Top-Down Influences in the Detection of Spatial Displacement in a Musical Scene

GEORGIOS MARENTAKIS, Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz
CATHRYN GRIFFITHS and STEPHEN MCADAMS, McGill University

We investigated the detection of sound displacement in a four-voice musical piece under conditions that manipulated the attentional setting (selective or divided attention), the sound source numerosity, the spatial dispersion of the voices, and the tonal complexity of the piece. Detection was easiest when each voice was played in isolation and performance deteriorated when source numerosity increased and uncertainty with respect to the voice in which displacement would occur was introduced. Restricting the area occupied by the voices improved performance in agreement with the auditory spotlight hypothesis as did reducing the tonal complexity of the piece. Performance under increased numerosity conditions depended on the voice in which displacement occurred. The results highlight the importance of top-down processes in the context of the detection of spatial displacement in a musical scene.

Categories and Subject Descriptors: H.1.2 [**Models and Principles**]: User/Machine Systems—*Human Information Processing*; H.5.1 [**Information Interfaces And Presentation**]: Multimedia Information Systems—*Audio input, output*; H.5.5 [**Information Interfaces and Presentation**]: Sound and Music Computing—*Methodologies and Techniques*

General Terms: Human Factors

Additional Key Words and Phrases: Spatial attention, auditory perception, 3d audio

ACM Reference Format:

Georgios Marentakis, Cathryn Griffiths, and Stephen McAdams. 2016. Top-down influences in the detection of spatial displacement in a musical scene. ACM Trans. Appl. Percept. 14, 1, Article 3 (July 2016), 19 pages.
DOI: http://dx.doi.org/10.1145/2911985

1. INTRODUCTION

Music in which the spatial arrangement of voices is considered explicitly as a compositional parameter has been composed since Giovanni Gabrieli in the 16th century. Important contributions to the field occurred in the beginning of the 20th century by Henry Brant and Charles Ives. More recent composers—such as Edgard Varèse, Karlheinz Stockhausen, Pierre Boulez, Iannis Xenakis, and Roger Reynolds—have extended this tradition significantly. The situation in which voices originate from the entire space surrounding the listener and their locations are dynamically manipulated during a

This work was supported by a Canada Research Chair (950-223484) and a grant from the Canadian Natural Sciences and Engineering Research Council (NSERC, RGPIN 2015-05280) to Stephen McAdams, an NSERC (RGPIN 127-2007) grant to Albert Bregman, and the Zukunftsfonds Steiermark Klangräume Project (PN:6067) awarded to Georgios Marentakis.
Authors' addresses: G. Marentakis, Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz, Inffeldgasse 10/3, 8010 Graz, Austria; email: marentakis@iem.at; C. Griffiths and S. McAdams, Schulich School of Music, 555 Sherbrooke Street West, Montreal, Quebec, Canada, H3A 1E3; emails: cathryn.griffiths@gmail.com, smc@music.mcgill.ca.

DOI: http://dx.doi.org/10.1145/2911985

musical piece is currently considered to be commonplace in electroacoustic music and some pop music [Reynolds 2002; Harley 1994].

Listening to music involves both selective and divided attention processes. Both are facilitated by the improved segregation of individual voices that occurs when their pitch and timbre are different [Gregory 1990; Sloboda and Edworthy 1981; Hartmann and Johnson 1991]. Dividing attention between voices in a musical piece works surprisingly well in comparison to other divided-attention tasks in hearing, such as dividing attention between two speakers. Other than implying a music-specific cognitive function, this advantage has been attributed to the use of structural properties of music [Bigand et al. 2000; Jones and Yee 2001; Sloboda and Edworthy 1981; Gregory 1990]. Divided-attention performance is facilitated significantly by relatedness of musical keys, metric position, and musical structure relationships between musical voices [Crawley et al. 2002; Sloboda and Edworthy 1981; Agres and Krumhansl 2008]. Furthermore, hierarchically organized musical sequences are easier to remember [Deutsch 1979].

The spatial separation of voices in a piece improves the perception of individual voices. This is because it results in different monaural[1] and binaural cues to each voice. Falling-interval jumps or timbral variations in a target voice in a polyphonic musical piece are easier to detect when the target and the distracting voices are spatially separated [Saupe et al. 2010; Janata et al. 2002]. The extent to which the spatial separation of voices facilitates divided attention in music is debatable. On the one hand, the improved segregation of voices due to spatial differences has been found to facilitate the identification of interleaved melodies [Hartmann and Rakerd 1989]. However, improved segregation may compromise the perception of global harmonic and rhythmic relations between voices [Bregman 1990, p. 502]. As a result, it can limit the extent to which structural properties in music can be used to support divided-attention tasks. In addition, increased spatial separation between voices results in increased spatial-attention load, as listeners need to monitor multiple locations simultaneously.

The impact of the *spatial separation between voices in divided attention in music* thus needs to be investigated in more detail. This research question is central to this article. It is examined here using the task of the detection of voice displacement in a spatially distributed musical piece. In the experiments presented next, listeners performed this task in conditions that manipulated their attentional setting and the number, spatial dispersion, and tonal complexity of the voices comprising the piece.

2. BACKGROUND

The role of spatial attention in audition has been hotly debated. Starting from the early experiments of Cherry [Cherry 1953; Cherry and Taylor 1953], a considerable number of studies have investigated the extent to which top-down influences from perceived spatial location affect auditory perception.

Initial supporting evidence comes from a variety of detection tasks, in which cueing the target location improves performance [Spence and Driver 2004; Mondor and Zatorre 1995; Mondor et al. 1998; Woods et al. 2001; Sach et al. 2000]. To the contrary, uncertainty with respect to target location results in poor recollection of messages from unexpected locations and worse keyword identification performance [Arbogast and Kidd 2000; Yost et al. 1996; Cherry 1953].

Further evidence comes primarily from speech-on-speech informational masking[2] experiments. These have been conducted with running speech stimuli, which do not overlap completely in time and frequency, and with processed speech stimuli with negligible frequency overlap. They demonstrated

[1]Spatial separation of voices results in an improved signal-to-noise ratio for a given voice in the ipsilateral (better) ear.

[2]Informational masking refers to the masking observed in situations in which target and masker signals overlap little in time and frequency or in which there exists uncertainty with respect to target and masker. It is considered to be an interference effect in short-term memory.

significant spatial unmasking which (1) is more pronounced for informational (speech) rather than energetic (noise) maskers, (2) occurs irrespective of the type of cues used to spatialize target and masker(s) (monaural, binaural energy or time cues, time-difference-panning) and (3) is robust to reverberation, in contrast to speech-on-noise unmasking. It is important to note that the proportion of spatial unmasking due to binaural cues increases with scene complexity[3] [Arbogast et al. 2002; Kidd et al. 1998; Gallun et al. 2005; Shinn-Cunningham et al. 2005; Hawley et al. 2004, 1999; Freyman et al. 2001; Brungart et al. 2005; Freyman et al. 1999; Kidd et al. 2005b; Yost et al. 1996; Bronkhorst and Plomp 1992]. As signals in the aforementioned experiments do not overlap completely in time and frequency, the spatial unmasking observed cannot be explained fully on the basis of bottom-up cues[4] [Bronkhorst and Plomp 1988; Gallun et al. 2008; Shinn-Cunningham 2008]. Spatial attention in the target location should also be considered in order to fully account for the spatial unmasking observed [Arbogast et al. 2002; Kidd et al. 1998; Gallun et al. 2005; Shinn-Cunningham et al. 2005; Ihlefeld and Shinn-Cunningham 2008b]. In agreement with this observation, uncertainty about the target [Kidd et al. 2005a; Allen et al. 2009; Arbogast and Kidd 2000] and the masker location [Allen et al. 2011; Fan et al. 2008; Jones and Litovsky 2008] reduces spatial release from informational masking[5].

More specifically, the role of spatial attention is to assist the process of grouping and integration of the acoustic features that comprise the target stream. In this way, it helps to "counteract[s] failures in across-time linkage of segments and failures in the selection of target segments and/or the target stream" [Ihlefeld and Shinn-Cunningham 2008b; Shinn-Cunningham 2008]. This is reasonable considering that, although cues to spatial location interfere little with concurrent auditory grouping [Darwin and Hukin 1997; Hukin and Darwin 1995; Culling and Summerfield 1995; Darwin 2008], they are important in sequential grouping or auditory stream formation [Darwin and Hukin 1999; Sach and Bailey 2004; Stainsby et al. 2011]. More recent findings point to interactions between neural mechanisms of monitoring source content and source location. Different neural mechanisms for processing the "what" and the "where" of a sound have been identified in neurophysiological studies [Maeder et al. 2001; Bizley and Cohen 2013; Alain et al. 2001]. Attention can modulate neural activity and improve perception in both source-content- and source-localization-related tasks [Ahveninen et al. 2006]. In this view, source-relevant cues to spatial location are used to determine source location. Spatial attention can then be directed in order to increase perceptual sensitivity to target features, including location.

Spatial attention, being a top-down cue, may be subject to resource limitations. On these grounds, the auditory spotlight hypothesis has been formulated, which claims that auditory spatial attention operates like a spatially tuned filter, with performance decreasing away from its focus, similar to what happens in vision [Eriksen and James 1986]. Teder-Sälejärvi and Hillyard [1998] found increased sensitivity and reduced reaction times for targets around an attended location and a deterioration in performance for targets away from the attended location. Using a phoneme identification task, Allen et al. [2009] also confirmed that performance deteriorates with increased distance from the attended location, independent of whether gaze was directed away from that location.

3. PREDICTIONS WITH RESPECT TO THE DETECTION OF SOUND DISPLACEMENT

On the grounds of the literature review presented earlier, the following predictions with respect to the detection of spatial displacement in a voice in a musical scene can be made here.

[3]Scene complexity is manifested by the number of maskers used and, after accounting for better-ear listening, its impact is much larger for speech-on-speech masking compared to noise-on-speech masking when two or more interferers are present.

[4]Bottom-up cues, such as the higher signal-to-noise ratio at the ipsilateral (better) ear and the interaural cues to each voice, can account for the spatial release from energetic masking.

[5]This release occurs in situations in which the time scale within which task-relevant features are available does not allow for reorienting spatial attention away from the masker location.

Detecting displacement in a single voice in isolation: In this case, there is no competition for spatial attention and all available acoustic features are target-relevant. Good performance is expected even in small displacements; factors such as uncertainty in the timing of the displacements might, however, reduce detection performance compared to detection of displacement using standard psychophysical procedures [Hartmann and Rakerd 1989; Saberi et al. 1991; Mills 1958].

Detecting displacement in a single voice in the presence of competing (distracting) voices: Similarly here, the competition for spatial attention is limited, as this is directed to the location of the target voice. Limitations emerge due to interference from the rest of the voices, which increases uncertainty about the target-relevant cues to spatial location. Increased spatial separation between voices facilitates detection. It provides bottom-up cues and enables the focusing of spatial attention to assist in the allocation of target-relevant features. Because the competition for spatial attention is limited here, its impact may be hard to quantify. It may, however, be indirectly estimated by comparing performance between selective and divided attention tasks.

Detecting displacement in any of the voices: This divided-attention task introduces uncertainty with respect to the voice and the location of a possible displacement. The impact of uncertainty could be approximated by the performance difference between this condition and the one mentioned previously. The difference should increase in proportion to the contribution of top-down processing in the performance of the task. When voices are spatially distant, the competition for spatial attention resources may be managed by broadening (or splitting or time-sharing) the spatial focus of attention so that all relevant locations are covered [Eriksen and James 1986]. In such a case, the benefit due to spatial attention should decrease when the size of the area occupied by the sound sources increases, in accordance with the auditory spotlight hypothesis [Best et al. 2006; Eriksen and James 1986; Ihlefeld and Shinn-Cunningham 2008a].

Alternatively, attention may be directed to a specific voice location, and changes in the location of other voices may be extracted from memory. This is a variant of the strategy that listeners employed in the context of dividing attention between competing speech messages [Best et al. 2006; Ihlefeld and Shinn-Cunningham 2008a; Kidd et al. 2005a]. Results essentially reflect the level of performance that can be achieved in the absence of facilitation due to spatial attention. Unattended messages occupying variable spatial locations will lie, as a rule, outside the spatial focus of attention.

Contributions from Tonal Complexity: The variations in pitch common in music increase uncertainty in the target-relevant acoustic features to sound location. This is because the frequency region in which relevant features are contained shifts as a function of tonal height. Given that the contribution of spatial attention to performance increases with the uncertainty of target-relevant features, it can be expected that the benefit due to spatial attention will be smaller in simplified musical pieces, with little variation in pitch. In a similar fashion, spatial attention load with increasing spatial separation between the voices will be less for such simplified musical pieces.

In summary, the following points can be made here:

(1) A possible degradation in performance due to divided attention instructions would imply a top-down processing bottleneck relating to source or location uncertainty.
   (a) A further degradation, when increasing the area occupied by the voices, would provide evidence for resource limitations in allocating spatial attention and support for the auditory spotlight hypothesis.
   (b) The absence of further degradation, due to increasing the area occupied by the voices, would rule out limitations specific to spatial attention and highlight resource limitations specific to source identity.

(c) The magnitude of the degradation should be proportional to the tonal complexity of the piece. Accordingly, the cost due to increasing the area occupied by the sources should be smaller in pieces with low tonal complexity, for example, with constant pitch in each voice.

(2) If no degradation in performance is observed due to divided-attention instructions, this would rule out an effect of top-down processing and indicate that the task was performed solely on the basis of bottom-up cues.

4. PRESENTATION OF THE EXPERIMENTS

The hypotheses formulated in Section 3 were examined by displacing voices in a four-voice musical piece and measuring displacement detection performance. Using real instrument sounds played by musicians in the experiments would have been impractical. They would not be exactly reproducible from trial to trial, and musicians would have to physically move while playing. We opted to use synthesized instruments (as in Saupe et al. [2010]) and sound spatialization, in accordance with electroacoustic music practice. Possible influences due to the use of a spatialization system in the results are discussed in Section 6. A variant of the Minimum Audible Angle estimation procedure [Hartmann and Rakerd 1989; Mills 1958; Grantham 1995] was used to estimate detection performance in each condition.

4.1 Musical Piece and Orchestration

After reviewing a number of scores, we decided to use a 17th-century four-voice chanson by Claudin de Sermisy ("Pour n'avoir onc faulse chose promise," originally composed for choir, duration 2min 4s). This piece contains a similar level of melodic and rhythmic variation in all voices and only the two inner voices cross registers at three brief moments[6]. The score was coded as a multichannel MIDI file, and the following (synthetic) instruments were selected to render each voice: flute (V1), clarinet (V2), English horn (V3), and French horn (V4). Voice numbering reflects pitch register order from highest to lowest (originally soprano, alto, tenor, and bass). The instruments were chosen to be maximally distinguishable in terms of timbre, while making sense from an orchestration point of view. The French horn is often combined with woodwinds, as in a classic wind quintet. In orchestration treatises, the French horn is often considered a bridging instrument between the brass and the woodwind families [Samuel 2002, p. 312].

4.2 Conditions in the Experiments

In the experiments, we manipulate the Attentional Setting (Selective or Divided Attention), and the number, Spatial Dispersion, and Tonal Complexity of the voices. Measurements were executed using three participant groups. A summary is provided in Table I. Early in the analysis of the experiments, it became apparent that detection performance varied significantly depending on the voice in which spatial displacement occurred. This factor had not been considered initially. The effect was followed up in the different conditions in the experiments, and an additional condition (C5 in Table I) was introduced to help us clarify whether this was a voice- or position-related effect. More details are provided in Sections 5.2 and 6.

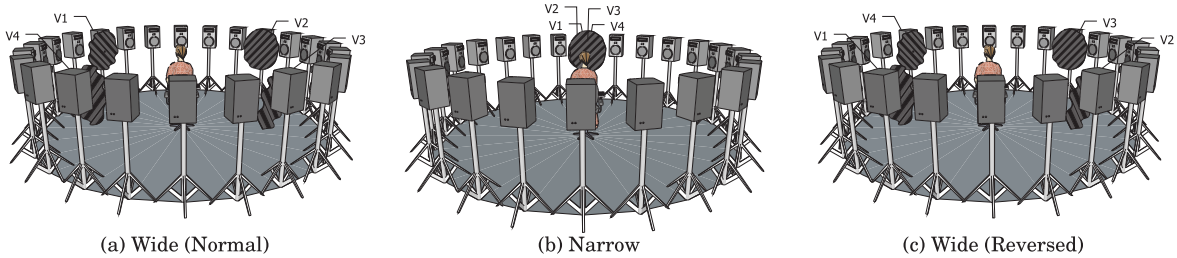
The Attentional Setting factor had three levels; participants detected spatial displacements (1) in a single target voice played in isolation, in which only one voice was active; (2) in a single target voice in the presence of the other three spatially fixed voices, in which all four voices were active; and (3) in any one of the voices in scenes comprising all four voices.

[6]However, timbre differences have been shown to facilitate voice tracking with crossing parts [Culling and Darwin 1993].

Table I. Presentation of the Conditions in the Experiment and the Associated Participant Groups

ID	Participant Group	N	Attentional Setting	Spatial Dispersion	Tonal Complexity	Displacements
1	A	1	Selective	Wide (Normal)	Normal	6°, 14°
2	C	1	Selective	Narrow	Normal	6°, 14°
3	A	4	Selective	Wide (Normal)	Normal	14°, 30°, 45°, 60°
4	A	4	Divided	Wide (Normal)	Normal	45°, 80°
5	B	4	Divided	Wide (Reversed)	Normal	45°, 80°
6	B	4	Divided	Wide (Normal)	Low	45°, 80°
7	C	4	Divided	Narrow	Normal	30°, 45°
8	C	4	Divided	Narrow	Low	30°, 45°

Note: ID is the condition number and N is the number of active voices.



(a) Wide (Normal) (b) Narrow (c) Wide (Reversed)

Fig. 1. Illustration of the Experimental Setup. Voices were displaced around their nominal position. Participants registered the displacement by pressing a button on the keyboard. Voices are illustrated here in the Wide (Normal (a) and Reversed (c)) versus the Narrow (b) configuration.

The Spatial Dispersion factor had three levels (see Figure 1): (1) Wide, Normal Locations with voices at −45° azimuth for flute (V1), 45° azimuth for the clarinet (V2), 135° azimuth for the English horn (V3) and −135° azimuth for the French horn (V4), (1a) Wide Reversed Locations[7], same as 1, but with reversed Voice locations (V4 with V1 and V3 with V2) and (2) Narrow, in which the nominal location for all voices was at 0° azimuth. Elevation was always 0°.

The Tonal Complexity factor had two levels: (1) Normal, in which the piece was rendered as it appeared in the score; and (2) Low, in which the rhythmic structure of the piece remained the same, but the voices each played a single pitch—G4 (a fundamental frequency of 391Hz) was assigned to V1 (flute), B4 (493Hz) to V2 (clarinet), G3 (195Hz) to English horn (V3) and E3 (164Hz) to French horn (V4).

4.3 Participants and Procedure

Experiments were performed on three occasions using three different participant groups (A, B, and C), in which the same apparatus was used. Group A consisted of 16 nonmusicians (11 female; mean age = 21 years, SD = 4.0 years) and 16 musicians[8] (7 female; mean age = 23, SD = 5.2 years), who completed measurements in Conditions 1, 3, and 4 in two experimental sessions on different days, each lasting about 1.5h. Condition 1 and half of Condition 3 (angular displacements of 14° and 30°) were completed in the first session; the remaining conditions and Condition 4 measurements were completed in the

[7]The Reversed Locations level of the Spatial Dispersion factor was introduced to investigate the effect of Voice in the Attentional Setting manipulations, as explained in Section 5.2, but is presented here to assist in the presentation of the overview.
[8]Musicians were practicing a musical instrument regularly; nonmusicians did not currently play a musical instrument and had less than 2 years training with a musical instrument during childhood.

second session. The order of conditions and the presentation sequence of instrumental voices within each condition were counterbalanced within each session.

Group A consisted of separate musician and nonmusician subgroups because Musicianship was also initially considered as a factor, together with Attentional Setting. Given the null effect of Musicianship (see Section 5.1) in Group A, Musicianship was not controlled in participant groups B and C, with the help of which the impact of Voice, Spatial Dispersion, and Tonal Complexity was examined.

Group B consisted of 16 participants (14 female; mean age = 24y, SD = 7.5), who completed the measurements in Conditions 5 and 6 in an hour-long session within which the two conditions were completed in a counterbalanced order.

Group C[9] consisted of 13 participants (3 remale; mean age = 31y, SD = 6.7), who completed Conditions 2, 7, and 8 in one session lasting 1h. Participants completed Condition 2 first, then Conditions 7 and 8 in counterbalanced order.

Participants in all groups were given a practice run to ensure that they understood the task prior to beginning the experiment.

4.4 Embedding Spatial Displacements in the Score

To measure spatial displacement detection performance, the voice(s) in each condition were displaced symmetrically on an arc around their condition-specific nominal position. Arc magnitude was condition-specific and had been determined in pilot experiments (see Table I and Figure 1). Voices were displaced on the arrival of a specific MIDI event. Twenty-four such events were embedded in the score[10]. Events were subdivided into two classes of 12 events each, each distinguished by its own MIDI event number. This design allowed for the estimation of detection performance for a specific voice, displacement magnitude, and condition in the experiment on the basis of 12 test and 12 catch trials each time the piece played. For displacement (test) trials, a detection within 2s succeeding the onset of a test trial was scored as a hit and no response within this time window was scored as a miss. A response within 2s succeeding a catch trial was scored as a false alarm and no response during this period was registered as a correct rejection.

The score was repeated according to the number of displacements tested in each condition. In conditions in which more than four iterations were necessary, a pause was given after a maximum of four repetitions. Specifically, there were two iterations per voice to allow for the two spatial displacements per voice in the condition with selective attention to a voice in isolation (C1) and four iterations per voice to allow for the four spatial displacements per voice in the condition with selective attention to a voice in the presence of distracters (C3). Each time the musical score iterated, the class that had signaled spatial displacements was switched with the one that signaled catch trials in order to prohibit event anticipation. In the divided-attention conditions, the same design results in 3 test trials per voice and 12 catch trials in total each time the piece iterated. The piece was repeated eight times to obtain measurements for the two displacement magnitudes for all voices. A hard-coded random map determined which voice would be displaced on the arrival of a displacement event. The map was constructed so that no voice's location would change more than twice in succession and that a different voice would be displaced at a given time point each time the score iterated.

[9]Because of availability problems, this set of measurements took place in a different room. A control test found no differences in performance between the two rooms (see Appendix).

[10]MIDI events were embedded in the score at predetermined locations at which (1) all four voices attacked the notes simultaneously, (2) no rest preceded any of the voices, (3) a response-time window of at least 2s after each event was available, and (4) the pitch of each voice reflected the initial register order. Event locations were identical for each voice and remained constant throughout all experimental conditions.

4.5 Apparatus

In all conditions, the synthesized voices were rendered as simulations of different instruments using the Synful Orchestra plug-in (Synful LLC, Woodland Hills, CA) [Lindemann 2007] for Max/MSP. Voices were spatialized using Vector-Based Amplitude Panning software [Pulkki 2001]. A Mac Mini computer (Apple Computer, Cupertino, CA) running Max/MSP software (Cycling '74, San Francisco, CA) controlled the experiment.

Participants were seated on a chair at the center of a circular array of 24 Genelec 8020A loudspeakers (Genelec, Iisalmi, Finland) with a radius of approximately 2m. They were provided with a computer keyboard with which they would indicate their responses throughout the experimental trials. Participants were instructed to press the Space key on the keyboard whenever they perceived sound displacement, irrespective of its direction. The levels of each voice were set to approximately 55dB SPL, as measured with a Bruel & Kjaer 2250-D sound-level meter positioned at the center of the loudspeaker array.

Due to availability problems, measurements for Groups A, B, and Group C occurred in different rooms, both acoustically treated, with similar dimensions and RT_{60} s. Room 1 dimensions were 7.2m (l)$\times$ 5.8m (w) $\times$ 2.4m (h) and Room 2 dimensions were 6m (l) $\times$ 4m (w) $\times$ 3m (h). RT_{60} [Farina and Tronchin 2013] at 63, 125, 250, 500, 1k, 2k, 4k, 8k, and 16 kHz, was 1.40, 0.70, 0.34, 0.32, 0.20, 0.18, 0.16, 0.15, 0.13s, respectively, for Room 1, and 1.7, 1.03, 0.60, 0.64, 0.55, 0.53, 0.54, 0.51, 0.42, 0.46, and 0.40s, respectively, for Room 2 at the same frequencies. Performance did not vary significantly across rooms (see Appendix).

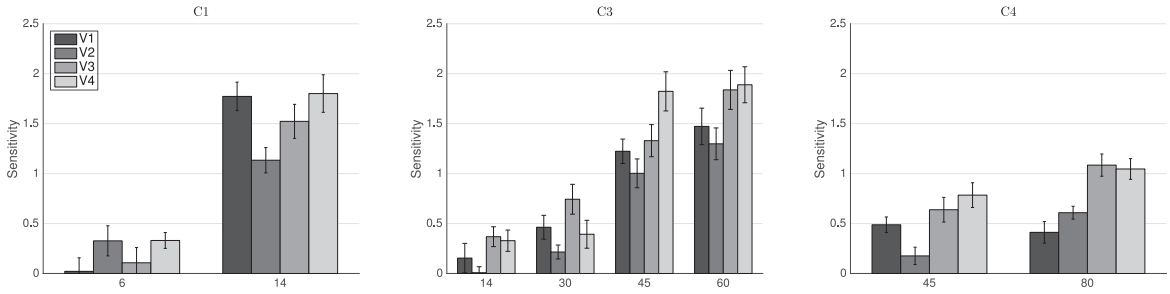
5. RESULTS

In this section, we focus on describing detection performance in the different conditions in the experiment using the measure of participants' sensitivity. Sensitivity is calculated as $d' = z(H) - z(FA)$, where z is the inverse of the cumulative normal distribution function and H is the hit and FA the false alarm rate [Macmillan and Creelman 2005]. In the Divided Attention conditions, it is not possible to allocate the occurrence of a false alarm to a specific voice, as all occur simultaneously. Therefore, a single global false alarm rate per displacement and condition was calculated and used in the analysis.

The statistical analysis is based on analyses of variance (ANOVAs). Given significant main effects or interactions, these are analyzed further using pairwise t-tests, corrected for multiple comparisons using the Bonferroni-Holm correction. When comparing conditions in which different participant groups were involved, the results obtained using the nonmusician dataset of Group A are reported in this article. As would be expected given the null effect of Musicianship in Section 5.1, the results of between-group comparisons are essentially the same irrespective of whether the musician or the nonmusician dataset for Group A is used in the comparisons.

5.1 Effect of Attentional Setting

The results presented here analyze and compare Conditions 1, 3, and 4 in which Participant Group A was employed and Attentional Setting was varied, from selective attention to a single voice in isolation (C1), to selective attention to a single voice in the presence of distracters (C3), to divided attention to any of the voices (C4). Spatial Dispersion was fixed in these conditions at the wide, normal locations level (as in Figure 1(a)) and Tonal Complexity was fixed at the normal level. As mentioned earlier, Participant Group A consisted of two subgroups of equal size, one comprising musicians and one nonmusicians. In all comparisons that follow, the main effect of musicianship and the interactions in which it was involved were not significant and are not discussed further.



(a) Attending to a single voice in isolation.

(b) Attending to a single voice in the presence of distracters.

(c) Detecting displacement in any of the voices.

Fig. 2. Sensitivity as a function of displacement, voice, and the attentional setting in Conditions 1, 3 and 4 (see Table I). Plotted data were averaged across Musicianship. Error bars represent the standard error of the mean.

5.1.1 *Displacement Effects within Attentional Settings.* Within each of the C1, C3, and C4 conditions, a three-way Displacement × Voice × Musicianship ANOVA was performed on sensitivity, with Voice and Displacement as within-subjects factors and Musicianship as a between-subjects factor. With reference to Figure 2, the effect of displacement was significant in all conditions and sensitivity increased significantly with increasing angular displacement, C1: $F(1,30) = 142.4$, $p < 0.001$; C3: $F(3,90) = 90.7$, $p < 0.001$; C4: $F(1,30) = 15.1$, $p = 0.001$. The effect of Voice and the interaction between Voice and Displacement was also significant in the increased source numerosity conditions (C3 and C4). These are analyzed in detail in Section 5.2.

5.1.2 *Performance between Attentional Settings.* Performance between pairs of the three conditions in which the attentional setting was varied (C1, C2, and C3) was compared using a Condition × Voice ANOVA on sensitivity for the single angular displacement that was shared between conditions (14° for C1 and C3, 45° for C3 and C4). Increasing source numerosity under selective attention instructions resulted in a significant reduction in sensitivity (C1 vs. C3 @14°, $F(1,30) = 167.03$, $p < 0.001$). Performance further deteriorated significantly under divided attention instructions (C3 vs. C4 @45°, $F(1,30) = 82.01$, $p < 0.001$). The analysis verified what is evident in Figure 2: increasing the number of active voices and introducing uncertainty with respect to which voice would move next resulted in decreased sensitivity.

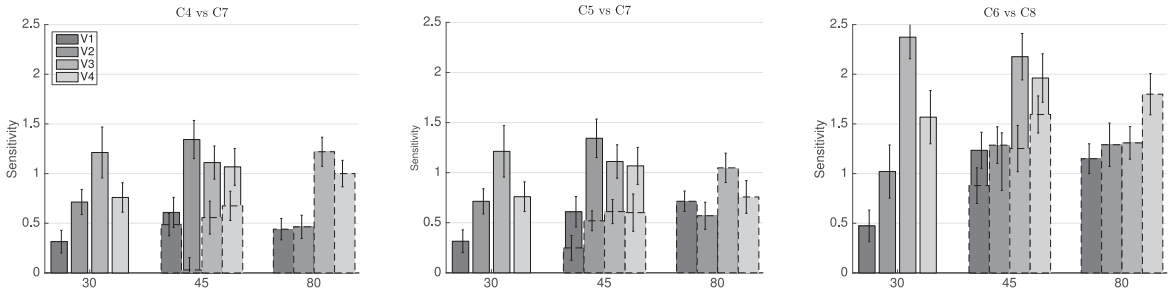
5.2 Effect of Voice

5.2.1 *Voice Effects within Attentional Settings.* As mentioned in Section 5.1.1, an effect of Voice was observed in C3, $F(3,90) = 7.9$, $p < 0.001$ and C4, $F(3,90) = 14.1$, $p < 0.001$. In C3, sensitivity for V4 and V3 was significantly higher than for V2, and sensitivity for V4 was significantly higher than for V1.

In C4, sensitivity for V4 and V3 was significantly higher than for both V2 and V1. In addition, the Voice × Displacement interaction was also significant for sensitivity: $F(3,90) = 5.5$, $p = 0.002$. Sensitivity was significantly higher for V4 and V3 than for V1 and V2 when displacement magnitude was 80° ($p < 0.005$), but not when it was 45°.

In summary, these results show that sensitivity to sound displacement was, as a rule, higher for the French (V4) and English (V3) horns in the back positions than for the flute (V1) and clarinet (V2) in the front positions and that, in C4, the sensitivity advantage increased with angular displacement.

5.2.2 *Further Observations.* At first glance, the effect of Voice and the interactions it was involved in can be attributed to either the location of the voices at the back of the listeners or to voice-specific



(a) Wide (Normal Locations) vs. narrow Spatial Dispersion, with normal Tonal Complexity.

(b) Wide (Reversed Locations) vs. narrow Spatial Dispersion, with normal Tonal Complexity.

(c) Wide vs. narrow Spatial Dispersion, with low Tonal Complexity.

Fig. 3. Illustration of impact of Spatial Dispersion on sensitivity. Wide (wide dashed bars) and narrow (narrow solid bars) show Spatial Dispersion conditions. Error bars represent the standard error of the mean.

aspects, such as the interaction between their melodic contours, their register, or the timbre of the individual voices. To clarify whether location or other voice-specific characteristics account for the observed effect, C5 was introduced, in which the Location of the voices was reversed (as in Figure 1(c)). A null effect when comparing Conditions 4 and 5 would indicate that the effect could be attributed to voice-related characteristics and not location.

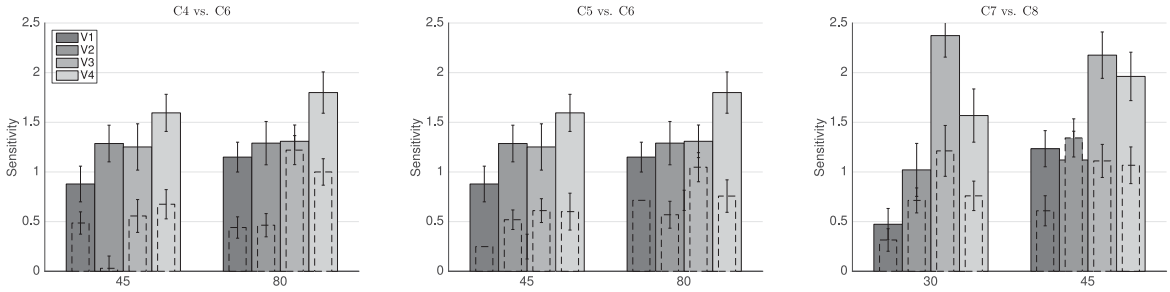
To compare Conditions 4 and 5, a Location (Normal vs. Reversed) $\times$ Voice $\times$ Displacement ANOVA was performed on sensitivity, with Location as between-subjects and Voice and Displacement as within-subjects factors. There was no effect of Location on sensitivity, whereas the effect of Voice persisted ($F(3,90) = 12.9$, $p < 0.001$), with V4 and V3 yielding significantly higher sensitivity ($p < 0.01$) compared to V1 and V2.

The effect of Voice on sensitivity was significant in all divided-attention conditions in the experiments, irrespective of Spatial Dispersion or Tonal Complexity setting, in C5: $F(3,45) = 3.5$, $p = 0.022$; in C6: $F(3,45) = 6.6$, $p = 0.001$; in C7: $F(3,36) = 7.2$, $p = 0.001$; and in C8: $F(3,36) = 14.7$, $p < 0.001$. With small variations, V3 and V4 resulted in significantly higher sensitivity compared to the other two voices. Specifically, although there was never a difference between V3 and V4 in sensitivity, in C5, sensitivity was significantly higher for V3 compared to V1 and V2; in C6, sensitivity was significantly higher for V4 compared to V1; and in both C7 and C8, both V3 and V4 yielded significantly higher sensitivity than V1.

The null effect of Experiment and persistence of the Voice effect in the aforementioned comparisons rules out an advantage due to location, spatial dispersion, or melodic trajectory. It implies that the effect of Voice may be attributed to other factors, such as auditory salience, that may relate to register or timbre.

5.3 Effect of Spatial Dispersion

The effect of Spatial Dispersion in the Divided Attention conditions was investigated by comparing detection performance between the wide and narrow Spatial Dispersion conditions at the 45° angular displacement that was shared between conditions. Figures 3(a) and 3(b), which illustrate the effect of Spatial Dispersion in conditions in which Tonal Complexity was normal (i.e., C4 and C5 vs. C7), show a tendency for increased sensitivity in the narrow compared to the wide Spatial Dispersion conditions. However, Figure 3(c), in which the same comparison is illustrated for the case in which Tonal Complexity was low (i.e., C6 vs. C8), shows that the magnitude of the advantage is smaller in this case.



(a) High and low Tonal Complexity conditions with wide Normal Locations.

(b) High and low Tonal Complexity conditions with wide Reversed Locations.

(c) High and low Tonal Complexity conditions with narrow Spatial Dispersion.

Fig. 4. a-c: Illustration of the effect of Tonal Complexity on sensitivity. Wide solid bars indicate performance, with low and narrow dashed bars with normal Tonal Complexity. Error bars represent the standard error of the mean.

Sensitivity at the common angular displacement of 45° between the two conditions in each panel of Figure 3 was analyzed statistically using a Voice × Spatial Dispersion ANOVA, with Voice as within-subjects and Spatial Dispersion as between-subjects factors.

In both comparisons in which Tonal Complexity was normal, sensitivity to spatial displacement when dividing attention across voices was significantly higher in the narrow compared to the wide Spatial Dispersion condition: C4 vs. C7: $F(1,27) = 13.4$, $p = 0.001$; C5 vs. C7: $F(1,27) = 10.2$, $p < 0.001$. When Tonal Complexity was low, a similar tendency was observed. However, the effect was not significant: C6 vs. C8: $F(1,27) = 2.5$, $p = 0.12$.

In summary, having all voices located within a smaller spatial region makes spatial displacement detection easier. However, the magnitude of the effect diminished when Tonal Complexity was reduced.

5.4 Effect of Tonal Complexity

The effect of the Tonal Complexity in the Divided Attention conditions was investigated by comparing detection performance between the low (constant-pitch) and high (variable-pitch) Tonal Complexity conditions. Figure 4 illustrates the effect of Tonal Complexity for the cases in which Spatial Dispersion was wide (Figures 4(a) and 4(b), C4 and C5 vs. C6) and the cases in which it was narrow (Figure 4(c), C7 vs. C8). In all cases, reducing Tonal Complexity resulted in increased sensitivity.

A Tonal Complexity × Displacement × Voice ANOVA was used to compare sensitivity in each pair of conditions in each panel of Figure 4. Displacement and Voice were treated as within-subjects factors. Tonal Complexity was treated as a within-subjects factor in the wide Spatial Dispersion comparisons and as a between-subjects factor in the narrow Spatial Dispersion Condition comparison. The effect of Tonal Complexity was significant in all comparisons, C8 vs. C7: $F(1,12) = 29.5$, $p < 0.001$, C5 vs. C6: $F(1,30) = 18.9$, $p < 0.001$, C4 vs. C6: $F(1,30) = 21.1$, $p < 0.001$. This indicates that reducing tonal complexity resulted in increased sensitivity irrespective of Spatial Dispersion.

5.5 Variations in Hit and False Alarm Rates

Here, some notable observations with respect to the hit and false alarm rate in each condition in the experiments are summarized, in order to prepare the reader for the discussion in Section 6.

Hit rate in the experiments varied pretty much in the same way as did sensitivity. Hit rate increased significantly with voice displacement magnitude, decreased due to increasing scene numerosity or attentional load, and was significantly higher for the voices in which a detection advantage was observed. Hit rate increased when spatial dispersion was reduced, but not significantly. Hit rate increased

significantly when Tonal Complexity was reduced, when Spatial Dispersion was wide, but not when Spatial Dispersion was narrow.

A high false-alarm rate was observed in all conditions (range between 10% and 40%). False-alarm rate was not influenced significantly by variations in the attentional setting of the participants, or by the numerosity, spatial dispersion, or tonal complexity of the musical scene. Even changes in displacement magnitude barely affected false-alarm rate. A significant reduction was observed only for the large displacements of C3, $F(3,90) = 10.56$, $p < 0.001$, and when reducing the Tonal Complexity of the piece in the conditions in which spatial dispersion was narrow, C7 vs. C8: $F(1,12) = 22.2$, $p = 0.001$.

Therefore, it is evident that most of the variation in sensitivity can be explained by the variation in hit rate. The magnitude of the differences in false-alarm rate in the different conditions was small and was only seldom found to be significant. The persistent high false-alarm rate in the experiments indicates that participants systematically reported voice displacements when none actually occurred. This is discussed in more detail in Section 6.

6. DISCUSSION

In summary, the results provide evidence that the factors investigated in the experiment— Attentional Setting, Spatial Dispersion, and Tonal Complexity—influence detection of spatial displacement in music. In addition, it has been found that displacement performance is affected by the voice in which displacement occurs. Displacement was easier to detect for some voices in comparison to the rest when the number of musical sources increased.

Here, we discuss and elaborate on the way that the experimental manipulations influenced performance in relation to the hypotheses formulated in Section 3. Possible influences due to the use of a spatialization system in the results are discussed at the end of the section. To help with the discussion of the results in the following, each of the aforementioned factors is discussed in a separate section.

Although, initially, Musicianship was considered as a factor, no effect of it was observed in the Attentional Setting manipulations. This factor was subsequently excluded. Musical training has resulted in superior performance when participants were asked to detect changes in a melody [Agres and Krumhansl 2008; Crawley et al. 2002] or perform a selective attention masked threshold task [Oxenham et al. 2003]. The null effect of musicianship observed here may be because musical training does not normally involve training in spatial hearing tasks.

6.1 Impact of Attentional Setting

The attentional setting of the listener had a profound influence on the results. Even in the simple case in which displacement in a single voice played in isolation was detected, approximate detection thresholds estimated by linear interpolation are on the order of 10°. This is much larger than the typical thresholds of around 3° for laboratory stimuli located at similar azimuth [Mills 1958]. The following factors may have contributed to this result. The first relates to the uncertainty with respect to the exact timing of the displacements. As mentioned in Section 4.4, the timing of the displacements was only approximately periodic. Timing uncertainty has been known to influence detection performance in both speech and spatial detection tasks. Although its effect is smaller compared to location uncertainty, it cannot be discounted as a factor influencing the results [Kitterick et al. 2010; Gatehouse and Akeroyd 2008; Chandler et al. 2005]. Second, thresholds reported here are estimated using signal detection theory that takes not only hit rate, but also false-alarm rate, into account. The high false-alarm rate that was observed in the experiments, which persisted throughout the different conditions, arguably contributed to the increased thresholds reported here. Our working hypothesis concerning the origin of the false-alarm rates is that they relate, at least in part, to the Tonal Complexity of the piece (see Section 6.2).

The results showed a significant performance deterioration under selective attention instructions when the source numerosity was increased. Estimated thresholds increased from 10° to approximately 43°, 50°, 35°, and 35° for V1 to V4, respectively, when all voices played simultaneously. The source of the deterioration is arguably the interference in the cues to sound location in the attended voice due to the remaining voices. The observed performance deterioration may constitute the operational definition of informational masking in our experiments, similar to the observations in the speech-on-speech masking experiments in the literature [Arbogast et al. 2002; Kidd et al. 1998; Gallun et al. 2005; Shinn-Cunningham et al. 2005; Ihlefeld and Shinn-Cunningham 2008b; Bronkhorst and Plomp 1988; Gallun et al. 2008; Shinn-Cunningham 2008; Freyman et al. 2001; Brungart et al. 2005; Hawley et al. 2004, 1999; Yost et al. 1996; Bronkhorst and Plomp 1992]. The significant further deterioration in performance in the Divided Attention condition, evident when comparing performance to the Selective Attention with Distracters condition at 45°, indicates a further limitation due to the use of a top-down process such as attention in the Selective Attention with Distracters condition.

More specifically, the origin of the resource limitations observed when comparing divided and selective attention performance with all voices active can be attributed to the uncertainty in the timing, identity, and location of the voice in which displacement would occur [Kidd et al. 2005a; Kitterick et al. 2010; Allen et al. 2009; Chandler et al. 2005]. The cost of location uncertainty could be approximated by the difference in performance between the narrow and the wide Spatial Dispersion and Divided Attention conditions, whereas that of voice identity uncertainty is gleaned by the difference in performance between Divided Attention in the narrow Spatial Dispersion condition and the Selective Attention with Distracters condition. On this basis, averaged across voice and assuming a similar contribution of timing uncertainty in all conditions, performance deteriorated by 60% from selective to divided attention, 40% of which was due to location and 20% due to voice identity uncertainty. It should be noted, however, that nonlinear interactions between timing and location uncertainty may need to be considered to fully account for the deterioration in detection performance due to location uncertainty calculated earlier [Chandler et al. 2005]. However, the estimation of such nonlinear effects was outside the scope of this study.

The design of the experiments may have emphasized the effect of spatial attention by limiting the ability of listeners to perform the task on the basis of monaural cues. Although spatial location in azimuth cannot be determined accurately on the basis of monaural cues [Hawley et al. 1999], in theory spatial displacement could. In such a case, listeners could have detected a displacement not on the basis of a sound location change, but on the basis of differences in the energy received by each ear. The locations of the voices and the magnitude of the displacements were, however, such that monaural differences would be relatively small. Both the nominal and the displaced locations of the four voices were roughly symmetric and located in both hemispheres, and voices were either displaced within the same quadrant or around midline. This may have increased listeners' reliance on binaural cues when all voices were active [Hawley et al. 2004; Bronkhorst and Plomp 1992]. For example, had the voices been displaced to enter another hemisphere in the wide Spatial Dispersion condition, the importance of monaural cues may have increased, as energy balance between the two ears would have shifted radically.

6.1.1 *Evidence in Support of the Auditory Spotlight Hypothesis.* There was a significant detection advantage that emerged in the Divided Attention conditions when the spatial dispersion of the voices was decreased. This result argues in support of the auditory spotlight hypothesis and emphasizes the use of top-down processing in the experiments in agreement with Best et al. [2006], Allen et al. [2009], and Teder-Sälejärvi and Hillyard [1998]. In the context of our task, narrowing the area occupied by the voices resulted in improved processing and better detection performance for all sounds within the

region of spatial focus. In agreement with the literature [Hawley et al. 2004, 1999], the magnitude of the benefit appears to depend on the acoustic complexity of the musical scene. The magnitude of the benefit was reduced by about 40% when Tonal Complexity was reduced, resulting in a nonsignificant main effect of Spatial Dispersion in this setting.

This result may have important implications for understanding listening in spatial music. It shows that the bottleneck imposed on spatial attention when voices are spatially distant results in significant deterioration in the ability to detect changes in the spatial configuration away from the auditory spotlight. In this study, the detection task was a spatial one. Therefore, no safe predictions can be made with respect to the implications of the auditory spotlight hypothesis for nonspatial tasks, such as the perception of structural aspects in spatial music. Overall, very good divided attention performance in music has been measured in experiments [Bigand et al. 2000; Jones and Yee 2001; Sloboda and Edworthy 1981; Gregory 1990]. This has been attributed to the utilization of structural properties of music in the performance of the divided-attention task. The results of the study that we present here, however, raise the question of whether these observations can be replicated when the spatial dispersion of the voices in the musical piece is increased. This is an important research question for music perception and cognition that needs to be investigated in the future, as it has potential impact on listening to orchestral and spatialized electroacoustic music.

6.2 Impact of Tonal Complexity

The Tonal Complexity of the piece had a significant impact on the ability of the listeners to detect spatial displacement in the Divided Attention conditions. This is an interesting result that further highlights the use of top-down processing in the examined task. As mentioned earlier, variations in tonal height result in cues to spatial location for the individual voices that shift along the frequency axis. This poses an additional demand on cognitive resources, as the cues contributing to the localization of each voice need to be reconsidered each time the tonal height of the voices changes. Reducing the Tonal Complexity of the piece eases the difficulty with which this task is performed and, in this way, yields a performance improvement. In accordance with this observation, the benefit due to spatial attention is reduced when Tonal Complexity is decreased.

False-alarm rate in the experiments was relatively high (range between 10% and 40%) and dependent on experimental factors. This implies that it cannot be simply attributed to attentional lapses or response noise. A compelling explanation is that false alarms originate in the interference with the localization cues for the target voice caused by the distracting voices. This interference increases with the Tonal Complexity of the piece. When observing the common 45° angular displacement of the Divided Attention conditions at a descriptive level, false-alarm rate was reduced, on average, from 31% to 24% when Spatial Dispersion was reduced, to 23% when Tonal Complexity was reduced, and to 14% when both Tonal Complexity and Spatial Dispersion were reduced. Similarly, hit rate increased from 46% to 58% when Spatial Dispersion was reduced, to 64% when Tonal Complexity was reduced, and to 64% when both Spatial Dispersion and Tonal Complexity were reduced. Finally, sensitivity increased from 0.43 to 1.03 when Spatial Dispersion was reduced, to 1.25 when Tonal Complexity was reduced and to 1.62 when both Spatial Dispersion and Tonal Complexity were reduced. Interestingly, the effect of reducing Tonal Complexity and Spatial Dispersion combines to reduce false alarms, but not to increase hit rate. No further advantage in hit rate due to reducing spatial attention load was observed when Tonal Complexity was low. This may be interpreted as suggesting more specifically that, although the benefit from spatial attention in allocating target-relevant features was reduced when the complexity of the acoustic scene was reduced, spatial attention still acted to reduce interference from distracting voices in the experiments, even when Tonal Complexity was reduced.

6.3 Effect of Voice

The null effect of Voice in the isolated selective-attention condition implies that the effect is specific to the selective and divided-attention conditions with increased source numerosity. There, the advantage for V3 and V4 (the English and French horns) persisted, with small variations, irrespective of the Tonal Complexity and Spatial Dispersion manipulations. The null effect of Condition, when reversing the location of the voices, ruled out a location-specific effect for voices at the back of the listeners, which may have activated the orienting and acoustic startle reflexes [Solokov 1963; Yeomans and Frankland 1995]. The replication of the effect in the low Tonal Complexity conditions rules out a local or global influence from the melodic trajectory of the voices as a source for this effect. The observed voice effect may therefore be attributed to either a timbre- and/or register-related advantage that appears when all voices are active, perhaps related to auditory salience. The observed advantage appeared for V3 and V4, which occupied the lower registers and was strongest for the lowest register voice (V4, French horn). This could indicate binaural interference [Heller and Richards 2010; Henning 1980; McFadden and Pasanen 1976; Croghan and Grantham 2010], in which interference with the localization of a high-frequency target due to a low-frequency distracter occurs. As a rule, this effect is much weaker in the opposite direction. This effect may have been accentuated by the fact that low-frequency ITDs are most reliably reproduced by the spatialization algorithm used in the experiments, as elaborated in the next section. An alternative, but not mutually exclusive, explanation may relate to instrument-specific aspects such as variations in spectral spread or the attack time. The design of the experiments does not allow for a conclusive remark with respect to the origin of this effect.

6.4 Implications of Using a Sound Spatialization System

The use of a spatialization system in the experiment was motivated by the widespread use of such algorithms not only in contemporary music but also in other auditory display applications. It is, in general, impractical to provide a virtual auditory space without relying on virtualization. The spatialization algorithm that we used was VBAP [Pulkki 2001]. In two dimensions, this is essentially amplitude panning using the tangent law. A dense loudspeaker array was used consisting of 24 loudspeakers with a loudspeaker separation of 15°. Sounds in the experiment were displaced either directly between two loudspeakers (displacements of 30° and 60°) or between virtual locations that activated different loudspeaker pairs (45° and 80°). In the latter case, sounds were positioned roughly in the middle of the loudspeaker pair. Overall, sensitivity varied smoothly with displacement even though sounds were displaced at or in-between loudspeakers. Two points are worth discussing further.

First, the number of active loudspeakers in VBAP depends on the desired sound location and is 1 when this coincides with a loudspeaker location and 2 when it lies between two loudspeaker locations. This introduces spectral coloration. Its influence on the results of this study is arguably small, if any. This is because performance in each sound displacement was estimated independently of the rest by displacing symmetrically around their nominal position. Because of symmetry and the displacement magnitudes chosen in the experiment, sounds oscillated between positions that were similar in their location within a loudspeaker pair for each displacement; for example, both on a loudspeaker or both (roughly) at the middle of a loudspeaker pair. Consequently, spectral coloration due to the spatialization algorithm was fixed throughout the estimation procedure.

Second, in experiments using narrow-band noise signals, it was found that sound localization in VBAP is influenced by band center frequency. Under anechoic conditions, VBAP reproduces reliable low-frequency interaural-time-difference (ITD) cues below 1.1kHz and reliable high-frequency interaural-level-difference (ILD) cues above 2.6kHz. Interaural cues in between, however, are not reproduced accurately. Nevertheless, the resulting sound localization when using broadband signals, as

in this study, is accurate, as a rule [Pulkki and Karjalainen 2001]. This is because, in VBAP, as well as in the real world, low-frequency ITDs dominate sound localization judgments when listeners are presented with conflicting localization cues [Pulkki and Karjalainen 2001; Wightman and Kistler 1992]. Therefore, the use of an amplitude-panning algorithm may have acted to increase the importance of ITDs in the perception of the sound locations in the experiment and, in turn, to emphasize spatial attention influences, as these manifest themselves most strongly for ITDs in complex acoustic settings.

7. CONCLUSIONS

We investigated the perception of sound displacement in a musical scene under conditions that manipulated the effect of attentional setting (selective or divided), the number of sound sources in the musical scene, the spatial dispersion of the voices, and the tonal complexity of the piece. Our results have shown a strong effect of attentional manipulations and source numerosity. Detection under selective-attention instructions deteriorated heavily when the number of sources in the scene increased and when uncertainty was introduced with respect to which voice would move or the location at which displacement would occur. In the presence of interfering voices, detection of spatial displacement was found to depend on the voice on which displacements occur, a factor that was attributed to both the timbre and possibly the register of the displaced voice. Reducing tonal complexity further improved detection performance irrespective of the spatial dispersion of the voices. Facilitating spatial attention by restricting the spatial extent occupied by the voices improved detection performance significantly, in agreement with the auditory spotlight hypothesis. However, the magnitude of the improvement decreased when the tonal complexity of the piece was reduced. The results indicate that spatial attention facilitated auditory grouping and reduced interference from competing voices in forming localization judgments based on the available frequency-dependent interaural cues for all of the voices within the spatial focus of attention.

APPENDIX

To check for a possible room effect, detection performance for the four voices in isolation was measured in both rooms (C1, C2 in Table I). Two three-way mixed analyses of variance with Displacement and Voice as within-subjects factor and Room as a between-subjects factor were performed, comparing detection performance in Room 2 with nonmusician and musician performance in Room 1. In both cases, no significant effect of Room emerged, $F(1,27) = 1.422$, $p = 0.243$; and $F(1,27) = 0.856$, $p = 0.363$, respectively. No significant effects were observed when comparing hit and false-alarm rates with the aforementioned test. Given the null effect of Room and the similarities of the two rooms, we conclude that room had a negligible influence.

ACKNOWLEDGMENTS

We would like to thank Michel Vallières for helping us with the selection of the musical piece. Albert Bregman provided helpful discussions on this research project.

REFERENCES

K. R. Agres and C. L. Krumhansl. 2008. Musical change deafness: The inability to detect change in a non-speech auditory domain. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, B. C. Love, K. McRae, and V. M. Sloutsky (Eds.). Cognitive Science Society, Austin, TX, 969–974.

J. Ahveninen, I. P. Jääskeläinen, T. Raij, G. Bonmassar, S. Devore, M. Hämäläinen, S. Levänen, F.-H. Lin, M. Sams, B. G. Shinn-Cunningham, and others. 2006. Task-modulated "what" and "where" pathways in human auditory cortex. *Proceedings of the National Academy of Sciences* 103, 39, 14608–14613.

C. Alain, S. R. Arnott, S. Hevenor, S. Graham, and C. L. Grady. 2001. "What" and "where" in the human auditory system. *Proceedings of the National Academy of Sciences* 98, 21, 12301–12306.

K. Allen, D. Alais, and S. Carlile. 2009. Speech intelligibility reduces over distance from an attended location: Evidence for an auditory spatial gradient of attention. *Perception and Psychophysics* 71, 1, 164–173.

K. Allen, D. Alais, B. G. Shinn-Cunningham, and S. Carlile. 2011. Masker location uncertainty reveals evidence for suppression of maskers in two-talker contexts. *Journal of the Acoustical Society of America* 130, 4, 2043–2053.

T. L. Arbogast and G. Kidd. 2000. Evidence for spatial tuning in informational masking using the probe-signal method. *Journal of the Acoustical Society of America* 108, 4, 1803–1810.

Tanya L. Arbogast, Christine R. Mason, and G. Kidd. 2002. The effect of spatial separation on informational and energetic masking of speech. *Journal of the Acoustical Society of America* 112, 5, 2086–2098.

V. Best, F. J. Gallun, A. Ihlefeld, and B. G. Shinn-Cunningham. 2006. The influence of spatial separation on divided listening. *Journal of the Acoustical Society of America* 120, 1506–1516.

E. Bigand, S. Forêt, and S. McAdams. 2000. Divided attention in music. *International Journal of Psychology* 35, 6, 270–278.

J. K. Bizley and Y. E. Cohen. 2013. The what, where and how of auditory-object perception. *Nature Reviews Neuroscience* 14, 10, 693–707.

A. S. Bregman. 1990. *Auditory Scene Analysis*. MIT Press, Cambridge, MA.

A. W. Bronkhorst and R. Plomp. 1988. The effect of head-induced interaural time and level differences on speech intelligibility in noise. *Journal of the Acoustical Society of America* 83, 4, 1508–1516.

A. W. Bronkhorst and R. Plomp. 1992. Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing. *Journal of the Acoustical Society of America* 92, 6 , 3132–3139.

D. S. Brungart, B. D. Simpson, and R. L. Freyman. 2005. Precedence-based speech segregation in a virtual auditory environment. *Journal of the Acoustical Society of America* 118, 5, 3241–3251.

D. W. Chandler, D. Grantham, and M. R. Leek. 2005. Effects of uncertainty on auditory spatial resolution in the horizontal plane. *Acta Acustica United with Acustica* 91, 3, 513–525.

C. Cherry. 1953. Some experiments on the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America* 25, 975–979.

C. Cherry and W. K. Taylor. 1953. Some further experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America* 26, 5, 554–559.

E. J. Crawley, B. E. Acker-Mills, R. E. Pastore, and S. Weil. 2002. Change detection in multi-voice music: The role of musical structure, musical training and task demands. *Journal of Experimental Psychology: Human Perception and Performance* 28, 2, 367–378.

N. B. H. Croghan and D. W. Grantham. 2010. Binaural interference in the free field. *Journal of the Acoustical Society of America* 127, 5, 3085–3091.

John F. Culling and C. J. Darwin. 1993. Perceptual separation of simultaneous vowels: Within and across-formant grouping by F0. *Journal of the Acoustical Society of America* 93, 6, 3454–3467.

J. F. Culling and Q. Summerfield. 1995. Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay. *Journal of the Acoustical Society of America* 98, 2, 785–797.

C. J. Darwin. 2008. Spatial hearing and perceiving sources. In *Auditory Perception of Sound Sources*. Springer, 215–232.

C. J. Darwin and R. W. Hukin. 1997. Perceptual segregation of a harmonic from a vowel by interaural time difference and frequency proximity. *Journal of the Acoustical Society of America* 102, 4, 2316–2324.

C. J. Darwin and R. W. Hukin. 1999. Auditory objects of attention: The role of interaural time differences. *Journal of Experimental Psychology: Human Perception and Performance* 25, 617–629.

D. Deutsch. 1979. Binaural integration of melodic patterns. *Perception and Psychophysics* 25, 5, 399–405.

C. W. Eriksen and J. D. St. James. 1986. Visual attention within and around the field of focal attention: A zoom lens model. *Perception and Psychophysics* 40, 4, 225–240.

W. L. Fan, T. M. Streeter, and N. I. Durlach. 2008. Effect of spatial uncertainty of masker on masked detection for nonspeech stimuli. *Journal of the Acoustical Society of America* 124, 1, 36–39.

Angelo Farina and Lamberto Tronchin. 2013. 3D sound characterisation in theatres employing microphone arrays. *Acta Acustica United with Acustica* 99, 1, 118–125.

R. L. Freyman, U. Balakrishnan, and K. S. Helfer. 2001. Spatial release from informational masking in speech recognition. *Journal of the Acoustical Society of America* 109, 5, 2112–2122.

R. L. Freyman, K. S. Helfer, D. D. McCall, and R. K. Clifton. 1999. The role of perceived spatial separation in the unmasking of speech. *Journal of the Acoustical Society of America* 106, 6, 3578–3588.

F. J. Gallun, N. I. Durlach, H. S. Colburn, B. G. Shinn-Cunningham, V. Best, C. R. Mason, and G. Kidd Jr. 2008. The extent to which a position-based explanation accounts for binaural release from informational masking. *Journal of the Acoustical Society of America* 124, 1, 439–449.

F. J. Gallun, C. R. Mason, and G. Kidd. 2005. Binaural release from informational masking in a speech identification task. *Journal of the Acoustical Society of America* 118, 3, 1614–1625.

S. Gatehouse and M. A. Akeroyd. 2008. The effects of cueing temporal and spatial attention on word recognition in a complex listening task in hearing-impaired listeners. *Trends in Amplification* 12, 2, 145–161.

D. Grantham. 1995. Spatial hearing and related phenomena. In *Hearing*, B. C. J. Moore (Ed.). Academic Press, San Diego, CA, 297–345.

A. H. Gregory. 1990. Listening to polyphonic music. *Psychology of Music* 18. 163–170.

M. A. Harley. 1994. *Space and Spatialization in Contemporary Music*. Ph.D. Dissertation. Schulich School of Music, Montreal, Quebec.

W. M. Hartmann and D. Johnson. 1991. Stream segregation and peripheral channeling. *Music Perception* 9, 2, 155–183.

W. M. Hartmann and B. Rakerd. 1989. On the minimum audible angle – A decision theory approach. *Journal of the Acoustical Society of America* 85, 5, 2031–2041.

M. L. Hawley, R. Y. Litovsky, and H. S. Colburn. 1999. Speech intelligibility and localization in a multi-source environment. *The Journal of the Acoustical Society of America* 105, 6 (1999), 3436–3448.

M. L. Hawley, R. Y. Litovsky, and J. F. Culling. 2004. The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *Journal of the Acoustical Society of America* 115, 2, 833–843.

L. M. Heller and V. M. Richards. 2010. Binaural interference in lateralization thresholds for interaural time and level differences. *Journal of the Acoustical Society of America* 128, 1, 310–319.

B. Henning. 1980. Some observations on the lateralization of complex waveforms. *Journal of the Acoustical Society of America* 68, 2, 446–454.

R. W. Hukin and C. J. Darwin. 1995. Effects of contralateral presentation and of interaural time differences in segregating a harmonic from a vowel. *Journal of the Acoustical Society of America* 98, 3, 1380–1387.

A. Ihlefeld and B. G. Shinn-Cunningham. 2008a. Spatial release from energetic and informational masking in a divided speech identification task. *Journal of the Acoustical Society of America* 123, 6, 4380–4392.

A. Ihlefeld and B. G. Shinn-Cunningham. 2008b. Spatial release from energetic and informational masking in a selective speech identification task. *Journal of the Acoustical Society of America* 123, 6, 4369–4379.

P. Janata, B. Tillmann, and J. J. Bharucha. 2002. Listening to polyphonic music recruits domain-general attention and working memory circuits. *Cognitive, Affective, and Behavioral Neuroscience* 2, 2, 121–140.

G. L. Jones and R. Y. Litovsky. 2008. Role of masker predictability in the cocktail party problem. *Journal of the Acoustical Society of America* 124, 6, 3818–3830.

M. R. Jones and W. Yee. 2001. Attending to Auditory Events: The role of temporal organisation. In *Thinking in Sound: The Cognitive Psychology of Human Audition*, S. McAdams and E. Bigand (Eds.). Oxford University Press, New York, NY, 69–112.

G. Kidd, T. L. Arbogast, C. R. Mason, and F. J. Gallun. 2005a. The advantage of knowing where to listen. *Journal of the Acoustical Society of America* 118, 6, 3804–3815.

G. Kidd, C. R. Mason, A. Brughera, and W. M. Hartmann. 2005b. The role of reverberation in release from masking due to spatial separation of sources for speech identification. *Acta Acustica United with Acustica* 91, 3, 526–536.

G. Kidd, C. R. Mason, T. L. Rohtla, and P. S. Deliwala. 1998. Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. *Journal of the Acoustical Society of America* 104, 1, 422–431.

P. T. Kitterick, P. J. Bailey, and A. Q. Summerfield. 2010. Benefits of knowing who, where, and when in multi-talker listening. *Journal of the Acoustical Society of America* 127, 4 (2010), 2498–2508.

E. Lindemann. 2007. Music synthesis with reconstructive phrase modeling. *IEEE Signal Processing Magazine* 24, 2, 80–91.

N. A. Macmillan and C. D. Creelman. 2005. *Detection Theory: A User's Guide*. Lawrence Erlbaum Associates, Inc., Mahwah, NJ.

P. P. Maeder, R. A. Meuli, M. Adriani, A. Bellmann, E. Fornari, A. Pittet, and S. Clarke. 2001. Distinct pathways involved in sound recognition and localization: A human fMRI study. *Neuroimage* 14, 802–816.

D. McFadden and E. G. Pasanen. 1976. Lateralization at high frequencies based on interaural time differences. *Journal of the Acoustical Society of America* 59, 3, 634–639.

A. W. Mills. 1958. On the minimum audible angle. *Journal of the Acoustical Society of America* 30, 4, 237–246.

T. A. Mondor and R. J. Zatorre. 1995. Shifting and focusing auditory spatial attention. *Journal of Experimental Psychology: Human Perception and Performance* 21, 2, 387–409.

T. A. Mondor, R. J. Zatorre, and N. A. Terrio. 1998. Constraints on the selection of auditory information. *Journal of Experimental Psychology: Human Perception and Performance* 24, 1, 66–79.

A. J. Oxenham, B. J. Fligor, C. R. Mason, and G. Kidd. 2003. Informational masking and musical training. *Journal of the Acoustical Society of America* 114, 3, 1543–1549.

V. Pulkki. 2001. *Spatial Sound Generation and Perception by Amplitude Panning Techniques*. Ph.D. Dissertation. Helsinki University of Technology, Helsinki, Finland.

V. Pulkki and M. Karjalainen. 2001. Localization of amplitude-panned virtual sources I: Stereophonic panning. *Journal of the Audio Engineering Society* 49, 9, 739–752.

R. Reynolds. 2002. *Form and Method: Composing Music (The Rothschild Essays)*. Routledge, New York, NY.

K. Saberi, L. Dostal, T. Sadralodabai, and D. Perrott. 1991. Minimum audible angles for horizontal, vertical and oblique orientations: Lateral and dorsal planes. *Acustica* 75, 57–61.

A. J. Sach and P. J. Bailey. 2004. Some characteristics of auditory spatial attention revealed using rhythmic masking release. *Perception and Psychophysics* 66, 8, 1379–1387.

A. J. Sach, N. I. Hill, and P. J. Bailey. 2000. Auditory spatial attention using interaural time differences. *Journal of Experimental Psychology: Human Perception and Performance* 26, 2, 717–729.

A. Samuel. 2002. *The Study of Orchestration*. W. W. Norton and Company, Inc, New York, London.

K. Saupe, S. Koelsch, and R. Rübsamen. 2010. Spatial selective attention in a complex auditory environment such as polyphonic music. *Journal of the Acoustical Society of America* 127, 1, 472–480.

B. G. Shinn-Cunningham. 2008. Object-based auditory and visual attention. *Trends in Cognitive Sciences* 12, 5, 182–186.

B. G. Shinn-Cunningham, A. Ihlefeld, Satyavarta, and E. Larson. 2005. Bottom-up and top-down influences on spatial unmasking. *Acta Acustica United with Acustica* 91, 6, 967–979.

J. Sloboda and J. Edworthy. 1981. Attending to two melodies at once: The effect of key relatedness. *Psychology of Music* 9, 39–43.

E. N. Solokov. 1963. Higher nervous functions: The orienting reflex. *Annual Review of Physiology* 25, 1, 545–580.

Charles Spence and Jon Driver. 2004. *Crossmodal Space and Crossmodal Attention*. Oxford University Press, Oxford, UK.

T. H. Stainsby, C. Füllgrabe, H. J. Flanagan, S. K. Waldman, and B. C. J. Moore. 2011. Sequential streaming due to manipulation of interaural time differences. *Journal of the Acoustical Society of America* 130, 2, 904–914.

W. A. Teder-Sälejärvi and S. A. Hillyard. 1998. The gradient of spatial auditory attention in free field: An event-related potential study. *Perception and Psychophysics* 60, 7, 1228–1242.

F. L. Wightman and D. J. Kistler. 1992. The dominant role of low-frequency interaural time differences in sound localization. *Journal of the Acoustical Society of America* 91, 3, 1648–1661.

D. L. Woods, C. Alain, R. Diaz, D. Rhodes, and K. H. Ogawa. 2001. Location and frequency cues in auditory selective attention. *Journal of Experimental Psychology: Human Perception and Performance* 27, 1, 65–74.

J. S. Yeomans and P. W. Frankland. 1995. The acoustic startle reflex: Neurons and connections. *Brain Research Reviews* 21, 3, 301–314.

W. A. Yost, R. H. Dye Jr, and S. Sheft. 1996. A simulated "cocktail party" with up to three sound sources. *Perception and Psychophysics* 58, 7, 1026–1036.

Received October 2015; revised March 2016; accepted April 2016