The spatial unmasking of speech: Evidence for better-ear listening

Barrie A. Edmonds and John F. Culling
School of Psychology, Cardiff University, Tower Building, Park Place, Cardiff, CF10 3AT, United Kingdom

(Received 12 January 2006; revised 31 May 2006; accepted 23 June 2006)

Speech reception thresholds (SRTs) were measured for target speech presented concurrently with interfering speech (spoken by a different speaker). In experiment 1, the target and interferer were divided spectrally into high- and low-frequency bands and presented over headphones in three conditions: monaural, dichotic (target and interferer to different ears), and swapped (the low-frequency target band and the high-frequency interferer band were presented to one ear, while the high-frequency target band and the low-frequency interferer band were presented to the other ear). SRTs were highest in the monaural condition and lowest in the dichotic condition; SRTs in the swapped condition were intermediate. In experiment 2, two new conditions were devised such that one target band was presented in isolation to one ear while the other band was presented at the other ear with the interferer. The pattern of SRTs observed in experiment 2 suggests that performance in the swapped condition reflects the intelligibility of the target frequency bands at just one ear; the auditory system appears unable to exploit advantageous target-to-interferer ratios at different ears when segregating target speech from a competing speech interferer. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2228573]

PACS number(s): 43.66.Rq, 43.66.Ba, 43.66.Pn [GDK] Pages: 1539–1545

I. INTRODUCTION

Listeners often experience the compelling impression that they can pick up any of a range of voices around them by focusing their attention on the appropriate direction (Cherry, 1953). This introspection is supported by the fact that listeners achieve better recognition of concurrent voices when these voices come from different directions. The phenomenon, known as spatial unmasking, can be measured as the improvement in the masked threshold for the detection or identification of a target sound as a function of the angle of separation, in the horizontal plane, between the target and an interfering sound.

The perceived direction of a sound source in the horizontal plane is largely determined by two binaural cues (Rayleigh, 1876; 1907). Interaural level difference (ILD) refers to the difference in sound level at the two ears caused by different distances to the source (i.e., the inverse-square law) and, at high frequencies, by the occluding effect of the head (i.e., head shadow). Interaural time difference (ITD) refers to the brief time lag in the signal at the more distant ear caused by the finite speed of sound. ITD is the dominant cue to sound direction (Wightman and Kistler, 1992). These cues also contribute to the spatial unmasking of speech (Bronkhorst and Plomp, 1988). However, the fact that sound localization and spatial unmasking exploit the same cues need not imply a common mechanism. Rather, it is thought that the auditory system exploits ITD and ILD independently to bring about monaural and binaural improvements in the target-to-interferer ratio (Bronkhorst and Plomp, 1988; Culling *et al.*, 2006; 2004; Edmonds and Culling, 2005a; Hawley *et al.*, 2004; Zurek, 1993). ILD is mainly associated with monaural gains in audibility at the ear with the better target-to-interferer ratio, while ITD gives rise to improvements in the audibility of a target sound through binaural interaction (Colburn, 1977; Durlach, 1972; Jeffress, 1948; Levitt and Rabiner, 1967; Rabiner *et al.*, 1966).

Zurek's (1993) model of spatial unmasking predicts the effect of spatial separation on speech intelligibility by calculating the target-to-interferer ratio due to head shadow and binaural unmasking across a series of 1/3-octave frequency bands and comparing these estimates against the articulation index (French and Steinberg, 1947). Here, we focus on how Zurek's (1993) model estimates the monaural contribution to speech intelligibility. The total monaural estimate of the model can be defined as either the contribution of the monaural frequency bands providing the better target-to-interferer ratios regardless of ear (which Zurek termed the "better-bands" rule) or the contribution of the monaural frequency bands at the ear providing the better overall target-to-interferer ratio (which Zurek termed the "better-ear" rule). The predicted difference in speech intelligibility for the better-bands and better-ear rules is negligible in normal listening situations, as head shadow will tend to cause one ear to have the better target-to-interferer ratio in all frequency bands. The model is agnostic as to whether the auditory system employs a better-bands rule or a better-ear rule. Nonetheless, the better-bands rule assumes that the human auditory system is free to select monaural information independently from frequency channel to frequency channel across the two ears. However, the authors are unaware of any attempts to empirically test between the predicted effects of the better-ear and better-bands rules on speech intelligibility.

The experiments reported here compare the predictions of the better-bands and better-ear rules by exploring the effect of splitting stimuli spectrally across the two ears using a

paradigm similar to the one described by Edmonds and Culling (2005b). High- and low-frequency bands of target and interfering speech materials were presented using infinite ILDs (i.e., each frequency band of each signal was presented to one ear only) in order to maximize the target-to-interferer ratio across the ears. We reasoned that if listeners really can exploit the better-bands rule then it should be possible to present the high- and low-frequency bands of a target utterance to different ears in the presence of a concurrent interferer without observing any impact on speech intelligibility. However, if listeners are only able to attend to the ear providing the best overall target-to-interferer ratio (i.e., they employ the better-ear rule), then performance could be poorer in this condition than when target and interferer are simply presented to different ears.

II. EXPERIMENT 1

A. Participants

Nine Cardiff University undergraduate students were recruited and awarded course credit in return for their participation. All participants reported normal hearing and spoke English as their first language.

B. Stimuli

Sentences from the MIT recordings of the male speaker CW reading the Harvard Sentence Lists (IEEE, 1969) were used as target items and sentences from the speaker DA (another male speaker from MIT recordings of the Harvard sentence lists) were used as interferer items. These sentence materials were filtered into high- and low-frequency bands using a 512-point, linear-phase, finite impulse response (FIR) filter with very steep (>1000-dB/octave) cutoffs. High- and low-frequency cutoffs spanned a splitting frequency, leaving a spectral notch of 1 equivalent rectangular bandwidth (ERB) (Moore and Glasberg, 1983). These spectral notches were the same as those employed by Edmonds and Culling (2005b): 700–802 Hz (i.e., a splitting frequency of *750 Hz*), 1411–1594 Hz (i.e., a splitting frequency of *1500 Hz*), and 2822–3287 Hz (i.e., a splitting frequency of *3000 Hz*)[1]. These splitting-frequency manipulations allowed the high- and low-frequency regions of the speech materials to be presented in different interaural configurations. Three interaural configurations of target and interferer were used: *monaural*, *dichotic*, and *swapped*. In the monaural configuration the target speech and interferer were both presented to one ear. In the dichotic configuration the high- and low-frequency bands of a target sentence were presented to the left ear and the high- and low-frequency bands of an interfering sentence were presented to the right ear. Last, in the swapped configuration a high-frequency target band and a low-frequency interferer band were presented to the left ear while a low-frequency target band and a high-frequency interferer band were presented to the right ear. Thus, a total of nine conditions (3 interaural configurations × 3 splitting frequencies) was used in this experiment.

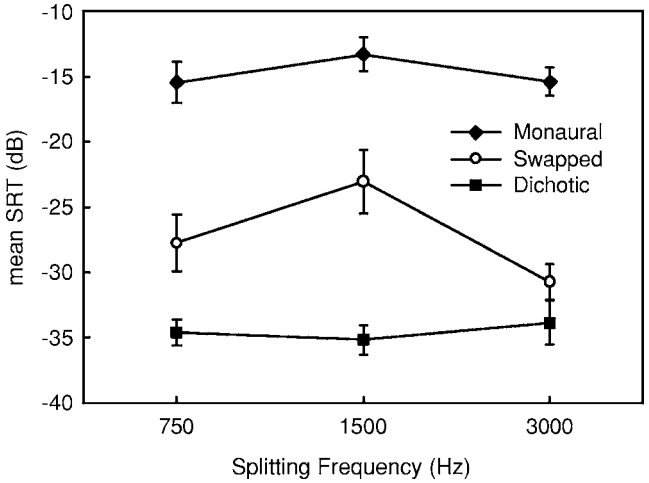

FIG. 1. Mean SRTs of the monaural (diamonds), dichotic (squares), and swapped (circles) conditions as a function of splitting frequency (750, 1500, and 3000 Hz). Error bars show standard error.

C. Procedure

Stimuli were presented to the listener using a TDT AP2 array processor via a TDT psychoacoustics rig (DD1, FT6, PA4, HB6) over Sennheiser HD 590 headphones in a single-walled IAC sound-attenuating booth. Speech reception thresholds (SRTs) were measured using a 1-up/1-down adaptive threshold method (Levitt, 1971) in which a set of ten target sentences was presented against a fixed-level (~70-dB) interferer sentence (a novel interfering sentence was used for each SRT measurement). The initial target-to-interferer ratio was −28 dB. The first stimulus was repeated, each time with a 4-dB-more-intense target sentence, until the listener judged that they could understand half of the sentence and attempted to transcribe it on a computer terminal. When the listener pressed "return" at the end of their transcription, the actual target sentence appeared below the listener's transcript with five keywords in capitals. The listener self-marked his or her identification of the five keywords. If more than two keywords were correctly identified, the level of the subsequent target sentence was reduced by 2 dB. Otherwise it was increased by 2 dB. The mean of the last eight target-to-interferer ratios derived in this way was taken as the SRT.

D. Results and discussion

Figure 1 shows the mean SRTs for the listeners in experiment 1. SRTs were highest (i.e., worst performance) in the monaural condition, lowest (i.e., best performance) in the dichotic condition, and of intermediate level in the swapped condition. SRTs in the swapped condition appeared to vary with splitting frequency, producing a chevron-like pattern of SRTs (i.e., the SRTs were lower for this condition at the 750 and 3000-Hz splitting frequencies than they were when this interaural configuration was presented with a 1500-Hz splitting frequency).

A two-way repeated-measures analysis of variance was performed on the SRTs for all nine conditions (3 interaural configurations × 3 splitting frequencies). There was a significant main effect of interaural configuration [$F(2,16)$

$=114.71$, $p<0.001$] and splitting frequency [$F(2,16)=3.82$, $p<0.05$]. There were no other significant F ratios. There was no significant interaction between interaural configuration and splitting frequency [$F(4,32)=21.20$, $p=0.088$]. Tukey *posthoc* comparisons of interaural configuration confirmed that all three conditions differed significantly from each other [$q>7$, $p<0.001$]. *Posthoc* analysis of splitting frequency revealed that only SRTs in the 1500 and 3000-Hz conditions were significantly different [$q=3.75$, $p<0.05$].

If the monaural advantage described by Zurek (1993) is due to the selection of information using the better-bands rule (i.e., the frequency channel at either ear providing the most favorable target-to-interferer ratio) rather than the better-ear rule (i.e., the ear providing the best target-to-interferer ratio over all frequency channels), then one would expect there to be no difference in the SRTs measured for the dichotic and swapped conditions. However, there was a significant difference between the SRTs of the dichotic and swapped conditions; this result suggested that listeners exploit the better-ear rule rather than the better-bands rule.

Although there was no significant interaction between splitting frequency and interaural configuration in this study, we noticed a chevron-shaped pattern of SRTs in the swapped condition, which suggests that listeners were particularly disadvantaged by the 1500-Hz splitting frequency in this condition. The fact that performance improves when a larger proportion of information is presented to one ear or the other suggests that SRTs in the swapped condition were determined solely by the target information present in the frequency channels at the ear providing the most information about the target. As the 1500-Hz splitting frequency removes target speech that is roughly in the middle of the speech range in informational terms (Fletcher and Galt, 1950), neither ear provides the listener with sufficient information to achieve optimal SRTs. We decided to further investigate the apparent effect of splitting frequency on the intelligibility of speech in the swapped condition in a second experiment.

III. EXPERIMENT 2

This experiment explored the contribution of high- and low-frequency speech bands to speech intelligibility. It was conducted to confirm whether or not listeners combine any information about the target that is presented to different ears or whether they rely solely on the information available at one ear. Two new conditions were designed to yield SRTs that varied systematically with splitting frequency. One ear was designated a "target-only" ear and received either a high-frequency target band (high-contribution condition) or a low-frequency target band (low-contribution condition); the proportion of target speech at the target-only ear in the high-contribution and low-contribution conditions depended on the splitting frequency. The remaining proportion of target speech (e.g., the low-frequency band in the high-contribution condition) was presented to the contralateral ear with the high- and low-frequency interferer bands. We reasoned that the intelligibility of the target speech in these conditions should increase (i.e., yield lower SRTs) as the range of frequencies available at the target-only ear increased. Consequently, if the better-ear rule is employed then one would expect the pattern of SRTs in the swapped condition to closely match the pattern of SRTs found in one or the other of these new conditions at different splitting frequencies. For example, the target-to-interferer ratio at the better ear in the swapped condition might mirror those of the low-contribution condition when the splitting frequency is high (i.e., 3000 Hz) and those of the high-contribution condition when the splitting frequency is low (i.e., 750 Hz). However, if listeners have any ability to exploit the better-bands rule rather than the better-ear rule, then performance in the swapped condition should be better than that achieved for these high-contribution and low-contribution conditions, as more frequency bands will be available to the listener.

A. Participants

A new group of nine undergraduate students from Cardiff University was recruited and rewarded with course credit for their participation. Again, all participants reported normal hearing and spoke English as their first language.

B. Stimuli

The target and interfering speech materials were again presented as high- and low-frequency bands divided at splitting frequencies of *750*, *1500*, and *3000 Hz* in three interaural configurations: *swapped* (the left ear received a high-frequency target band and a low-frequency interferer band; the right ear received a low-frequency target band and a high-frequency interferer band), *high-contribution* (a high-frequency target band was presented to the left ear; the high- and low-frequency interferer bands and a low-frequency target band were presented to the right ear), and *low-contribution* (a low-frequency target band was presented to the right ear; the high- and low-frequency interferer bands and a high-frequency target band were presented to the left ear).

C. Results and discussion

Figure 2 shows how the swapped-condition thresholds relate to the high-contribution and low-contribution condition SRTs. These thresholds vary systematically with increasing splitting frequency in all three conditions; again, a chevron-like pattern of SRTs was observed in the swapped condition. When a splitting frequency of 750 Hz was employed the low-contribution condition SRTs were highest (i.e., intelligibility was poorest), while SRTs in the high-contribution and the swapped conditions were lowest. When a splitting frequency of 1500 Hz was used, SRTs in all three conditions were intermediate and indistinguishable. When the 3000-Hz splitting frequency was employed, SRTs in the low-contribution and swapped conditions were lowest, while those of the high-contribution condition were highest.

A two-way repeated measures analysis of variance was performed on the SRTs, with two within-subjects factors (3 interaural configurations $\times 3$ splitting frequencies). There was a significant main effect of interaural configuration

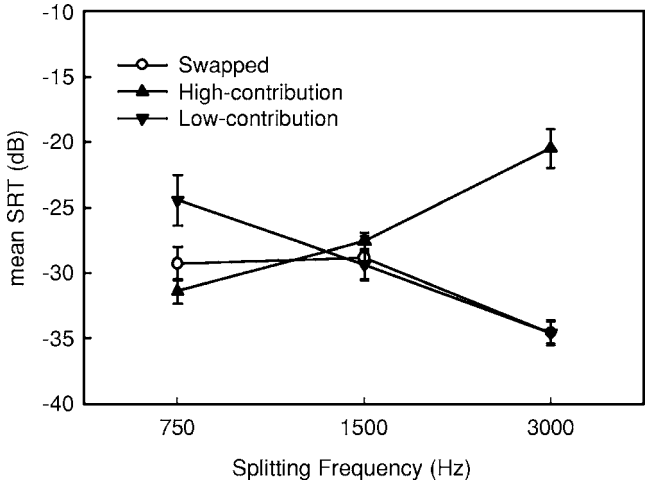

FIG. 2. Mean SRTs of the swapped (circles), low-contribution (upright triangles), and high-contribution (inverted triangles) conditions as a function of splitting frequency (750, 1500, and 3000 Hz). Error bars show standard error.

$[F(2,16)=9.88, p<0.001]$ and a significant interaction between interaural configuration and splitting frequency $[F(4,32)=22.24, p<0.001]$.

Tukey *posthoc* tests for the interaction between interaural configuration and splitting frequency revealed the following differences. For comparisons of interaural configuration within the 750-Hz level of splitting frequency, the low-contribution condition SRTs were significantly different from both the high-contribution condition $[q=5.7, p<0.001]$ and swapped-condition $[q=3.97, p<0.05]$ SRTs; there was no significant difference between the SRTs of the swapped and high-contribution conditions $[p>0.05]$ at this splitting frequency. Comparisons of interaural configuration within the 1500-Hz level of splitting frequency revealed no significant differences between the SRTs of the three interaural configurations. For comparisons of interaural configuration within the 3000-Hz level of splitting frequency, the mean SRT of the high-contribution condition was significantly different from both the swapped and the low-contribution condition SRTs $[q>11.58, p<0.001]$; the SRTs of the swapped and low-contribution conditions were not significantly different $[p>0.05]$ at this splitting frequency.

Comparisons of splitting frequency within interaural configuration revealed the following differences. SRTs for the swapped condition at a splitting frequency of 3000 Hz were significantly different from those at 1500 and 750 Hz $[q>4.3, p<0.05]$. SRTs for the high-contribution condition at a splitting frequency of 3000 Hz were significantly different from those at 1500 and 750 Hz $[q>5.8, p<0.001]$. SRTs for the low-contribution condition at all three splitting frequencies were significantly different from one another $[q>4.0, p<0.05]$.

The overall similarity of mean SRTs in the swapped condition to those observed in the high-contribution and low-contribution conditions at different splitting frequencies suggests that the intelligibility of speech in the swapped condition was dependent upon the frequency bands at the ear providing the better target-to-interferer ratio. We conclude that this effect is consistent with the better-ear rule of listening and not the better-bands rule.

It is tempting to suppose that the high thresholds observed in the low-contribution condition with 750-Hz splitting frequency and the high-contribution condition with 3000-Hz cutoffs may reflect performance based solely on those narrow frequency regions. However, given the restricted range of frequencies available, it is quite possible that the better ear was, in each case, the one with the interfering voice and that these thresholds represent masked thresholds based on a relatively wide frequency band.

IV. GENERAL DISCUSSION

Zurek's (1993) model of spatial unmasking describes how speech intelligibility is dependent on monaural and binaural contributions to the target-to-interferer ratio across a series of frequency bands. We tested, in two experiments, whether the monaural advantage to spatial unmasking reflects an all-or-nothing strategy for exploiting information at one ear or the other, or whether the auditory system is able to take advantage of the ear with the better target-to-interferer ratio within each frequency channel. These experiments were designed as a simple test of the two strategies (i.e., better ear or better bands) that Zurek (1993) described in his model. In experiment 1, speech intelligibility was better in the dichotic condition than in the swapped condition. This result is inconsistent with the better-bands rule, which would predict that thresholds in these two conditions should be identical. Experiment 2 showed that performance in the swapped condition was virtually indistinguishable from the best thresholds observed in the high-contribution and low-contribution conditions at a given splitting frequency; this suggests that speech intelligibility in the swapped condition was dependent on the target frequency channels at just one ear. Consequently, we suggest, these data indicate clear support for a better-ear interpretation of Zurek's model. While this conclusion is not particularly novel (i.e., it only confirms a prediction of Zurek's model), it is most definitely surprising, as a number of related studies suggest that the better-bands rule should have prevailed.

Brungart *et al.* (2005), for instance, used the Zurek's better-bands rule to explain some unexpected effects in their study of precedence-based segregation (see Sec. IV C for further details). They manipulated the extent to which the "echo" of a masking stimulus was delayed and found that this manipulation brought about an improvement in the intelligibility of target speech heard against broadband noise at some delays and not others. After an inspection of their stimuli, they concluded that the addition of a delayed masker at certain delay intervals introduced spectral notches in the spectrum of the masker, which allowed listeners to exploit advantages in the target-to-interferer ratio at the frequencies of these notches. As these notches appeared in different frequency regions in both ears, Brungart *et al.* reasoned that, in order for listeners to take full advantage of these notches, the auditory system might exploit information at the ear providing the most favorable target-to-interferer ratio within each

frequency band (i.e., the better-bands rule). A simulation of Zurek's (1993) model of spatial unmasking confirmed that these gains in intelligibility, afforded by the changes in target-to-interferer ratio at these notches, was consistent with the increases in performance observed for these conditions. It would be interesting to know if the better-ear rule would also have been able to account for the changes in performance that were observed by Brungart *et al.* (2005), as the results of the current investigation (against speech interference) suggest that listeners are only able to exploit notches in the masker spectrum at only one of the ears. Unfortunately, Brungart *et al.* (2005) did not report whether this was the case or not. Moreover, our own data were collected using speech as an interferer, rather than the speech-shaped noise used by Brungart *et al.* Below, we consider how the results of the current dichotic listening experiments relate to other bodies of research in the literature.

A. Band-independent processing of ITD

The idea that the auditory system is able to select and integrate information from different frequency channels has been successfully demonstrated binaurally (Akeroyd, 2004; Culling and Summerfield, 1995; Edmonds and Culling, 2005b). It is somewhat intriguing, therefore, that the results of the current investigation suggest that the auditory system does not employ a similar strategy for monaurally exploiting target-to-interferer ratios. Previously, in an analogous set of experiments to the ones described here, Edmonds and Culling (2005b) looked for evidence of across-channel processing in the spatial unmasking of speech using ITD. Three experiments were designed to test whether the segregation of spatially separated sounds is dependent on the consistency of ITD across different frequency bands, in particular, whether or not the binaural gain in speech intelligibility was constrained to the exploitation of a single ITD across frequency. It was found that, as long as the target and interferer had a different ITD in each frequency channel, speech intelligibility was unaffected. As integration of ITD across frequency is important for localizing sounds, this result suggested that improvements in speech intelligibility for these stimuli reflected the benefit of binaural unmasking and that perceived location was a relatively unimportant cue. The fact that the monaural and binaural components of spatial unmasking display such strikingly different (i.e., better-ear vs band-independent processing) characteristics suggests that the mechanisms underlying the exploitation of ITD and ILD might be separate and distinct.

B. The spectral fusion of sounds across the ears

Early work in the field of dichotic listening (e.g., Broadbent, 1955; Broadbent and Ladefoged, 1957; Cutting, 1976) suggests that listeners can fuse together different spectral regions of a sound source when they are presented separately but simultaneously to opposite ears to form a single auditory image. Broadbent and Ladefoged (1957), for example, presented listeners with a dichotic speech stimulus in which the first and second formants of a synthesized sentence were simultaneously presented to different ears. They found that as long as the two formants shared the same fundamental frequency ($F0$), listeners would describe hearing a single voice in a single location just as they might report for an unmodified stimulus.

While the nature of these spectral fusion experiments is quite different from those of the current investigation, a number of parallels can be drawn. Both sets of experiments require the listener to piece together different parts of a speech stimulus that has been spectrally divided across the ears. In the spectral fusion experiments described above, the speech stimulus is split at a point dividing the first and second formants. Here, a specific splitting frequency divides the speech into two halves; although this manipulation does not take into account the formant structure of the stimuli, $F0$ information is preserved. Thus, although the target speech is spectrally split across the two ears both frequency regions of the target share the same $F0$. Consequently, one might expect spectral fusion to occur for the swapped-target stimuli in the current paper; if this was the case, SRTs should reflect contribution of all target frequency bands and not just those at one ear. Furthermore, predictions about the effect of spatial cues could be based on differences between the perceived locations of the two fused auditory images.

We can make a number of different predictions about how listeners might have performed, in the swapped condition of the current paper, if spectral fusion of swapped speech bands occurred. First, spectral fusion across the two ears might cause listeners to perceive the target and interferer as having the same location (e.g., in the center of the head). If this was the case, thresholds might match those observed in the baseline condition. Second, the fused bands of target speech might be perceived to have a slightly different location from the interferer (e.g., both heard slightly off-center, but in opposite hemispheres). Third, spectral fusion across the ears might result in the target and interferer being perceived to have distinctly different locations (e.g., target at one ear and interferer at the opposite ear). If this was the case, the competing voices should be easy to segregate and thresholds should be similar to those in the dichotic condition. The data are obviously inconsistent with the first and third scenarios, as thresholds in the swapped condition fell somewhere in between those achieved for the baseline and dichotic conditions. The second scenario sounds appealing, at least until one attempts to explain the data of experiment 2 using the same scenario. Using this scenario, the target and interferer of the high- and low-contribution conditions in experiment 2 should be perceived by listeners to be well separated (i.e., the target should be perceived to be off-center and the interferer at the opposite ear). This spatial separation, however, does not provide any additional benefit to the listener; the thresholds of the high- and low-contribution conditions accurately predict performance in the swapped condition. This suggests that the benefit of spectral fusion observed in "quiet" does not extend to more complex listening situations (e.g., involving the segregation of target speech from a concurrent speech interferer), and perhaps indicates a problem with grouping[2] by fundamental frequency ($F0$) across the ears in the swapped condition.

The possibility that failure to group by $F0$ may underlie

the pattern of results would be consistent with the findings of Culling and Darwin (1993). They found that correct identification rates for pairs of simultaneous vowels with different *F*0's was largely unaffected by a manipulation in which the *F*0's of the two vowels were swapped between the first and second formants. They thus found little evidence for across-formant grouping by *F*0. Bird and Darwin (1998) later showed that a similar manipulation of competing sentences was ineffective at disrupting their intelligibility. These experiments both attempted to disrupt performance by confusing the grouping process. The present experiment may offer a complementary form of evidence because, in this case, a lack of across formant grouping by *F*0 may have impaired listeners' ability to perform a task in which such grouping was needed.

C. Spatial release from informational masking

A number of studies in the spatial unmasking literature suggest that attending to perceived location can be particularly important for providing a release from informational masking (Brungart, 2001; Brungart *et al.*, 2001; 2005; Freyman *et al.*, 2001; 2004; 1999; Kidd *et al.*, 2005; 1994). Informational masking is thought to reflect a failure to segregate or a propensity to confuse sounds (especially speech sounds); it is thought to produce an excess of masking (i.e., a rise in the threshold for identifying a target over and above that attributable to energetic masking). The impact of informational masking on the segregation of sound sources can be avoided, however, if the listener is provided with some means to focus their attention on the target item. Freyman *et al.* (1999), for example, demonstrated how perceived location can help listeners focus their attention on the target utterance and obtain a release from the informational masking. In Freyman's study, an illusory spatial separation between target and masker (presented over spatially separated loudspeakers) was achieved with the addition a delayed copy of the masker (i.e., to simulate a room reflection) at the target loudspeaker. The "echo" invoked the precedence effect (Litovsky *et al.*, 1999; Zurek, 1987), which creates the illusion of spatial separation while substantially disrupting the cues for binaural unmasking. Under these circumstances, an improvement in speech intelligibility against speech maskers was observed but not against noise maskers; this effect was characterized as a release from informational masking.

Previously, Edmonds and Culling (2005a; 2005b) found little support for the idea that perceived location is important for spatial unmasking. However, the present paper appears to suggest that listeners could only exploit information at one location (i.e., ear), as listeners were foiled by the swapped condition of experiment 1 (i.e., producing much higher threshold than one might expect if perceived location was not important to spatial unmasking). Given the design of the experiments described here, one could choose to interpret the results of the current paper in terms of release from informational masking. The results of the swapped condition are a prime candidate for such interpretation, as there is no energetic masking in this condition (i.e., the target and interferer are never physically mixed in a given frequency channel). The elevated thresholds in the swapped condition (i.e., compared to those observed in the dichotic condition), for example, could be characterized as perceptual interference in the segregation of the competing speech streams. Alternatively, if one keeps to the low-level (i.e., nonattentional) framework of Zurek's (1993) model, the SRTs in the swapped condition can be explained as the contribution of target frequency bands at each ear to the intelligibility index. That is, the listener simply listens to the ear providing the better level of intelligibility.

V. CONCLUSIONS

When segregating competing voices, listeners do not benefit from favorable target-to-interferer ratios in different frequency bands if these bands are presented to different ears; the SRTs observed in the current investigation appear to have been determined solely by the quality of target information presented to a single ear. In terms of Zurek's (1993) model, this result is more consistent with the better-ear rule than the better-bands rule for monaural listening. Further investigation will be required to determine if this better-ear mechanism is differentially sensitive to different forms of masking, as it is unclear how grouping by fundamental frequency and perceived location, for example, affect this process.

ACKNOWLEDGMENTS

Barrie Edmonds was supported by the UK EPSRC (Grants GR/M96155 and GR/S11794).

[1]The 1-ERB-wide notch was originally introduced by Edmonds and Culling (2005b) in order to reduce the possibility of confounding binaural interactions in frequency channels close to the splitting frequency. However, as the current investigation employs infinite ILDs and not ITDs the spectral notch is not necessary, but was kept for the sake of consistency across these similarly themed research papers.

[2]Grouping is distinct from fusion. Fusion is the binding together of different elements into a single auditory image. Grouping, as the term implies, involves identifying different sound elements as belonging to distinct auditory objects.

Akeroyd, M. A. (**2004**). "The across frequency independence of equalization of interaural time delay in the equalization-cancellation model of binaural unmasking," J. Acoust. Soc. Am. **116**, 1135–1148.

Bird, J., and Darwin, C. J. (**1998**). "Effects of a difference in fundamental frequency in separating two sentences," in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, Q. Summerfield, and R. Meddis (Whurr, London), pp. 263–269.

Broadbent, D. E. (**1955**). "A note on binaural fusion," Q. J. Exp. Psychol. **7**, 46–47.

Broadbent, D. E., and Ladefoged, P. (**1957**). "On the fusion of sounds reaching different sense organs," J. Acoust. Soc. Am. **29**, 708–710.

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

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D. S., Simpson, B. D., and Freyman, R. L. (**2005**). "Precedence-based speech segregation in a virtual auditory environment," J. Acoust. Soc. Am. **118**, 3241–3251.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.

Cherry, C. (**1953**). "Some experiments on the recognition of speech, with

one and with two ears," J. Acoust. Soc. Am. **25**, 975–979.
Colburn, H. S. (**1977**). "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise," J. Acoust. Soc. Am. **61**, 525–533.
Culling, J. F., and Darwin, C. J. (**1993**). "Perceptual separation of simultaneous vowels: Within and across-formant grouping by *F*0," J. Acoust. Soc. Am. **93**, 3454–3467.
Culling, J. F., and Summerfield, Q. (**1995**). "Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay," J. Acoust. Soc. Am. **98**, 785–797.
Culling, J. F., Edmonds, B. A., and Hodder, K. I. (**2006**). "Speech perception from monaural and binaural cues," J. Acoust. Soc. Am. **119**, 559–565.
Culling, J. F., Hawley, M. L., and Litovsky, R. Y. (**2004**). "The role of head-induced interarual time and level differences in the speech reception threshold for multiple interfering sound sources," J. Acoust. Soc. Am. **116**, 1057–1065.
Cutting, J. E. (**1976**). "Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening," Psychol. Rev. **83**, 114–140.
Durlach, N. I. (**1972**). "Binaural signal detection: Equalization and cancellation theory," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), pp. 369–462.
Edmonds, B. A., and Culling, J. F. (**2005a**). "The role of head-related time and level cues in the unmasking of speech in noise and competing speech," Acta. Acust. Acust. **91**, 546–553.
Edmonds, B. A., and Culling, J. F. (**2005b**). "The spatial unmasking of speech: Evidence for within-channel processing of interaural time delay," J. Acoust. Soc. Am. **117**, 3069–3078.
Fletcher, H., and Galt, R. H. (**1950**). "The perception of speech and its relation to telephony," J. Acoust. Soc. Am. **22**, 89–151.
French, N. R., and Steinberg, J. C. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119.
Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Am. **109**, 2112–2122.
Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2004**). "Effect of number of masking talkers and auditory priming on informational masking in speech recognition," J. Acoust. Soc. Am. **115**, 2246–2256.
Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.
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.
IEEE (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 225–246.
Jeffress, L. A. (**1948**). J. Comp. Physiol. Psychol. **41**, 35.
Kidd, G., Mason, C. R., Brughera, A., and Hartmann, W. M. (**2005**). "The role of reverberation in release from masking due to spatial separation of sources for speech identification," Acta. Acust. Acust. **91**, 526–536.
Kidd, G. J., Mason, C. R., Deliwala, P. S., and Woods, W. S. (**1994**). "Reducing informational masking by sound segregation," J. Acoust. Soc. Am. **95**, 3475–3480.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Levitt, H., and Rabiner, L. R. (**1967**). "Predicting binaural gain in intelligibility and release from masking for speech," J. Acoust. Soc. Am. **42**, 820–829.
Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). "The precedence effect," J. Acoust. Soc. Am. **106**, 1633–1654.
Moore, B. C., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory-filter bandwidths and excitatory patterns," J. Acoust. Soc. Am. **74**, 750–753.
Rabiner, L. R., Laurence, C. L., and Durlach, N. I. (**1966**). "Further results on binaural unmasking and the EC model," J. Acoust. Soc. Am. **40**, 62–70.
Rayleigh, L. (**1876**). "On perception of the direction of a source of sound," Nature (London) **14**, 32–33.
Rayleigh, L. (**1907**). "On our perception of sound direction," Philos. Mag. **8**, 214–232.
Wightman, F. L., and Kistler, D. J. (**1992**). "The dominant role of low-frequency interaural time differences in sound localization," J. Acoust. Soc. Am. **91**, 1648–1661.
Zurek, P. M. (**1987**). "The precedence effect," in *Directional Hearing*, edited by W. A. Yost and G. Gourevitch (Springer, New York), pp. 85–105.
Zurek, P. M. (**1993**). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. A. Studebaker and I. Hochberg (Allyn and Bacon, Boston), pp. 255–276.