech and speech-like signals," *Acta Acust. Acust.* **99**, 268–282.
- Rennies, J., Schepker, H., Holube, I., and Kollmeier, B. (2014a). "Listening effort and speech intelligibility in listening situations affected by noise and reverberation," *J. Acoust. Soc. Am.* **136**, 2642–2653.
- Rennies, J., Warzybok, A., Brand, T., and Kollmeier, B. (2014b). "Modeling the effects of a single reflection on binaural speech intelligibility," *J. Acoust. Soc. Am.* **135**, 1556–1567.
- Sarampalis, A., Kalluri, S., Edwards, B., and Hafter, E. (2009). "Objective measures of listening effort: Effects of background noise and noise reduction," *J. Speech Lang. Hear. Res.* **52**, 1230–1240.
- Sato, H., Morimoto, M., and Wada, M. (2012). "Relationship between listening difficulty rating and objective measures in reverberant and noisy sound fields for young adults and elderly persons," *J. Acoust. Soc. Am.* **131**, 4596–4605.
- Schepker, H., Haeder, K., Rennies, J., and Holube, I. (2016). "Perceived listening effort and speech intelligibility in reverberation and noise for hearing-impaired listeners," *Int. J. Audiol.* **55**, 738–747.
- Schepker, H., Rennies, J., Holube, I., and Kollmeier, B. (2011). "Kontinuierliche Messung und Vorhersage von Höranstrengung" ("Continuous measurement and prediction of listening effort"), in *Proceedings of the 36th Conference of the Deutsche Gesellschaft für Akustik e.V.*, Düsseldorf, Germany, pp. 589–590.
- Schulte, M., Wagener, K., Vormann, M., Dillier, N., and Büchler, M. (2008). "Report tests for listening effort," public deliverable D-7-4 of the HearCom project, URL: <http://hearcom.eu/about/DisseminationandExploitation/deliverables.html> (Last viewed December 5, 2013).
- Smeds, K., Wolters, F., and Rung, M. (2015). "Estimation of signal-to-noise ratios in realistic sound scenarios," *Am. J. Audiol.* **26**, 183–196.
- Steeneken, H. J. M., and Houtgast, T. (1980). "A physical method for measuring speech-transmission quality," *J. Acoust. Soc. Am.* **67**, 318–326.
- van Schoonhoven, J., Schulte, M., Boymans, M., Wagener, K. C., Dreschler, W. A., and Kollmeier, B. (2016). "Selecting appropriate tests to assess the benefits of bilateral amplification with hearing aids," *Trends Hear.* **20**, 1–16.
- Verhey, J. L., and Heeren, W. (2015). "Categorical scaling of partial loudness in a condition of masking release," *J. Acoust. Soc. Am.* **138**, 904–915.
- Wagener, K., Brand, T., and Kollmeier, B. (1999). "Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Evaluation des Oldenburger Satztests" ("Development and evaluation of a German sentence test III: Evaluation of the Oldenburg sentence test"), *Z. Audiol.* **38**, 86–95.
- Warzybok, A., Rennies, J., Brand, T., Doclo, S., and Kollmeier, B. (2013). "Effects of spatial and temporal integration of a single early reflection on speech intelligibility," *J. Acoust. Soc. Am.* **133**, 269–282.
- Wendt, D., Hietkamp, R. K., and Lunner, T. (2017). "Impact of noise and noise reduction on processing effort: A pupillometry study," *Ear Hear.* **38**, 690–700.
- Xia, J., Noorale, N., Kalluri, S., and Edwards, B. (2015). "Spatial release of cognitive load measured in a dual-task paradigm in normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **137**, 1888–1898.
- Yost, W. A. (1997). "The cocktail party problem: Forty years later," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Hillsdale, NJ), pp. 329–347.
- Zekveld, A. A., Kramer, S. E., and Festen, J. M. (2010). "Pupil response as an indication of effortful listening: The influence of sentence intelligibility," *Ear Hear.* **31**, 480–490.
- Zekveld, A. A., Kramer, S. E., and Festen, J. M. (2011). "Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response," *Ear Hear.* **32**, 498–510.
- Zekveld, A. A., Rudner, M., Kramer, S. E., Lyzenga, J., and Rönneberg, J. (2014). "Cognitive processing load during listening is reduced more by decreasing voice similarity than by increasing spatial separation between target and masker speech," *Front. Neurosci.* **8**(88), 1–11.