Spatial selective attention in a complex auditory environment such as polyphonic music

Katja Saupe[a)]
Department of Neurobiology, Institute of Biology II, University of Leipzig, Talstrasse 33, Leipzig D-04103, Germany

Stefan Koelsch[b)]
Junior Research Group Neurocognition of Music, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Stephanstrasse 1a, Leipzig D-04103, Germany

Rudolf Rübsamen
Department of Neurobiology, Institute of Biology II, University of Leipzig, Talstrasse 33, Leipzig D-04103, Germany

(Received 4 June 2009; revised 13 November 2009; accepted 13 November 2009)

To investigate the influence of spatial information in auditory scene analysis, polyphonic music (three parts in different timbres) was composed and presented in free field. Each part contained large falling interval jumps in the melody and the task of subjects was to detect these events in one part ("target part") while ignoring the other parts. All parts were either presented from the same location (0°; *overlap condition*) or from different locations (−28°, 0°, and 28° or −56°, 0°, and 56° in the azimuthal plane), with the target part being presented either at 0° or at one of the right-sided locations. Results showed that spatial separation of 28° was sufficient for a significant improvement in target detection (i.e., in the detection of large interval jumps) compared to the overlap condition, irrespective of the position (frontal or right) of the target part. A larger spatial separation of the parts resulted in further improvements only if the target part was lateralized. These data support the notion of improvement in the suppression of interfering signals with spatial sound source separation. Additionally, the data show that the position of the relevant sound source influences auditory performance. © *2010 Acoustical Society of America.* [DOI: 10.1121/1.3271422]

PACS number(s): 43.75.Cd, 43.66.Pn, 43.66.Jh [DD] Pages: 472–480

I. INTRODUCTION

Our everyday listening environment is often highly complex, with many sounds occurring at the same time. Sitting in an office, for example, you might hear the telephone ringing, people talking, and traffic noise outside. Such a setting was exemplified by Cherry (1953) as a cocktail party situation. Because all the surrounding sound signals arrive at the cochlea as a composite, a preliminary analysis of the incoming sound is required to divide the auditory input into distinct perceptual objects (also referred to as auditory scene analysis; see Bregman, 1990, for review). To select relevant information from concurrent, irrelevant sound streams, spectral, temporal, and spatial cues are analyzed and integrated (for a review of selective attention to auditory objects see Alain and Arnott, 2000). As long as only two sound sources are present, the influence of spatial information on the segregation of auditory objects is often relatively small if other stimulus parameters are available instead (Butler, 1979; Deutsch, 1975; Shackleton *et al.*, 1994; Yost *et al.*, 1996). However, the benefit from spatial information increases significantly with three simultaneously active sound sources (Eramudugolla *et al.*, 2008; Hawley *et al.*, 2004; Yost *et al.*, 1996). Previous studies investigating spatial auditory attention with more than two sound sources often presented the competing stimuli successively rather than simultaneously (e.g., Münte *et al.*, 2001; Nager *et al.*, 2003; Teder-Sälejärvi *et al.*, 1999). Treisman (1964) was one of the first to present up to three sources simultaneously for the investigation of selective filtering in auditory attention. While retaining a fixed attended sound source, the number (0–2) and simulated spatial location of irrelevant sound sources, as well as the simulated distance between sound sources, were varied. However, because of the dichotic presentation of sound stimuli through earphones instead of free field stimulation, the acoustical percept was somewhat unnatural with sound sources being either located directly at the two ears or along an intracranial axis between the two ears. Yost *et al.* (1996) created a natural listening condition by using up to three simultaneously active sound sources in free field and used spoken words, letters, and numbers as acoustic stimuli. Divided attention to all simultaneously active stimuli was tested for different positions and different spatial separations of the sound sources. Several follow-up studies investigated the masking influence of task-irrelevant acoustic stimuli on the speech reception threshold in a multi-source environment (Culling *et al.*, 2004; Hawley *et al.*, 1999, 2004; Kidd *et al.*, 2005; Peissig and Kollmeier, 1997).

[a)]Author to whom correspondence should be addressed. Present address: Institute of Psychology I, University of Leipzig, Seeburgstrasse 14-20, Leipzig D-04103, Germany. Electronic mail: saupe@rz.uni-leipzig.de
[b)]Present address: Department of Psychology, Pevensey Building, University of Sussex, Falmer, Brighton, BN1 9QH, UK.

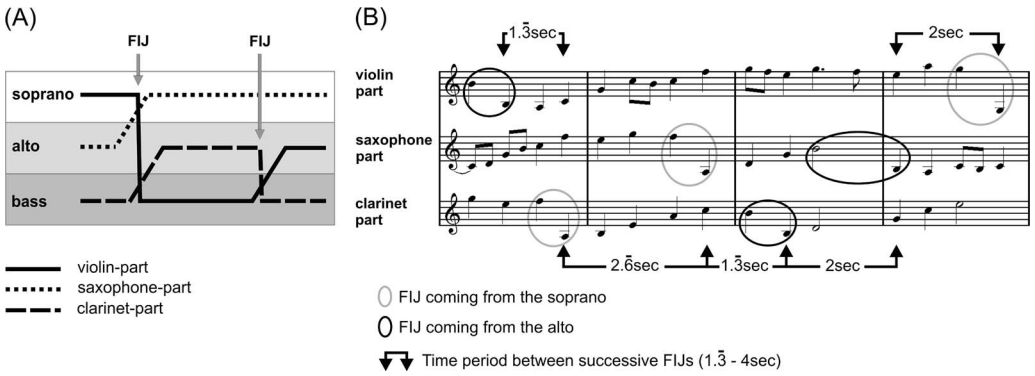

FIG. 1. (A) Example of the melody course of the three parts. The violin part (target part) is initially in the soprano register, saxophone part in the alto, and the clarinet part in the bass register. A FIJ in the melodic contour of at least 1 octave in the violin part (FIJ in the soprano) brought that part into the bass register; isochronal, the saxophone part turns into the soprano and the clarinet part into the alto register. Later in this example, a FIJ occurs in the clarinet part, which is in the alto register before that FIJ, and in the bass register after that FIJ. The different registers are indicated by different shades of gray; (B) an excerpt sheet of music. Note that the same number of FIJs occurred in each part. Targets were defined as FIJs occurring in the violin part, and distracters were FIJs occurring in the saxophone and clarinet parts.

While the cocktail party phenomenon in a complex listening environment has mostly been described for language comprehension, listening to music often requires similar processes. Polyphonic music (i.e., multi-part music) can also contain several simultaneously active sound sources, creating multiple auditory streams. When different parts are played, for example, by different instruments, it is possible to focus attention selectively to one instrument and to follow the melody played by this instrument (Janata *et al.*, 2002). A first step in investigating the contribution of spatial information in selective listening to musical patterns was the "scale illusion" (Deutsch, 1975). In this paradigm, dichotic tonal sequences consisting of the repetitive presentation of an ascending and a descending scale were presented dichotically in a way that adjacent tones of the scales were switched from ear to ear. Most of the subjects heard instead of alternating scales, two streams, one from high- to medium-pitched and back to high notes, and a second from low- to medium-pitched and back to low notes. This indicates that stimuli were channeled mainly by pitch range rather than by the ear of input. Butler (1979) extended this paradigm to the use of other melodic materials and demonstrated that the effects can also be transferred to free field stimulation. Later, it has been shown that introducing differences in timbre can cause a degradation of the scale illusions (Smith *et al.*, 1982).

It has also been shown that a decrement in the integration of melodic patterns occurs if the tones of the pattern were distributed pseudorandomly between the ears compared to when presented binaurally (Deutsch, 1979). When a lower frequency tone (drone) was simultaneously presented to the ear opposite to that receiving the melody component, performance was largely improved again but not if the drone and the melody tone were delivered to the same ear. The authors suggested competition between two organizing principles: "Where input is to one ear at a time, localization cues are very compelling, so that linkages are formed on the basis of ear input and not frequency proximity. However, when both ears receive input simultaneously, an ambiguity arises as to the sources of these inputs, so that organization by frequency proximity becomes a more reasonable principle."

The present study specifically investigates the influence of spatial information for selective attention to one of three simultaneously presented melody parts of polyphonic music (see also Fig. 1). This phenomenon bears some similarities to spatial release from masking for which two mechanisms are currently discussed (Bronkhorst, 2000; Durlach, 1963; vom Hövel, 1984): (i) the binaural unmasking of lower frequencies facilitated due to different interaural time differences (ITDs) between competing sound sources; (ii) "best ear" listening, i.e., a benefiting signal-to-noise ratio at the ear ipsilateral to the target sound source and contralateral to the interfering sound sources caused by the headshadow effect. If binaural unmasking is effective, an improvement in attending to the relevant stream with increasing distance of the sources (i.e., increasing differences between ITDs related to the competing sound sources) would be expected. Indeed, Drennan *et al.* (2003) demonstrated an increase in the ability to segregate two competing speech sounds with increasing angle between the sources in the acoustic free field, as well as with increasing interaural time differences under headphone conditions.

In the present study, all three melodic parts were either presented from the same or from different locations, with the to-be-attended part being presented either at 0° or at one of the right-sided locations (i.e., +28° or +56° in the azimuthal plane). Due to an increasing influence of binaural unmasking, we hypothesized that the detection of targets occurring in the attended part would improve with increasing spatial separation of the different parts. An additional aspect of the present study was to find out whether the position of the relevant part, either frontal or lateral to the subject, influences it target detections.

II. METHODS

A. Subjects

The data of 20 right-handed and normal-hearing subjects (nine females) aged 22–30 years (mean age of 25.8 years) were included in this study (two female subjects were excluded after the practice blocks because they were unable to

detect the targets). All subjects were non-musicians, i.e., they had no formal musical training (apart from normal school education). None of the participants had a history of a neurological disease or injury. All subjects participated on a voluntary basis, gave written informed consent, and received monetary reimbursement.

B. Stimuli

Eight polyphonic music pieces, each consisting of three parts, and each with a length of approximately 3:45 min (ranging from 3:41 to 3:53 min) were created using the software CUBASE SX 2.01 (Steinberg Media Technology GmbH, Hamburg, Germany). Each part had a different computer-generated timbre (Violin, Saxophone, and Clarinet) and was synthesized into a single wav-file. The assignment of the three timbres to the different parts was the same for each subject and each music piece. The three single wav-files of one composition were merged into one multi-channel wav-file using MATLAB 7.1 (The MathWorks, Nattick, MA). The three different parts of one composition played in three different registers (soprano, alto, and bass). The register, in which a part plays, is defined by the pitch of the part in relation to the pitch of the other two parts (with the soprano playing the highest and the bass playing the lowest pitch).

The melodic contour of each part was non-monotonically ascending until interrupted by a target/distractor, which is a falling interval jump (FIJ) in the melody of at least 1 octave. The FIJs occurred in all three parts and always turned the part in which a FIJ occurred into the bass register [Fig. 1(A)]. Due to the general pattern of melodic ascent interrupted by sudden descents (FIJs), each part played in different registers during each composition [see Fig. 1(A) for a schematic illustration].

An important requirement concerning the stimulus design was that target detection should only be possible by selectively attending to the relevant part, instead of global listening to the overall sound of the music pieces. To control for this (i) FIJs did not violate the harmony of the overall sound (i.e., they did not induce dissonances); (ii) FIJs occurred only in a part, which was in the soprano or in the alto register (to avoid large changes in the frequency range of the overall sound); and (iii) no discernible breaks occurred in the melodic contour of any register even when a crossover of parts occurred due to a FIJ. For example, in Fig. 1(A), during the first FIJ the melodic contour in the soprano register does not change noticeably although the violin part has been "replaced" by the saxophone part, and likewise the saxophone by the clarinet and clarinet by the violin in the alto and bass registers, respectively.

If downward movements in the melodic contour occurred apart from FIJs, they only spanned an interval of at most three semitones and were therefore hardly confused with targets [Fig. 1(B)].

Each part in each piece of music contained 38 FIJs (19 while the part was in the soprano, and 19 while the part was in the alto register). The time period between successive FIJs (irrespective of the part) was 1333–4000 ms [Fig. 1(B)], with the first one in any part occurring not earlier than 4000 ms after the onset of the music piece.

The duration of single tones used for the compositions was 333–2667 ms, which is equivalent to an eighth note and a whole note played at a tempo of 90 beats/s. The tones directly followed each other, i.e., there was no interstimulus interval. The second tone of a FIJ always occurred on a beat and its duration was at least 667 ms (equivalent with a quarter note).

The three parts of a music piece were similar in style and contained the same components (tone durations, frequencies, and rhythmic patterns) in the same proportion. Our stimuli were major-minor tonal music and—except for the frequent crossings of parts—composed by following the classical theory of harmony (e.g., Hindemith, 1940).

Stimulus intensity was 50 dB sensation level (SL), i.e., 50 dB above the individual sensation threshold.

C. Task

The subjects were instructed to focus their attention on the melodic contour of the violin part (which was always the target part) and to detect the FIJs occurring in this part (targets). Subjects indicated detection as fast as possible by pressing a button on a response box. FIJs in the distractor parts (i.e., in the saxophone and clarinet parts) were to be ignored.

D. Apparatus and procedure

Testing was performed in an echo- and sound-attenuated room with walls, ceiling, and floor covered by acoustic foam with 5 cm^3 wedges. The setup consisted of a semicircular platform of 2.3 m in radius, raised 1.13 m above the floor, with speakers (Control IG JBL) positioned at 0° azimuth, 28° to the left and to the right (−28° and +28°, respectively), and 56° to the left and to the right (−56° and +56°, respectively). The complete setup was covered by black gauze, so the speakers were not visible to the subjects. During the tests, the subjects were comfortably seated on an adjustable chair in the center of the loudspeaker array. Subjects were asked to focus a fixation point at 0°. Head movements were prevented by fixing the head to the backrest. During testing, subjects were observed from a neighboring control room through a semitransparent mirror.

By use of the software PRESENTATION Version 9.07 (Neurobehavioral Systems, Inc., Albany, CA) single instrumental parts could be assigned to each of the five loudspeakers through an eight-channel soundcard (SB Audigy 2Zs Audio). Five experimental conditions were defined: an overlap condition
(0°/0°/0°), in which all instrumental parts were presented through one loudspeaker at 0°, and four separation conditions, in which the three different parts were presented from three different loudspeakers. The four separation conditions were subdivided into two with slight (+28°/−28°/0° and 0°/−28°/+28°) and two with wide (+56°/−56°/0° and 0°/−56°/+56°) speaker separations. Part locations are de-

TABLE I. Loudspeaker configurations for the different stimulus conditions. Part locations are depicted in the order: violin part (target part)/saxophone part/clarinet part as degrees from front.

Conditions	Violin/saxophone/clarinet
Overlap condition	0°/0°/0°
Separation conditions, target part frontal	0°/−28°/+28°
	0°/−56°/+56°
Separation conditions, target part lateral	+28°/−28°/0°
	+56°/−56°/0°

picted in the order: violin part (target part)/saxophone part/clarinet part as degrees from front.

Three different compositions (pseudorandomly chosen for each subject from the set of the eight compositions) were presented for each of the five loudspeaker configurations (see Table I). The order of the resultant 15 experimental blocks was randomized between subjects. During short breaks between the blocks, the subjects were informed whether the violin part (target part) will be presented up front or from the right side in the upcoming presentation.

Prior to data acquisition, subjects performed training blocks to assure that they were able to detect the targets. The training was subdivided into three phases: first and second, only the violin part was presented from 0° and subjects were instructed to just listen to the part (first phase) or to indicate target detection by pressing a button in response to targets (second phase). Subjects were qualified for the experiment only if they showed a success rate of at least 85% in the second phase. At last, a composition with all three parts was presented in the 0°/−56°/+56° loudspeaker configuration, with the target part being 10 dB louder than the distracting parts (to familiarize the subjects with the polyphonic sound of the compositions). For each phase of the training a different compositions out of the experimental set was chosen, which was balanced over subjects.

E. Data analysis

Behavioral responses were analyzed for the frequency of hits, selection errors, and detection errors. The respective response categorization was based on the reaction times to targets and distracters. Key presses 200–1100 ms after occurrence of targets were categorized as "hits," after distracters as "selection errors," and key presses outside these time windows as "detection errors."

For a better comparability of the conditions, all respective response categories were related to each other by means of a non-dimensional qualifier, in the following denoted as q-index. The range of the q-index is confined to an interval (0,…,1), with "1" indicating accurate detection of all targets and no responses in the selection and detection error categories. On the contrary, the q-index becomes "0" if at least one of the following three cases was observed: (i) no target was detected, (ii) the subject reacted to all distracters, or (iii) if the total number of responses equals the number of detection errors. This is formalized by the following equation:

$$q = \frac{H}{N_H} \cdot \frac{(N_{ES} - ES)}{N_{ES}} \cdot \frac{(n - ED)}{n}, \quad (1)$$

with H representing the response number of hits, ES is the response number of selection errors, ED is the number of detection errors, N_H is the maximum attainable number of hits, N_{ES} is the maximum attainable number of selection errors, and n the total of responses ($n=H+ES+ED$).

Differences in the q-index and in the reaction times between the overlap and separation conditions were quantified by paired t-tests under consideration of the alpha level Bonferroni corrections. For the separation conditions, the q-indices and the reaction times of the hits were subjected to two-way analyses of variance (ANOVAs) (factors: separation × position of the target part); Bonferroni multiple comparisons post hoc tests were used for comparisons between the conditions. The Greenhouse–Geisser correction was applied when the assumption of sphericity was violated. Differences were assessed as significant at an alpha level of 0.05.

The acoustic stimulation lasted for several minutes and was not subdivided into separate trials. In addition, subjects could respond at any time. As a consequence, a considerable number of hits could be achieved by an excessively high response rate. Because the number of responses was not limited, it was necessary to calculate the chance level, which takes these aspects into consideration. Therefore, the number of responses in the hits category and the q-index were tested against chance expectation with the chance level calculated based on an urn model (Johnson and Kotz, 1977). For that, each of the three answer categories (H=hits, ES=selection errors, and ED=detection errors) was assigned to a defined number (N_H, N_{ES}, and N_{ED}) of objects. N_H corresponds to the number of targets that are included in the target part (violin part), while N_{ES} corresponds to the number of distracters that are included in the saxophone and clarinet parts. N_{ED} will be ascertained, assuming that the number of n-responses (the total number of responses measured for a real subject) is uniformly distributed over the whole stimulus duration, with no relation to the stimulus structure, in a hypothetical subject (Poisson process, see Pitman, 1993). The distributions of the responses in the hits category correspond to a distribution while drawing n samples (without replacement), which is known in the literature as a hypergeometrical distribution (Pitman, 1993; Sachs, 1992).

III. RESULTS

A. Response rate

1. Influence of spatial separation of the parts

In all five conditions the target detection (number of hits and values for the q-index) was above chance level. The means in the three response categories for all five conditions are indicated in Table II. For a better comparability of the subjects' performance, the respective response categories were combined for the calculation of the non-dimensional qualifier, which is the q-index (see Sec. II for details) [Fig. 2(A)]. A higher q-index was found for the separation conditions than for to the overlap condition [0°/0°/0° ×

TABLE II. Response rate of hits, selection errors, and detection errors for all five conditions of part locations (mean, with SD in parentheses). Part locations are depicted in the order: violin part (target part)/saxophone part/clarinet part as degrees from front. The maximal attainable number of hits was 114 and number of selection errors was 228.

Part location (deg from front)	Hits	Selection errors	Detection errors
	Mean (SD)	Mean (SD)	Mean (SD)
0/0/0	62.7(16.8)	12.2(6.9)	7.5(4.9)
0/−28/+28	94.9(14.1)	2.3(1.8)	4.3(3.3)
+28/−28/0	93.7(14.3)	5.8(3.2)	3.8(3.1)
0/−56/+56	93.7(14.9)	2.9(−3.0)	4.9(5.3)
+56/−56/0	101.2(11.5)	2.9(2.7)	2.6(2.5)

separation conditions: $t_{0°/-28°/+28}(19)=-13.44$, $p_{0°/-28°/+28°}<0.001$; $t_{0°/-56°/+56°}(19)=-10.92$, $p_{0°/-56°/+56°}<0.001$; $t_{+28°/-28°/0°}(19)=-10.27$, $p_{+28°/-28°/0°}<0.001$; and $t_{+56°/-56°/0°}(19)=-15.17$, $p_{+56°/-56°/0°}<0.001$]. In general, the sensitivity for target detection differed strongly between subjects, as indicated by the high standard deviation. Still, in all subjects, the same tendency for an increase in the q-index with spatial separation of the parts was observed (for single subject data, see Fig. 3).

2. Increasing spatial separation at different positions of the target part

For the different separation conditions, the influence of the separation and position factors of the target part (position) was evaluated based on the respective q-indices [Fig. 2(B)]. Two-way ANOVAs (factors: separation × position) resulted in a significant main effect of separation [$F(1,19)=5.48$, $p<0.05$] but not of position [$F(1,19)=3.45$, $p=0.79$], and in a significant interaction between the separation and position factors [$F(1,19)=9.71$, $p<0.01$]. The Bonferroni multiple comparisons *post hoc* tests indicated a significant increase in the q-index with increasing spatial separation for lateral target positions [(+28°/−28°/0°) × (+56°/−56°/0°): $t(19)=-3.93$, $p=0.005$] but no effect of separation for frontal target positions [(0°/−28°/+28°) × (0°/−56°/+56°): $t(19)=0.36$, $p=0.999$]. Also, an effect of position could not be observed for moderate signal separations [(0°/−28°/+28°) × (+28°/−28°/0°): $t(19)=0.93$, $p=0.999$], while larger separation led to better target detection of laterally presented parts [(0°/−56°/+56°) × (+56°/−56°/0°): $t(19)=-3.42$, $p<0.05$].

B. Reaction time

The reaction times (RTs) for hits (Fig. 4) were significantly longer in the overlap condition compared to the separation conditions [(0°/0°/0°) × separation conditions: $t_{0°/-28°/+28°}(19)=3.79$, $p_{0°/-28°/+28°}<0.005$; $t_{+28°/-28°/0°}(19)=4.65$, $p_{+28°/-28°/0°}<0.005$; $t_{0°/-56°/+56°}(19)=3.51$, $p_{0°/-56°/+56°}<0.01$; and $t_{+56°/-56°/0°}(19)=4.32$, $p_{+56°/-56°/0°}<0.005$]. Comparisons within the different separation conditions showed neither a main effect of separation [$F(1,19)=0.15$, $p=0.7$] nor of position [$F(1,19)=1.69$, $p=0.21$] and also no interaction of these two factors [separation × position: $F(1,19)=1.48$, $p=0.24$].

IV. DISCUSSION

Using a target detection task, the influence of spatial information for selective attention to one of multiple streams was investigated with polyphonic music.

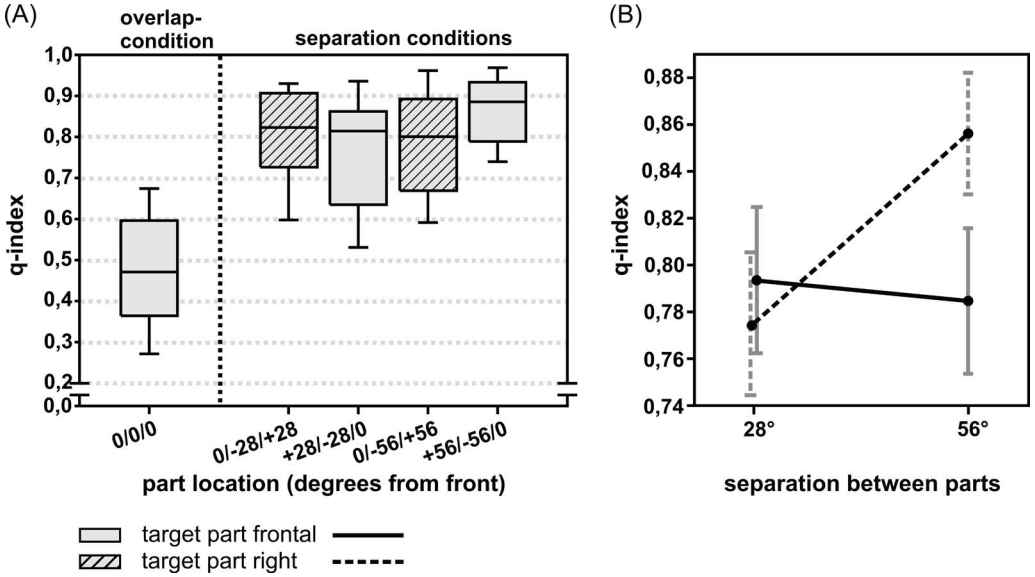

FIG. 2. (A) q-indices for performance in the different stimulus conditions and (B) the interactions of the q-indices between the factors separation and position of the target part in the separation conditions. The separation factor is divided into moderate (conditions 0°/−28°/+28° and +28°/−28°/0°) and large separations (conditions 0°/−56°/+56° and +56°/−56°/0°) of the three parts. The position factor is divided into frontal presentation of the target part (conditions 0°/−28°/+28° and 0°/−56°/+56°; solid line) and presentation of the target part on the right side (conditions +28°/−28°/0° and +56°/−56°/0°; hatched line). Box plots in (A) show medians, interquartile, and interdecile ranges. Target parts frontal are indicated by gray boxes, target parts lateral are indicated by hatched boxes. Mean values and standard error of the means are shown in (B). Part locations are depicted in the order: violin part (target part)/saxophone part/clarinet part as degrees from front.

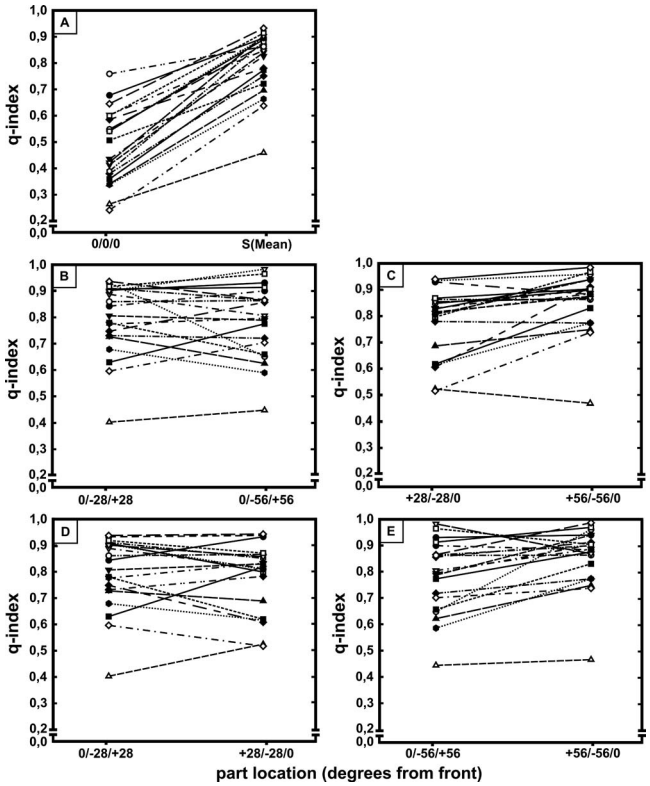

FIG. 3. Q-index per subject: (A) overlap condition (0°/0°/0°) compared to the mean of all separation conditions [S (mean)]; (B) increasing spatial separation of the three parts with target part frontal (condition 0°/−28°/+28° and 0°/−56°/+56°) and (C) lateral (condition +28°/−28°/0° and +56/−56/0°); comparison between target part frontal and lateral while same spatial separation of the parts for (D) moderate (condition 0°/−28°/+28° and +28°/−28°/0°) and (E) large separations of the parts (condition 0°/−56°/+56° and +56°/−56°/0°). Each line indicates the mean values for one subject. Part locations are depicted in the order: violin part (target part)/saxophone part/clarinet part as degrees from front.

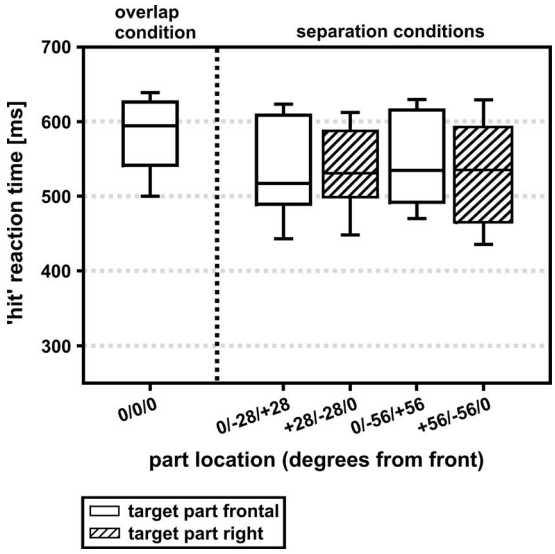

FIG. 4. Distribution of mean reaction time for hits in the overlap (0°/0°/0°) and the separation conditions (0°/−28°/+28°, +28°/−28°/0°, 0°/−56°/+56°, and +56°/−56°/0°). Median, interquartile, and interdecile values are shown. Part locations are depicted in the order: violin part (target part)/saxophone part/clarinet part as degrees from front.

A. Influence of spatial separation of the parts

In the overlap condition, 55% of the targets were detected, which is consistent with results of Janata *et al.* (2002), who also studied a selective attention task to polyphonic music. In that study, subjects were asked to detect timbral deviants in an attended stream during the presentation of three different instruments (different timbres). While in the selective attention condition in Janata *et al.*, 2002, timbral deviants only occurred in the attended stream, in the present study, the FIJs (in the melodic contour) occurred in all three parts. Still, only those in the violin part (targets) were task-relevant, whereas those in the saxophone and clarinet parts (distracters) had to be ignored. Moreover, the fact that FIJs did neither violate the frequency range nor the harmony of the overall sound impeded target recognition. Thus, targets could not be detected by integrative listening to global changes in the overall sound structure but only by selective listening to the relevant part and taking no account of the respective structures in irrelevant parts.

The possibility of explaining the target detection level by random responses could be ruled out for all subjects and for all conditions. Thus, in agreement with previous studies (Janata *et al.*, 2002; Yost *et al.*, 1996), the present results also show that, in a more challenging experimental setting, basal stream segregation of multiple sound sources is possible even without spatial separation of the sources just by the use of spectral information. However, additionally adding spatial information resulted in a significant increase in performance and decrease in reaction times. Even the relatively small spatial signal separation of 28° facilitates the allocation to different sound sources and improves listeners' ability to follow the relevant melodic contour. This result is consistent with previous findings based on the comprehension of spoken language, which showed a large influence of spatial separation for the processing of single sound sources in complex auditory environments (Hawley *et al.*, 1999, 2004; Peissig and Kollmeier, 1997; Yost *et al.*, 1996). Because in some of these studies the subjects had to reproduce the sentences/words after listening (Hawley *et al.*, 1999, 2004; Yost *et al.*, 1996), some constraints have to be made regarding the interpretation of these earlier results. Spoken language bears a lot of redundancy and this might have aided the performance of the subjects. On the other hand, reproduction of aurally perceived speech material relies on memory processes, for which was not controlled in these experiments. Our own endeavor was not hampered by any of these constraints and thus clearly disclosed the positive effects of spatial segregation on selective attention in complex acoustic environments.

B. Increasing spatial separation at different positions of the target part

Our results demonstrate that 28° spatial separation of the parts leads to significant improvement in target detection compared to the overlap condition, irrespective of the position (frontal or right) of the target part.

We hypothesized that, with increasing spatial separation of the relevant and the distracting parts, the influence of spatial unmasking further increases. As mentioned above, two mechanisms—best-ear listening and binaural unmasking—

are discussed to be responsible for spatial release from masking (Bronkhorst, 2000; Durlach, 1963; vom Hövel, 1984). During frontal presentation of the target part, masking signals were presented in both hemifields with equal distances to the target part. Note that, in each of the three parts, subsequent tones directly followed each other (with no silent gap between tones). This led to a continuous masking of the target signal from both sides during frontal presentation of the target part. For this reason, best-ear listening should not be very effective in this condition (even if it cannot be excluded entirely due to differences in the interfering parts). In contrast, during lateral presentation of the target signal, head-shadow effects might additionally have improved the signal-to-noise ratio for the relevant part, resulting in a monaural advantage at the best ear. Specifically, such a monaural advantage becomes effective if the interfering signals are in the same hemifield, instead of being located in different hemifields with the relevant sound source in-between (Hawley *et al.*, 1999, 2004; Peissig and Kollmeier, 1997; Zurek, 1992). Thus, during lateral presentation of the target part, both mechanisms—binaural unmasking and best-ear listening—should have been available for spatial unmasking, whereas during frontal presentation of the target part, binaural unmasking should be the dominant mechanism in the present experiment.

In agreement with our hypothesis, an increase in spatial separation of the parts resulted in a further improvement of target detection if the target part was lateralized (+56° to the right side). This goes in line with previous studies, which demonstrated an improvement of signal processing with increasing interaural time and intensity differences in acoustic speech material (Yost *et al.*, 1996) or noise-band material (Drennan *et al.*, 2003).

Previous reports showed that the ability to focus on a specific location is strongly reduced in the periphery compared to the central auditory space (Teder-Sälejärvi *et al.*, 1998, 1999). Also, spatial processing of sound sources, e.g., localization and spatial discrimination, is much better for frontal compared to lateral sound positions (Oldfield and Parker, 1984; Perrott *et al.*, 1993). Because of this enhanced frontal focus of selective auditory attention, one might expect that the processing of the relevant stream coming from the frontal loudspeaker is at least as good as when presented laterally. However, if in the present study the target part was presented frontally, no further improvement of target detection with increasing spatial separation from the irrelevant parts was observed. This lack of effect in the frontal target position suggests that, in the present experiment, the increase in spatial unmasking may be driven mainly by the best-ear listening.

However, signal modification by the pinnae might be the base for an alternative explanation of the results. Previous experiments demonstrated an amplification of the incoming signal, depending on the angle of incidence at the pinnae (Saberi *et al.*, 1991; Shaw, 1974). The highest amplification for the respective ear was demonstrated for lateral angles of 45° and 50°. Thus, presenting the target part from the lateral loudspeaker led to an additional increase in signal-to-noise ratio, which might help to separate the relevant part from the three concurrent streams. The other way around, during frontal presentation of the target part (with distracting parts on the left and right sides), the distracting parts will be amplified compared to the target part due to the more favorable angle of incidence at the pinnae. This effect should even increase with increasing spatial separation in the present experiment, which means that the intensity of the distracting parts increase compared to the target part. Interestingly, this would conflict with a possible benefit received by increasing binaural unmasking during the frontal presentation of the target part. But even if during frontal presentation of the target part, the signal-to-noise ratio decreases with increasing spatial separation (due to the pinnae related amplification of the distracting parts), the supportive influence of binaural unmasking should increase, and also the focus of auditory attention to frontal location should increasingly aid the selection of the target part at this position. Previous studies describe a gradual distribution of auditory attention (Mondor and Zatorre, 1995; Teder-Sälejärvi and Hillyard, 1998), which means that the amount of attention decreases with increasing distance of the sound source to the focus of attention. Because of the ability to specifically focus attention to the central auditory space, processing of the relevant part at frontal position (and ignoring of the lateral irrelevant parts) with increasing spatial separation should be facilitated. However, it is possible that the influence of these competing mechanisms (increasing binaural advantage and the gradient of attention on one side, monaural disadvantage on the other side) is equally strong in condition 0°/−56°/+56°. This might be the reason why during frontal presentation of the target part, we observed no difference in target detection with increasing spatial separation.

Also, impeded signal processing for frontal signals in a masking situation has been reported earlier. For example, deterioration in speech reception thresholds was documented for frontal target signals when the two masking signals are distributed in both hemifields, as compared to when they originate from one hemifield (Culling *et al.*, 2004; Hawley *et al.*, 1999, 2004; Peissig and Kollmeier, 1997). Similarly, using headphone stimulation, Treisman (1964) observed better word identification when the relevant message was presented to the right ear instead of being presented binaurally, i.e., perceived centrally. These results were explained by assuming object formation occurring outside the focus of attention, which improves with increasing similarity of the irrelevant sound sources, and which allows listeners to group them into a separate object distinct from the relevant sound source (Alain and Arnott, 2000; Alain *et al.*, 1996; Alain and Woods, 1993; Treisman, 1964).

No difference in performance was observed between frontal and lateral presentations of the target part during small separation angle of the parts (0°/−28°/+28° and +28°/−28°/0°, respectively). Considering the discussion above, one might expect better task performance during lateralized presentation of the target part also for small separation angles because the two distinct mechanisms (best-ear listening and the signal amplification by the pinnae) should be available in addition to binaural unmasking. The fact that this was not the case implies that these two monaural mecha-

nisms were only of minor significance in the small separation condition. However, it is still possible that the results indicate balanced effects of opposite mechanisms: While the focus of attention might, in addition to binaural unmasking, facilitate the processing of the frontal sound source, best-ear listening and the monaural advantage due to the spectral influence of the pinnae might be in favor of processing of the target part presented laterally.

In contrast to the target detection rate, response times seem to be unaffected by the position of the target part as soon as all three parts are presented from different locations.

V. SUMMARY AND CONCLUSIONS

Sustained selective attention to one musical instrument in polyphonic music is possible even without spatial separation, only by using spectral information. Introducing spatial separation between the relevant and the masking parts improved the selective processing of the relevant information. Increasing spatial separation from 28° to 56° between the different sound sources caused further improvements in target detection, if the relevant signal is presented laterally (with one masking signal presented from the opposite hemifield and one from 0°). If, under the identical arrangement of the sound sources, the relevant part is in front (with one masking signal presented from the right and one from the left side), no further improvement in target detection became evident with increasing separation of the distracters from 28° to 56° in the present experiment.

The competing influences of (i) best-ear listening, (ii) binaural unmasking, (iii) focus of attention, and (iv) pinna-related signal amplification on target detection were discussed. Still, the present stimulus design did not allow us to distinguish between the contributions of these alternatives. To make a distinction between the influences of binaural unmasking and monaural mechanisms, the influence of the interaural intensity differences (IIDs) and ITDs could be tested separately, following the experimental design employed by Culling *et al.* (2004). The usage of larger angles of separation would be helpful to prove the influence of the filter characteristics of the pinna. If the angle of incidence to the pinna has an effect, this influence should decrease with variations in the sound source around the angle of highest pinneal signal amplification.

Concluding, the present study demonstrates that the underlying mechanisms for solving the cocktail party problem are universal and can be utilized in selective listening to music as has been demonstrated for selective language processing earlier (e.g., Kidd *et al.*, 2005; Yost *et al.*, 1996).

ACKNOWLEDGMENTS

The authors wish to thank Frank-Steffen Elster for helping in composing the stimuli, Sven Gutekunst and Maren Grigutsch for the support in programming, Gerd Joachim Dörrscheidt for providing the calculation of the *q*-index and the coincidence level, and Sven Gutekunst and Peter Wolf for technical support, as well as Ramona Menger for acquisition and scheduling of subjects.

This work was supported by a stipend of the German Research Foundation (Postgraduate Training Program "Function of attention in cognition," Grant No. DFG 1182).

Alain, C., and Arnott, S. R. (**2000**). "Selectively attending to auditory objects," Front. Biosci. **5**, d202–212.

Alain, C., Ogawa, K. H., and Woods, D. L. (**1996**). "Aging and the segregation of auditory stimulus sequences," J. Gerontol. B Psychol. Sci. Soc. Sci. **51**, P91–P93.

Alain, C., and Woods, D. L. (**1993**). "Distractor clustering enhances detection speed and accuracy during selective listening," Percept. Psychophys. **54**, 509–514.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA).

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

Butler, D. (**1979**). "Further study of melodic channeling," Percept. Psychophys. **25**, 264–268.

Cherry, E. C. (**1953**). "Some experiments on the recognition of speech, with one and with 2 ears," J. Acoust. Soc. Am. **25**, 975–979.

Culling, J. F., 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.

Deutsch, D. (**1975**). "Two-channel listening to musical scales," J. Acoust. Soc. Am. **57**, 1156–1160.

Deutsch, D. (**1979**). "Binaural integration of melodic patterns," Percept. Psychophys. **25**, 399–405.

Drennan, W. R., Gatehouse, S., and Lever, C. (**2003**). "Perceptual segregation of competing speech sounds: The role of spatial location," J. Acoust. Soc. Am. **114**, 2178–2189.

Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206–1218.

Eramudugolla, R., McAnally, K. I., Martin, R. L., Irvine, D. R. F., and Mattingley, J. B. (**2008**). "The role of spatial location in auditory search," Hear. Res. **238**, 139–146.

Hawley, M. L., Litovsky, R. Y., and Colburn, H. S. (**1999**). "Speech intelligibility and localization in a multi-source environment," J. Acoust. Soc. Am. **105**, 3436–3448.

Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (**2004**). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," J. Acoust. Soc. Am. **115**, 833–843.

Hindemith, P. (**1940**). *Unterweisung im Tonsatz (The Craft of Musical Composition)* (Schott, Mainz).

Janata, P., Tillmann, B., and Bharucha, J. J. (**2002**). "Listening to polyphonic music recruits domain-general attention and working memory circuits," Cogn. Affect. Behav. Neurosci. **2**, 121–140.

Johnson, N. L., and Kotz, S. (**1977**). *Urn Models and Their Application: An Approach to Modern Discrete Probability Theory (Probability & Mathematical Statistics)* (Wiley, New York).

Kidd, G., Arbogast, T. L., Mason, C. R., and Gallun, F. J. (**2005**). "The advantage of knowing where to listen," J. Acoust. Soc. Am. **118**, 3804–3815.

Mondor, T. A., and Zatorre, R. J. (**1995**). "Shifting And focusing auditory spatial attention," J. Exp. Psychol. Hum. Percept. Perform. **21**, 387–409.

Münte, T. F., Kohlmetz, C., Nager, W., and Altenmüller, E. (**2001**). "Neuroperception—Superior auditory spatial tuning in conductors," Nature (London) **409**, 580.

Nager, W., Kohlmetz, C., Altenmüller, E., Rodriguez-Fornells, A., and Münte, T. F. (**2003**). "The fate of sounds in conductors' brains: An ERP study," Brain Res. Cogn. Brain Res. **17**, 83–93.

Oldfield, S. R., and Parker, S. P. A. (**1984**). "Acuity of sound localization—A topography of auditory space.1. Normal hearing conditions," Perception **13**, 581–600.

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.

Perrott, D. R., Costantino, B., and Cisneros, J. (**1993**). "Auditory and visual localization performance in a sequential discrimination task," J. Acoust. Soc. Am. **93**, 2134–2138.

Pitman, J. (**1993**). *Probability* (Springer-Verlag, New York).

Saberi, K., Dostal, L., Sadralodabai, T., Bull, V., and Perrott, D. R. (**1991**).

"Free-field release from masking," J. Acoust. Soc. Am. **90**, 1355–1370.
Sachs, L. (**1992**). *Angewande Statistik* (Springer-Verlag, Berlin).
Shackleton, T. M., Meddis, R., and Hewitt, M. J. (**1994**). "The role of binaural and fundamental-frequency difference cues in the identification of concurrently presented vowels," Q. J. Exp. Psychol. A, **47**, 545–563.
Shaw, E. A. G. (**1974**). *The External Ear* (Springer-Verlag, New York).
Smith, J., Hausfeld, S., Power, R. P., and Gorta, A. (**1982**). "Ambiguous musical figures and auditory streaming," Percept. Psychophys. **32**, 454–464.
Teder-Sälejärvi, W. A., and Hillyard, S. A. (**1998**). "The gradient of spatial auditory attention in free field: An event-related potential study," Percept. Psychophys. **60**, 1228–1242.
Teder-Sälejärvi, W. A., Hillyard, S. A., Röder, B., and Neville, H. J. (**1999**). "Spatial attention to central and peripheral auditory stimuli as indexed by event-related potentials," Brain Res. Cognit. Brain Res. **8**, 213–227.
Treisman, A. M. (**1964**). "The Effect of irrelevant material on the efficiency of selective listening," Am. J. Psychol. **77**, 533–546.
vom Hövel, H. (**1984**). "Zur Bedeutung der Übertragungseigenschaften des Außenohres sowie des binauralen Hörsystems bei gestörter Sprachübertragung (The influence of the transfer characteristics of the external ear and the binaural hearing aid during defective speech transmission)," in *Fakultät für Elektrotechnik* (RWTH, Aachen).
Yost, W. A., Dye, R. H., and Sheft, S. (**1996**). "A simulated 'cocktail party' with up to three sound sources," Percept. Psychophys. **58**, 1026–1036.
Zurek, P. M. (**1992**). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. A. Studebaker and I. Hochberg (Allyn & Bacon, Boston), pp. 255–276.