The mechanism and factors of the impact of spatial separation on spatial masking

Wang Ruijie
Aalto Universtiy
Master's Programme CCIS / AAT

`ruijie.1.wang@aalto.fi`

Abstract

The audibility of a target sound source in the presence of masking noise is enhanced when there's spatial separation between two sound sources, as opposed to when both sources are located at the same position. Traditional models of spatial unmasking attribute this improvement to both energy and binaural effects. However, empirical evidence suggests that these models fail to fully elucidate spatial unmasking in reverberation environments and overlook the influence of top-down factors. This paper reviews the mechanisms and key parameters influencing spatial unmasking in scenarios involving spatial separation, encompassing energy and binaural effects as well as attention mechanisms.

1 Introduction

The human eye demonstrates remarkable discriminatory abilities. This capability is often modeled in ocular optics by simulating a measurement point of 2900 on the eye. With a 9mm pupil, the resolution threshold is typically 175 microns [1]. Distances exceeding this threshold allow for accurate discrimination between two spatially separated objects. In contrast, the auditory system operates differently. Unlike the eye, which utilizes a lens and multiple receptors to distinguish light from various directions, the ear relies solely on the cochlea. The cochlea lacks direction-sensitive functions, which limits the resolution of the human auditory system compared to vision [2]. Due to this limitation, two sound sources that are spatially close can lead to spatial masking. Therefore, it is crucial to investigate the effects of spatial separation on auditory unmasking.

1.1 Spatial separation

Spatial separation in auditory perception refers to the differentiation in the physical locations of sound sources within three-dimensional space. This spatial disparity results in variances in the timing and intensity of sound signals as they reach the ears, known respectively as interaural time difference (ITD) and interaural intensity difference (ILD), or changes in energy perception in one ear. These discrepancies are crucial cues that enable our auditory system to pinpoint sound sources. The resulting binaural effect allows sounds from different locations to be perceived distinctly.

Spatial unmasking leverages these cues—ILD and ITD—enabling listeners to detect specific target signals within a frequency band. A prominent illustration of this is the cocktail party effect, which demonstrates how individuals can discern specific conversations in noisy environments, highlighting the importance of spatial separation. This effect encapsulates complex interactions involving multiple interfering sound sources, binaural effects, and attention mechanisms [2]. Notably, research by Carhart and colleagues has shown that the degree of spatial unmasking achieved through spatial separation is not contingent upon the number of interfering sound sources [3].

1.2 Spatial unmasking

In the traditional spatial unmasking model, masking signals typically diminish the audibility of a target sound source across various frequency bands. Two spatial factors come into play to counteract this reduction in audibility and achieve clearer perception: the energy effect and the binaural effect. The energy effect involves changes in the sound energy level at the listener's ear due to spatial separation of sources, while the binaural effect leverages differences in the sound received by each ear to enhance signal detection. Together, these factors facilitate the spatial unmasking of sounds, improving the listener's ability to discern the target sound in noisy environments.

1.2.1 Energy effect

The head shadow effect primarily influences the energy effect in spatial unmasking. This phenomenon occurs when the masking and target sound sources are spatially separated. For instance, if the masking source is positioned directly in front of a listener while the target sound source is moved to the right, the head shadow effect comes into play. Consequently, the energy ratio of the target to the masking source at the listener's right ear significantly increases compared to the left ear. This creates a scenario where the right ear, now receiving a stronger signal, effectively becomes the better ear. As a result, the listener perceives the sound from the right ear more distinctly than when listening with both ears.

The energy effect is further substantiated by the research of Best et al., who examined the influence of the head shadow effect on target sound sources with varying power spectral densities, focusing on zebra finches and human speech [4]. Their studies reveal that spatial unmasking varies significantly, between 6-10 dB, depending on the type of speech material. This variation is attributed to differences in spectral distribution among the speech materials.

For instance, the sound signal of zebra finches is predominantly concentrated in high frequencies, where the sound wavelengths are smaller, and the head shadow effect is more pronounced. This results in a notable enhancement of the better ear advantage, which is shown in Fig. 1. They observed an increase in broadband target-to-masker ratio (TMR) of approximately 18 dB in the better ear, which predominantly drives the observed spatial unmasking. Changes in the TMR in the better ear due to spatial separation are identified as the primary contributors to spatial unmasking for noise masking sources.

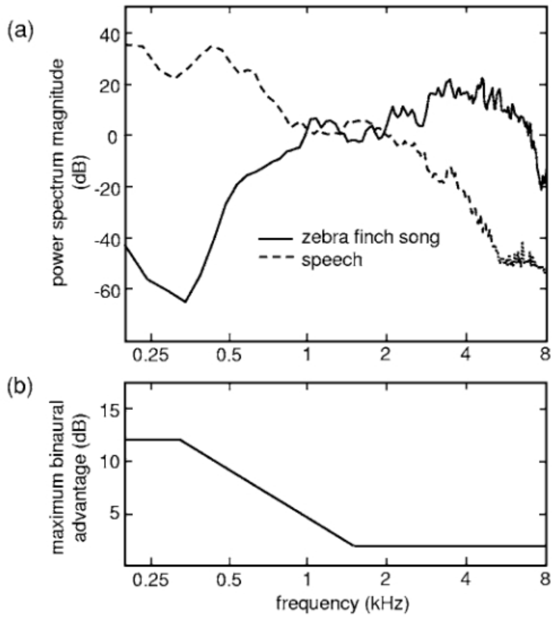


Figure 1: (a) Mean power spectral density plots of two different materials (b) Maximum binaural advantage as the function of frequency

1.2.2 Binaural hearing

Blauert conducted tests using identical configurations for both the masked and target signals to study the differences in intelligibility. His research quantified the level difference in binaural intelligibility (BILD), comparing monaural reference and binaural listening. He found that binaural listening typically offers a 5-7 dB advantage in terms of intelligibility. However, this advantage diminishes when the noise source and the target signal are positioned on the same side as the unblocked ear [5]. In such scenarios, the better ear effect predominates, indicating a stronger influence of the energy effect.

When the interband TMR is low, ITD between the ears becomes a crucial factor in spatial unmasking, activating neurons sensitive to ITD. Edmonds and his colleagues have demonstrated that ITD does not function uniformly across all frequency domains but operates distinctly within individual frequency bands. Their research explored the relationship between ITD settings in different frequency bands and the extent of spatial masking. They discovered that having ITD differences between the target signal and the masking sound in both high and low frequency bands leads to improved spatial unmasking scenarios. Notably, the degree of spatial masking is not influenced by whether the ITD in these bands is similar or different, either scenario can yield the same spatial unmasking effect. Edmonds's findings indicate that ITD at both high and low frequencies can achieve full binaural gain, affirming that ITD is effective independently within each frequency band in spatial unmasking [6].

Although the traditional spatial unmasking model is generally effective, it encounters limitations in reverberation environments. Research by Kidd and his colleagues has shown that the occurrence of spatial unmasking in such settings is closely tied to the frequency spectra of the target sound and masking noise [7]. Spatial unmasking is less pronounced when there is little spectral overlap between the masked noise and the target sound, even if the reverberation energy is minimal. Conversely, significant spatial unmasking can occur when masking noise with a sparse, non-overlapping spectral speech signal is used, despite high reverberation energy. This suggests that factors other than the energy effect become dominant when the spectra of the masking and target signals are similar. The researchers propose that this could be due to top-down spatial attention, which may help manage the competition between simultaneous signal sources for central auditory processing resources.

This paper critiques the traditional spatial unmasking model and elucidates the mechanisms of spatial separation in enhancing auditory clarity. The second section delves into how spatial separation of sound sources influences spatial unmasking and discusses the underlying mechanisms. The third section details the physical parameters that impact spatial unmasking. In the fourth section, the phenomena of spatial unmasking are summarized.

2 Mechanisms Behind Spatial Unmasking

2.1 Energy Concentration from Spatial Separation

Webster pointed out that an analysis of the energy situation within a narrow band is necessary when the spatial unmasking process does not require the signal to be extracted from the full-spectrum noise that may be present in the stimulus due to the cochlea dividing the incoming sound into different frequency channels for analysis[8]. Due to the head shadow effect, target sounds at different frequencies will produce different TMR variations within the better ear during spatial separation, meaning that the energy ratio of the target to the masker in the critical frequency band of the acoustically better ear affects the target's audibility in each band.

Shinn-Cunningham showed that when the target sound and the masker are somewhat spatially separated TMR usually increases, especially in the frequency range where the ear is most sensitive to sound, the critical band. The increased TMR not only increases the energy of the target sound, but also makes the sound easier to distinguish and understand. Within each frequency channel, the increase in narrowband TMR directly affects the audibility of the target sound in that channel. Specifically, for acoustically better ears, increasing the narrowband TMR for a given channel significantly increases the intelligibility of the target sound in that channel. This mechanism is a bottom-up stimulus-driven process that relies on the physical properties of the sound signal rather than the listener's attention or expectations[9].

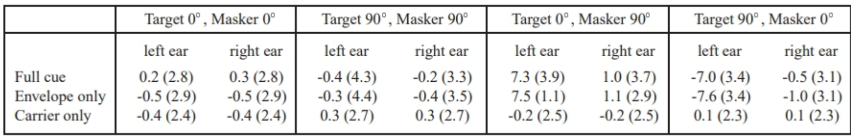
2.2 Decoupling of Binaural Signals

In addition to energy effects, binaural processing is a crucial mechanism used by the human auditory system to localize and discriminate sound sources. When sounds from different directions reach the two ears, the brain can process these differences to identify the source of the sound due to ITD and ILD caused by the propagation of the sound. These differences are significant in the context of spatial masking, as they help the listener to distinguish between the target sound and the background noise, and Shinn-Cunningham showed that when the target sound was present, even at low intensity, the overall firing rate of neurons was reduced by the addition of the target sound, a phenomenon known as interaural deafferentation. This phenomenon is known as "interaural decorrelation." The presence of the target sound alters the correlation between the sound signals received in both ears, thereby reducing the response of brainstem neurons to the masked acoustic signal. This alteration allows the brain to more easily detect the target sound from the masked sound, significantly when the target sound partially overlaps with the masked sound on the frequency spectrum[9].

This mechanism was explored in depth in Edmonds's study, in which he investigated the difference in spatial unmasking obtained by applying different delays to the low and high frequencies of the stimulus signal, respectively, and showed that maximum spatial unmasking was obtained only when the ITD was set for all frequencies of the stimulus signal, i.e., that the ITD was the main factor influencing interaural decorrelation and that this effect was valid both in the high frequencies and in the low frequencies[6]. This finding suggests that interaural decorrelation in spatial unmasking is mainly achieved by exploiting the independent ITD differences within each frequency channel. This suggests that our auditory system cannot only exploit ITDs and ILDs in their entirety, but also that these differences can be processed independently at a finer frequency level, providing a more refined spatial auditory analysis.

2.3 Enhanced Masking Signal Ratio through Spatial Separation

In many audiological studies, particularly those involving spatial unmasking, scholars have often focused on the average TMR over the entire frequency range. Shinn-Cunningham's experimental study provides important insight into this topic[9]. Her study specifically focused on changes in mean TMR between subjects' left and right ears under different spatial configuration conditions. It explored how these changes affected subjects' ability to recognise target sounds. In the absence of spatial separation between the target sound and the masking sound source, there was little significant difference in the mean TMR between the two ears. In this case, subjects were less likely to correctly recognise the target sound due to the relatively low TMR. However, the situation was significantly different once the target sound was spatially separated from the masking source. According to the data in the experiment (shown in Fig. 2), after the spatial separation, the TMR at the subject's better ear was significantly improved by 7.5 dB, and the subject's correct recognition rate was close to 100%, as shown in Fig.3 , demonstrating significant auditory improvement. Special measures were taken in her experimental design to finely process the spectra of the target sound and the masking source to ensure little overlap between the two in the spectrum. The advantage of this design is that it allows the researcher to accurately assess the effect of spatial separation on TMR and hearing performance without the interference of spectral overlap. This is critical because it demonstrates that even when the spectra of the target sound and the masking source do not overlap, spatial separation is still effective in improving TMR in the better ear and, as a result, significantly improves the listener's sound discrimination.



	Target 0° , Masker 0°		Target 90° , Masker 90°		Target 0° , Masker 90°		Target 90° , Masker 0°	
	left ear	right ear	left ear	right ear	left ear	right ear	left ear	right ear
Full cue	0.2 (2.8)	0.3 (2.8)	-0.4 (4.3)	-0.2 (3.3)	7.3 (3.9)	1.0 (3.7)	-7.0 (3.4)	-0.5 (3.1)
Envelope only	-0.5 (2.9)	-0.5 (2.9)	-0.3 (4.4)	-0.4 (3.5)	7.5 (1.1)	1.1 (2.9)	-7.6 (3.4)	-1.0 (3.1)
Carrier only	-0.4 (2.4)	-0.4 (2.4)	0.3 (2.7)	0.3 (2.7)	-0.2 (2.5)	-0.2 (2.5)	0.1 (2.3)	0.1 (2.3)

Figure 2: Average target-to-masker energy ratio (TMR) at the left and right ears

2.4 Tracking the Target Signal with Spatial Separation

Spatial separation not only plays a role in energy distribution within frequency bands and spatial binaural hearing, but it also plays a key role in the allocation of a tester's attention. Attention modelling suggests that when testers are able to perceive a difference in the spatial location of two sound sources, they can concentrate more on

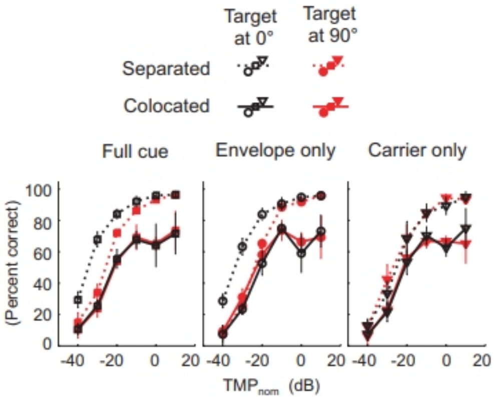


Figure 3: Performance improves with increasing target-to-masker level and with perceived spatial separation between target and masker for all spatial cue conditions

the target signal. This ability to focus attention is especially important when energy effects are not significant, such as when the TMR of the target to the masker between bands is small. Shinn-Cunningham examined the relationship between TMR at the better ear and target sound intelligibility under different spatial cues that lead to perceptual differences between the target and the masker, using a psychometric function[9]. Her study noted that for a given value of TMR at the best ear, the amount of spatial unmasking gain due to the attention factor caused by spatial separation was always the same even across spatial cues, that this gain was equivalent to a 5 dB increase in TMR at the better ear at threshold, as shown in Fig. 4, and that this factor could act on cases where the TMR was small, and that Shinn-Cunningham categorised this discrepancy as a top-down factor, with the difference in perceived location leading to subjects being able to focus more on the target sound, thus reducing the effect of masking noise. This finding demonstrates a departure from traditional models of spatial unmasking in that when the TMR is very small, traditional models suggest that there should be no spatial unmasking at this point, but the study shows that spatial unmasking still occurs at this point.

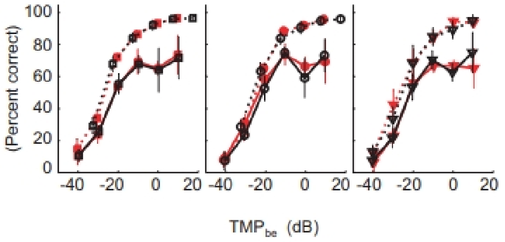


Figure 4: Performance improves with increasing better ear target-to-masker level and with perceived spatial separation between target and masker for all spatial cue conditions

In addition, Freyman's research further explored the complexity of attention factors in practice[10]. He found that although spatial unmasking is enhanced when there is a perceived difference in spatial location between the target sound and the masking sound, this enhancement in unmasking is not significant when the masking sound changes from a human voice to another type of noise. This suggests that in real-world environments, differences in multiple dimensions between the target and masked voices-such as fundamental frequency, timbre, speaker gender, and other factors-are also important for reducing interference, and that unmasking based on perceived location differences alone may not be sufficiently effectively reduce interference.

3 Parameters that affect the amount of spatial unmasking

Shinn-Cunningham's study explored the effects of different spatial cues on the spatial unmasking effect. By using three different stimulus signals: a full-cue signal containing all spatial cues, an envelope-only cue signal containing only ILDs and ITDs, and a carrier-only signal containing only IPDs, the processing of the signals is shown in Fig. 5. Shinn-Cunningham analysed the performance of these cues in different configurations. The results of the study, which is shown in Fig. 6, showed that stimuli containing only envelope cues performed similarly to stimuli containing ITD or ILD cues on spatial unmasking under higher TMR conditions, i.e., TMR above -20 dB. This finding is at odds with the predictions of traditional binaural models, which suggest that the spatial unmasking effect of envelope-only cues is typically smaller. However, at TMR below -20 dB, stimuli containing only envelope cues showed less spatial unmasking effect, further confirming the importance of ITD cues in low TMR conditions [9].

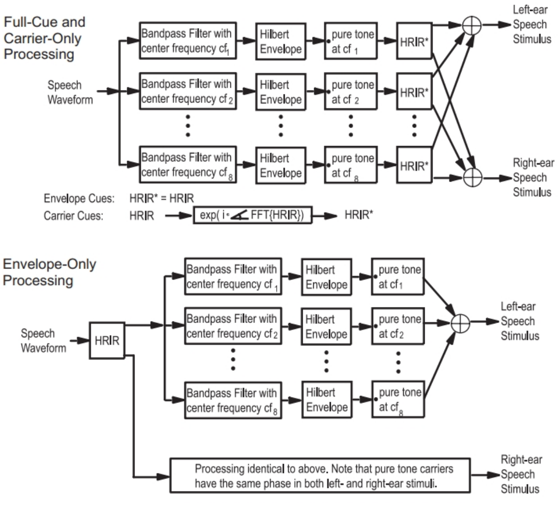


Figure 5: Three Different Signal Processing Methods to Model Spatial Unmasking

In addition, Shinn-Cunningham investigated the effect of narrowband target-to-noise ratio (TMRband) on the amount of spatial unmasking, as shown in Fig. 7. By evaluating the performance of different spatial cues in the masking scenarios of the above target band, the following target band, and the simultaneous presence

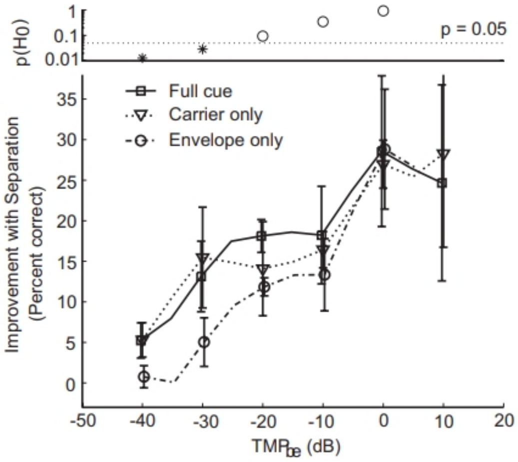


Figure 6: Boosting of spatial unmasking by different spatial cues after considering TMR for better ears

of the upper and lower target bands, it was shown that subjects demonstrated similar TMRband changes in all three masking scenarios. This finding suggests that although spatial cues are critical to the overall auditory unmasking effect, TMRband changes are less dependent on different spatial cues. This suggests that a more comprehensive range of spatial cues could be considered when designing hearing aids for specific hearing loss conditions, rather than being limited to specific types [9].

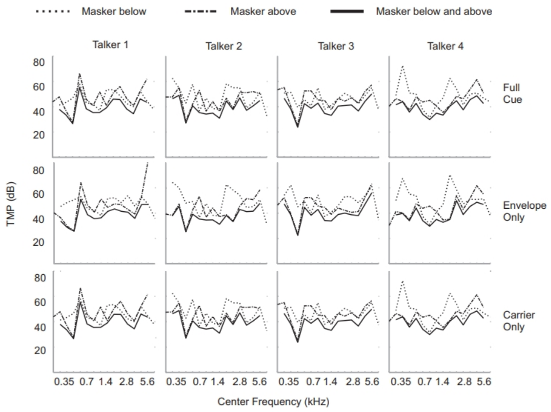


Figure 7: Subjects' narrowband TMR changes with different spatial cues

4 Conclusion and discussion

In summary, spatial unmasking driven by spatial separation can be attributed to four primary mechanisms. First, spatial separation enhances the TMR within the narrow interaural band of the better ear, leading to improved intelligibility of the target sound. This energy change is a bottom-up mechanism. Secondly, spatial separation modifies the ITD across frequency bands, causing spatial de-correlation of interaural signals. This bottom-up mechanism aids in discriminating low-level target signals under conditions of low TMR.

Moreover, spatial separation not only enhances the better ear effect by improving broadband TMR at the better ear, but it also facilitates spatial unmasking by leveraging binaural effects. Beyond these physical changes, spatial separation initiates a top-down mechanism that enables subjects to focus on a target sound even when it possesses lower energy, enhancing perceptual clarity.

The evolution of these findings marks a significant departure from traditional binaural models which predominantly emphasized the role of spatial cues. Recent research indicates that while spatial cues continue to be crucial, their influence on the amount of spatial unmasking is minimal when TMR exceeds -20 dB. However, at very low TMR levels, the type of spatial cue becomes critically important, leading to notable differences in spatial unmasking outcomes.

References

[1] Sergio Bonaque-González, Juan M Trujillo-Sevilla, Miriam Velasco-Ocaña, Óscar Casanova-González, Miguel Sicilia-Cabrera, Alex Roqué-Velasco, Sabato Ceruso, Ricardo Oliva-García, Javier Martín-Hernández, Oscar Gomez-Cardenes, et al. The optics of the human eye at 8.6 μm resolution. *Scientific Reports*, 11(1):23334, 2021.

[2] Ville Pulkki and Matti Karjalainen. *Communication acoustics: an introduction to speech, audio and psychoacoustics.* John Wiley & Sons, 2015.

[3] Raymond Carhart, Tom W Tillman, and Elizabeth S Greetis. Release from multiple maskers: Effects of interaural time disparities. *The Journal of the Acoustical Society of America*, 45(2):411–418, 1969.

[4] Virginia Best, Erol Ozmeral, Frederick J Gallun, Kamal Sen, and Barbara G Shinn-Cunningham. Spatial unmasking of birdsong in human listeners: Energetic and informational factors. *The Journal of the Acoustical Society of America*, 118(6):3766–3773, 2005.

[5] Jens Blauert. *Spatial hearing: the psychophysics of human sound localization.* MIT press, 1997.

[6] Barrie A Edmonds and John F Culling. The spatial unmasking of speech: evidence for within-channel processing of interaural time delay. *The Journal of the Acoustical Society of America*, 117(5):3069–3078, 2005.

[7] Gerald Kidd, Christine R Mason, Andrew Brughera, and William M Hartmann. 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, 2005.

[8] Frederic A Webster. The influence of interaural phase on masked thresholds i. the role of interaural time-deviation. *The Journal of the Acoustical Society of America*, 23(4):452–462, 1951.

[9] Barbara G Shinn-Cunningham, Antje Ihlefeld, Eric Larson, et al. Bottom-up and top-down influences on spatial unmasking. *Acta Acustica United with Acustica*, 91(6):967–979, 2005.

[10] Richard L Freyman, Uma Balakrishnan, and Karen S Helfer. Spatial release from informational masking in speech recognition. *The Journal of the Acoustical Society of America*, 109(5):2112–2122, 2001.