A Behavioral Study on the Effects of Rock Music on Auditory Attention

Letizia Marchegiani[1] and Xenofon Fafoutis[2]

[1] Language and Speech Laboratory, Faculty of Arts,
University of Basque Country
l.marchegiani@laslab.org
[2] Department of Applied Mathematics and Computer Science,
Technical University of Denmark
xefa@dtu.dk

Abstract. We are interested in the distribution of top-down attention in noisy environments, in which the listening capability is challenged by rock music playing in the background. We conducted behavioral experiments in which the subjects were asked to focus their attention on a narrative and detect a specific word, while the voice of the narrator was masked by rock songs that were alternating in the background. Our study considers several types of songs and investigates how their distinct features affect the ability to segregate sounds. Additionally, we examine the effect of the subjects' familiarity to the music.

Keywords: Auditory Attention, Speech Intelligibility, Cocktail Party Problem

1 Introduction

Colin Cherry coined the term *Cocktail party effect* to indicate the human ability to pay attention, in particularly noisy acoustic scenarios (like a cocktail party), to the speech of only one of the present talkers, ignoring the other sounds and voices around [6]. Knudsen defined attention as a filter between all the incoming stimuli, "selecting the information that is most relevant at any point in time" [14]. A long debate about the collocation of this filter along the perception process has raged for many years and several studies and experiments have been performed to understand how attentive mechanisms decide on the saliency of a stimulus.

Bregman claimed that the perceptual process is articulated in two phases: a preliminary separation of all the signals of the mixture in segments, on the base of the generating source, and a following grouping of the segments in streams [3]. Cusack *et al.* [7] and Carlyon [5] confirmed Bregman's findings and proved that the way in which the stimuli are organized as part of the same audio flow and the level of analysis performed on each of them, is broadly affected by attention. These assertions reduce the cocktail party effect mainly to a sound source segregation problem, opening a new perspective of investigation on which

factors could influence the segregation procedure and how this ability is related to the concept of saliency.

Cherry proposed that some specific cues could help the mental ability of isolating a single sound from the environment, such as different speaking voices, different genders of the competitive talkers (see also [9]), different accents and previous knowledge. The voice features which could facilitate the segregation process, like difference in the fundamental frequency, phase spectrum or intensity, are illustrated in [19]. The spatial location of the source also plays a crucial role (the so called *spatial unmasking*), as shown in [1] and [2]. Depending on to the nature of these factors, it is possible to analyze human attentive behavior under two different angles: a bottom-up and a top-down one. According to the bottom-up perspective, the sounds which pop out of the acoustic scene, such an ringing alarm, result to be salient. In the top-down perspective, on the other hand, saliency is driven by acquired predispositions, the presence of a task or a specific goal.

We are interested in top-down attention in a simulated cocktail party scenario, in which the listening capability is challenged by the presence of rock music in the background. We chose to begin our investigation with rock music because, in addition to its popularity, it is shown to be distracting from performing a task [18]. Studying how attention is influenced by music has significance in several domains. From one perspective, organizers of social events or DJs can choose background music with respect to its effect on the ability of the participants to communicate. Up to some extend, they might be able to direct their attention and their behavior. Furthermore, music composers can incorporate in their compositions features that attract the attention of their audience. Parente [22] explored the distracting efficacy of rock music and the influence of music preference, showing a positive effect of music liking on task performance. Later, North and Hargreaves [20] confirmed these results, making subjects play computer motor racing games either while accomplishing a backward-counting task or in the absence of it. They also demonstrated that arousing music determines a bigger confusion than less arousing music. The impact of loudness has been investigated in [26], while the effect of music tempo on reading abilities has been studied in [11].

In this paper, we analyze the distribution of attention in a noisy environment, in which the voice of interest is masked by alternating songs with specific features. In order to understand how these features affect speech intelligibility and the performance in a listening task, we carried out some behavioral experiments, asking our subjects to follow a narrative and push a button each time they hear a specific word. In particular, we investigate the influence of soft and hard rock songs, along with songs with high dynamics. The latter are songs that frequently alternate between soft and hard states multiple times throughout their duration. The effect of familiarity to the music is also examined. Our analysis is twofold. First, we investigate the influence of the temporal and spectral overlap of the narrative to the background music to exclude the case that the performance

doesn't depend on auditory attention, but on the inability of the subjects to listen to the speaker. Then, we analyze the influence of the songs.

The remainder of the paper is structured as follows. Section 2 presents the selected songs and the behavioral experiments. Section 3 analyses the experimental results. Lastly, Section 4 concludes the paper.

2 Experiment Setup

The experiment aims to identify how rock music influences the performance in tasks that require attention. In a nutshell, the participants were asked to focus their attention on a narrator and identify a specific word, while in the background different songs were alternating.

The narrative was a fairy tale, entitled *The Adventures of Reddy Fox* [4]. Specifically, we used the 14 minutes out of the first five chapters of the audiobook[3]. Since it is targeted to children, the fairy tale uses simple language that is relatively easily understood by non-native English speakers. Since it is relatively easy to lose attention while performing a trivial task, the subjects were asked to identify the word *'and'*. The selected word is very common and it can be easily missed. Thus, the participants' full attention is required to successfully perform the task. Furthermore, with such a common word we avoid bottom-up cues that depend on the rarity of the sound and the possible "surprise effect", as described in [10]. The duration of the narrative was 14 minutes. During the first 2 minutes of the narrative, there was no background music. During the remaining 12 minutes, 6 songs were alternating in the background. Each song played for 2 minutes. The original story was slightly modified so that the target word, *'and'*, appears 9 or 10 times in each 2-minute time slot, resulting to a total amount of 67 word appearances.

The carefully selected songs had particular properties that affect the attention in different ways. Our primary goal, is to identify the effect of dynamics in the songs. Since unexpected sensory stimuli tend to attract the attention [10], background music with high dynamics is expected to significantly disrupt the subjects. Additionally, we consider two categories of rock music with low dynamics. The former is soft rock songs, that are characterized by low emotional intensity, clean vocals, peaceful drumming and guitars without distortion sound effects. The latter is hard rock songs, that are characterized by high emotional intensity, high-pitched screaming vocals, intense drumming and guitars with distortion sound effects. Rock songs with high dynamics tend to alternate between soft and hard states multiple times throughout their duration. We note that the terms *soft* and *hard* do not refer to particular properties of the audio, such as the volume, but rather on the aggressiveness of the performance, as this is indicated by the musical terms *piano* and *forte*. The selected songs represent these three classes of rock music, that for the remainder of the paper will be referred to as *HD* (High Dynamics), *ND* (No Distortion) and *D* (Distortion).

[3] http://www.booksshouldbefree.com/book/the-adventures-of-reddy-fox-by-thornton-w-burgess

Table 1. The rock songs selected as background music and their respective properties.

Song Code	Artist	Song Title	Listeners	Dynamics	Soft / Hard
HD-NP	The Pixies	Gouge Away	320228	High	Both
HD-P	Nirvana	Smells Like Teen Spirit	1589584	High	Both
ND-NP	The National	Runaway	203268	Low	Soft
ND-P	Radiohead	Karma Police	1285583	Low	Soft
D-NP	Mother Love Bone	This is Sangrila	72672	Low	Hard
D-P	Guns 'n' Roses	Welcome to the Jungle	981998	Low	Hard

Apart from the song properties, we expect the subjects' familiarity to the songs to significantly affect the task performance. Such influence can be of various natures. For instance, a subject might feel the tendency to sing along with a favorite song or might have associated the song with specific memories. In order to identify this influence, we have selected two songs of each class, a popular and an unpopular. The unpopular songs aim to identify the influence of the song properties clean from the effects of familiarity. Then, the relative comparison to the popular songs will indicate the effects of the subject's familiarity to the songs. The popularity of the songs was assessed based on the statistics of the *Last.fm*[4] music social network. In particular, the popular songs were selected among songs that have more than 900000 unique listeners. The listeners of the unpopular songs are one order of magnitude less than the respective popular song. In an attempt to verify the validity of the song selection, the subjects were questioned to characterize their familiarity to the songs. For the remainder of the paper, a suffix on the code name of each song indicates its popularity. Specifically, *-P* indicates a popular song and *-NP* indicates an unpopular song.

Table 1 summarizes the selected songs with their respective properties. The fourth column shows the unique listeners in *Last.fm* at the time of the song selection. All songs are available in common audio / video streaming services. When mixed with the narrative, the volume of the songs was adjusted to the same level and the transition between two consecutive songs was smoothed out using fading. In particular, we adjusted the peak volume of all songs to $-6dBFS$ (while the narrative was adjusted to $-3dBFS$). Furthermore, we made sure that no word *'and'* appears in the transition between two different songs. The songs were mixed in two different orders between which, the subjects were divided. The purpose of this is to mitigate the influence of the subjects' fatigue on the results. Table 2 shows the song order as mixed with the narrative. The last column shows the total number of appearances of the word *'and'* for each 2-minute slot.

Prior to the actual experiment, the subjects were asked to do a 1-minute test experiment to get familiar with their task. The test experiment was using a different narrative and song from the actual experiment. During the actual experiment, the time the subject was clicking the button was recorded. Lastly, the subjects were allowed to pause the experiment. After the completion of the experiment, the subjects were asked to characterize their familiarity to the songs. In particular, they were asked to choose from the following options:

[4] http://www.last.fm

Table 2. Song order as mixed with the narrative.

Narrative Time	**Oder 1**	**Order 2**	**Words**
0:00-2:00	No Music	No Music	9
2:00-4:00	HD-P	ND-P	9
4:00-6:00	ND-NP	HD-NP	10
6:00-8:00	D-NP	D-P	9
8:00-10:00	D-P	D-NP	10
10:00-12:00	ND-P	ND-NP	10
12:00-14:00	HD-NP	HD-P	10

- Not familiar. I have never listened to the song before.
- Barely familiar. It reminds me something, but I'm not able to recognize it.
- Quite familiar. I have listened to the song enough times and I know it sufficiently.
- Very familiar. I know the song very well and I'm able to recognize it. I have listened to it many times.

According to the answers of each subject, $0-3$ points were assigned to each song (0 represents zero familiarity). The normalized average value among all the subjects defines the *Familiarity Index (FAM* $\in [0,1]$*)* of each song.

A total amount of 22 subjects (similarly to previous works on selective attention [8][23][7][17]), between $25-35$ years old, with no hearing, language or attentional impairment, participated in the experiment (11 subjects per song order). Their task performance, their answers to the post-questionnaire and occasional short interviews suggest that all the subjects understood their task at a sufficient level and conducted the experiment in silent environments using headphones.

3 Experimental Results and Analysis

For each subject, we consider as hits any word identification that has a timestamp within 3 seconds from the actual word appearance in the narrative. All other word identifications are considered false alarms and are excluded from the results. Figure 1 shows the total number of appearances of the target word in the narrative, as well as the total number of hits and false alarms for each song, aggregated over all the 22 subjects. The relatively high performance when no background music was present, shows that the subjects were able to perform the task. For each one of the 67 word appearances, Figure 2 shows the ratio of subjects who successfully identified the word over the total number of subjects. The analysis of the results continues as follows. First, we aim to identify if there is a significant correlation between the subjects' performance and the temporal and spectral overlap of the narrative and the background music. Assuming that such correlation doesn't exist, the relative performance variation in presence of different music can only depend on the properties of the songs.

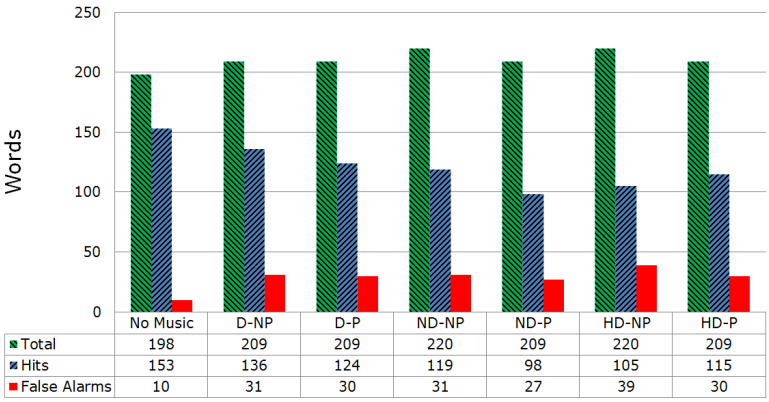

Fig. 1. Total number of word appearances and number of hits and false alarms per song aggregated all over all the subjects.

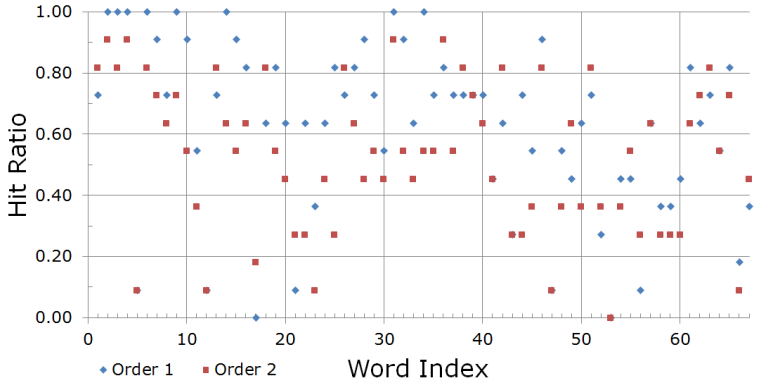

Fig. 2. Hit ratio over the total number of subjects for each of the 67 word appearances.

3.1 Audibility: Spectral and Temporal Overlap

We compute the spectral and temporal overlaps introduced by the musical background, making use of the concept of Ideal Binary Mask (IBM). Wang [24] first proposed the idea of IBM as the aim of Computational Auditory Scene Analysis (CASA), in terms of extrapolation of a target signal from a mixture. Further investigations [25][13] have shown that these masks can be exploited to improve the speech reception threshold and, more generally, speech intelligibility, both in impaired and normal-hearing listeners. In [15] these results has been confirmed by exploring in more detail some of the factors which could affect these improvements. As highlighted in [24], IBMs are defined according to the nature of the signal of interest and their performance is similar to the way the human auditory system functions in the presence of masking. These characteristics are

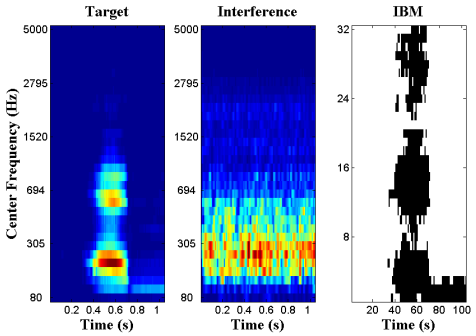

Fig. 3. Example of IBM, obtained with SNR=0 dB, LC=-4 dB, windows length=20 ms, frequency channels=32. The ones are indicated by the black bins, the zeros by the white bins.

crucial for the perceptual representation and analysis of different acoustic scenarios. In [17], IBMs are used to calculate the masking between two narratives uttered by a speech synthesizer in a monaural combination. We follow the same approach to estimate spectral and temporal overlaps between the story and the songs and their relative effect on speech intelligibility.

A binary mask is a binary matrix in which 1 represents the most powerful time-frequency regions of the target signal compared to an interference signal, according to a local criteria (LC). If $T(t, f)$ and $I(t, f)$ denote the target and interference time-frequency magnitude, then the IBM is defined by the following formula.

$$\mathrm{IBM}(t, f) = \begin{cases} 1, & \text{if } T(t, f) - I(t, f) > LC \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

In Figure 3, an example of the IBM relative to one of the 'and' in the story is shown. The spectrograms of a target sound signal (the story) is compared to an interference signal and the regions of the target with the highest energy are kept in the resultant IBM. As interference signal, we use a Speech Shaped Noise (SSN) of reference. The time frequency (T-F) representation is based on the model of the human cochlea, by the use of gammatone filtering (see [16]). The parameters controlling the structure of the binary masks are, apart from the LC, the windows length (WL) and the number of frequency channels (FC).

We estimate the masking between each audio frame containing the word 'and' in the story and the respective frame in the song sequence. We use the definition of overlaps given in [17], which are based on the comparison between the IBMs correspondent to each pair of frames. The spectral overlap is determined by the co-occurrence of black bins in the two binary masks over the total number of time-frequency bins. The temporal overlap is obtained by compressing the IBMs over frequency, assigning value 1 if there is at least a black slot in one of the relative frequency bins and 0 otherwise (0 is considered as silence). The resulting binary vectors, named Compressed Ideal Binary Masks (CIBM) are

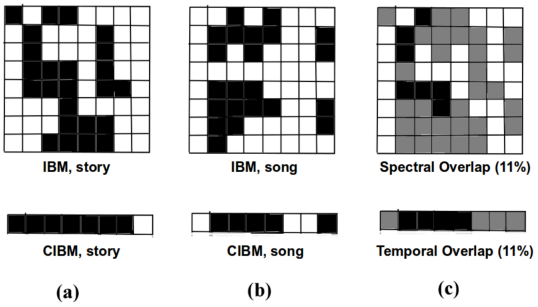

Fig. 4. An example of spectral and temporal overlap estimation. Only black regions represent overlapped parts on (c).

compared and the amount of temporal overlap is given again by the number of co-occurrence of black bins on the CIBMs over the total number of bins in the vectors. Figure 4 illustrates the temporal and spectral overlap definitions.

Initially, we compute the overlap between each word *'and'* and the background music using IBMs with the following parameters: $SNR = 0\ dB$, $LC = -4\ dB$, $WL = 20\ ms$ and $FC = 32$. We consider the total number of times each word *'and'* has been correctly detected as a measure of speech intelligibility. The results suggest small positive correlation between the spectral overlap (0.08 for the first order and 0.056 for the second) and the subjects' performance, as well as small negative correlation between the temporal overlap (-0.103 for the first order and -0.056 for the second) and the subjects' performance. The results are validated using a permutation test with 10000 resamples, at 5% significance level, which indicates no significant correlation ($p > 0.22$). We, then, optimize the parameters of the IBMs (LC, WL and FC) keeping $SNR = 0$ to maximize the correlation and apply again a permutation test with 10000 resamples at the same significance level. The test shows no significant correlation even in the case of optimized parameters ($p > 0.11$). Therefore, there is no significant correlation between the masking level and the ability of the subjects to identify the requested words and the difference in the performance of the subjects can only be attributed to the song properties.

3.2 Analysis of Song Influence

Using the answers of the post-questionnaire regarding the familiarity of each subject to each song, we calculate the *Familiarity Index (FAM)* of each song, as described in Section 2. Table 3 shows the familiarity index of each song in comparison to the number of unique *Last.fm* listeners that we used to define their popularity. The results suggest that our subject's familiarity to the songs matches their popularity. An ANOVA test on FAM shows significant ($p < 10^{-10}$) difference between popular and unpopular songs.

Table 3. The familiarity of the subjects to the songs matches their popularity.

Song Code	Listeners	FAM
HD-NP	320228	0.41
HD-P	1589584	0.8
ND-NP	203268	0.33
ND-P	1285583	0.55
D-NP	72672	0.15
D-P	981998	0.79

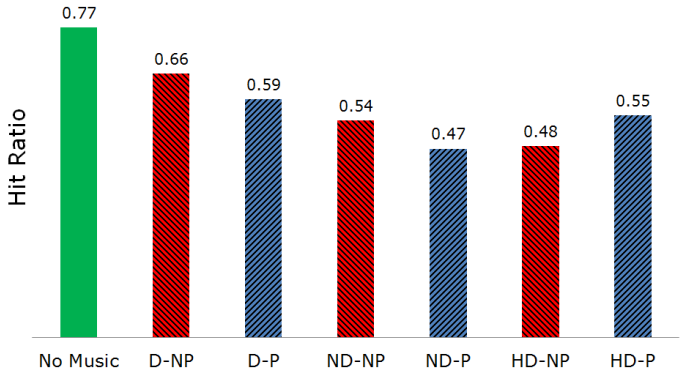

Fig. 5. Average hit ratio of the subjects for each song.

Figure 5 shows the average hit ratio of the 22 subjects for each song. We note that the performance variation between different songs is at the same order of magnitude as the difference of the performance between no music and music, which indicates its significance. Furthermore, we performed an ANOVA test which shows that the difference between the various backgrounds is significant ($p < 0.0001$).

Since the unpopular songs can be characterized as unfamiliar to the subjects, a comparison between them would expose the influence of background music on attention solely based on the song properties. Observe that the subjects' performance in the song with high dynamics (HD-NP) is significantly lower than the respective songs with low dynamics (D-NP and ND-NP). High dynamics in music are shown to attract significantly the attention of the subjects. Since the subjects are unfamiliar with the song, the frequent and sudden changes in the song's dynamics are unexpected and, thus, distract the subjects from their task. The relative comparison between the two songs with low dynamics suggests that hard rock music (D-NP) attracts the attention at a lower level compared to softer rock music (ND-NP). This phenomenon happens because distorted music is perceived by the human mind as more noisy. Thus, the human mind is significantly more capable to differentiate it from the narrator's voice and ignore

Table 4. List of common mistakes.

Time	Subjects	Actual Text
6:03	12	"End of"
12:28	11	"in broad"
5:00	9	"As she"
8:19	8	"that he"

it. On the soft song, on the other hand, the background music is much more similar to the narrator's voice and it is harder for the human mind to separate them. Indeed, the greater the difference between the features of two sounds, the easier the segregation process is [6]. An ANOVA test shows a significant effect of the style of the songs on task performance ($p = 0.018$).

Next, we compare the performance between the popular and unpopular song of each type to identify the influence of the subjects' familiarity to the songs on attention. We note that it is hard to generalize how familiarity affects a specific subject. Indeed, the answers to the post-questionnaire suggest that familiarity generated emotions of different nature to different subjects. For example, some subjects stated that songs gave them the tendency to sing or hum along. Other subjects found the songs annoying or answered that songs made them remember past experiences. When a song becomes an emotional trigger, familiarity is expected to negatively affect the subject's performance. However, overexposure to specific sensory stimuli, such as a song, can lead to a state of apathy or indifference to it [10]. Such a state would have the opposite effect on task performance. Nevertheless, our results indicate that in the songs with low dynamics (D-NP, D-P, ND-NP, D-NP), the subjects' familiarity to the music acts as an emotional trigger that attracts the attention. Interestingly, the results in the songs with high dynamics (HD-NP, HD-P) indicate the opposite. Given the subjects' familiarity to the song (HD-P), the frequent and sudden changes in the song's dynamics cannot be considered unexpected. Contrary to the respective unpopular song (HD-NP), the sudden changes in the dynamics are anticipated by the subjects who are more capable to keep their attention on their task.

Lastly, we noticed that there are some common mistakes among the subjects. Table 4 summarizes how many subjects did the specific common mistake. The last column indicates what the narrator actually said instead of the word *'and'* as perceived by some of the subjects. The coherent confusion, that can be attributed to the phonetic similarities of the words, suggests that some subjects were focused on catching words, rather than semantically interpreting the meaning of what they were listening to. Attentive mechanisms are responsible of allocating resources, assigning saliency and deciding on the level of analysis required for each stimulus, according to task difficulty. Therefore, it would be interesting to understand if subjects' behavior was a strategy to better accomplish the task or if the complexity of the task did not allow them to follow the

story. It should be also noted that there were no common mistakes that were associated to the appearance of the word *'and'* in the lyrics of the songs.

4 Conclusion and Future Work

We performed behavioral experiments to investigate the distribution of attention in a simulated cocktail party scenario, characterized by the presence of rock music in the background. The subjects were asked to identify a specific words from a narrative while different songs were sequentially playing in the background. We showed that some specific features of the songs result to be more confusing than others while performing the assigned task, giving hints about the distracting power of some particular kinds of songs (D, ND, HD). Further analysis could be carried out in the future to analyze more specifically the nature of these features. Moreover, previous works (e.g. [21]) proved that attention can be highly influenced by the emotional state induced in the subject by a stimulus. With regards to arousal aspects, for example, provocative stimuli that are able to induce surprise or fears, are easily detectable even in situations in which the subject is exposed to a strong cognitive load because of another task that requires attention. Other investigations [12] provided a characterization of emotional associations which could be generated by music and triggered by particular acoustic features, drawing to a classification of songs on the base of these associations. Therefore, we plan to explore how the emotional character of the songs (considering both arousal and valence effects) can influence task performance. Such a study would also provide more conclusive results regarding the effects of familiarity.

References

1. Arbogast, T.L., Mason, C.R., Kidd Jr, G.: The effect of spatial separation on informational and energetic masking of speech. J. of the Acoustical Society of America 112, 2086 (2002)
2. Arbogast, T.L., Mason, C.R., Kidd Jr, G.: The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners. J. of the Acoustical Society of America 117, 2169 (2005)
3. Bregman, A.S.: Auditory Scene Analysis: The perceptual organization of sound. The MIT Press (1994)
4. Burgess, T.W.: The Adventures of Reddy the Fox. Little Brown and Company (1923)
5. Carlyon, R.P.: How the brain separates sounds. Trends in Cognitive Sciences 8(10), 465–471 (2004)
6. Cherry, E.C.: Some experiments on the recognition of speech, with one and with two ears. J. of the Acoustical Society of America 25, 975 (1953)
7. Cusack, R., Deeks, J., Aikman, G., Carlyon, R.P., et al.: Effects of location, frequency region, and time course of selective attention on auditory scene analysis. J. of Experimental Psychology-Human Perception and Performance 30(4), 643–655 (2004)

8. Darwin, C.J., Brungart, D.S., Simpson, B.D.: Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. The Journal of the Acoustical Society of America 114, 2913 (2003)
9. Drullman, R., Bronkhorst, A.W.: Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation. J. of the Acoustical Society of America 107, 2224 (2000)
10. Itti, L., Baldi, P.: Bayesian surprise attracts human attention. Advances in Neural Inform. Process. Syst. 18, 547 (2006)
11. Kallinen, K.: Reading news from a pocket computer in a distracting environment: effects of the tempo of background music. Comput. in Human Behavior 18(5), 537–551 (2002)
12. Kim, Y.E., Schmidt, E.M., Migneco, R., Morton, B.G., Richardson, P., Scott, J., Speck, J.A., Turnbull, D.: Music emotion recognition: A state of the art review. In: Proc. ISMIR. pp. 255–266. Citeseer (2010)
13. Kjems, U., Boldt, J.B., Pedersen, M.S., Lunner, T., Wang, D.: Role of mask pattern in intelligibility of ideal binary-masked noisy speech. J. of the Acoustical Society of America 126, 1415 (2009)
14. Knudsen, E.I.: Fundamental components of attention. Annu. Reviews Neuroscience 30, 57–78 (2007)
15. Li, N., Loizou, P.C.: Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction. J. of the Acoustical Society of America 123, 1673 (2008)
16. Lyon, R.: A computational model of filtering, detection, and compression in the cochlea. In: Proc. IEEE Int. Conf on Acoust., Speech, and Signal Process. (ICASSP). vol. 7, pp. 1282–1285. IEEE (1982)
17. Marchegiani, L., Karadogan, S.G., Andersen, T., Larsen, J., Hansen, L.K.: The role of top-down attention in the cocktail party: Revisiting cherry's experiment after sixty years. In: Proc. 10th Int. Conf. on Machine Learning and Applications and Workshops (ICMLA). vol. 1, pp. 183–188. IEEE (2011)
18. Mayheld, C., Moss, S.: Effect of music tempo on task performance. Psychological Rep. 65(3f), 1283–1290 (1989)
19. Moore, B.C., Gockel, H.: Factors influencing sequential stream segregation. Acta Acustica United with Acustica 88(3), 320–333 (2002)
20. North, A.C., Hargreaves, D.J.: Music and driving game performance. Scandinavian J. of Psychology 40(4), 285–292 (1999)
21. Öhman, A., Flykt, A., Esteves, F.: Emotion drives attention: detecting the snake in the grass. J. of Experimental Psychology: General 130(3), 466 (2001)
22. Parente, J.A.: Music preference as a factor of music distraction. Perceptual and Motor Skills 43(1), 337–338 (1976)
23. Shinn-Cunningham, B.G., Ihlefeld, A.: Selective and divided attention: Extracting information from simultaneous sound sources. In: International Community for Auditory Display (ICAD) (2004)
24. Wang, D.: On ideal binary mask as the computational goal of auditory scene analysis. Speech Separation by Humans and Machines 60, 63–64 (2005)
25. Wang, D., Kjems, U., Pedersen, M.S., Boldt, J.B., Lunner, T.: Speech intelligibility in background noise with ideal binary time-frequency masking. J. of the Acoustical Society of America 125, 2336 (2009)
26. Wolfe, D.E.: Effects of music loudness on task performance and self-report of college-aged students. J. of Research in Music Educ. 31(3), 191–201 (1983)