

REVIEW ARTICLE

An overview on informational masking

Marzieh Amiri^{1,2}, Farnoush Jarollahi^{2*}

¹- Department of Audiology, School of Rehabilitation Sciences, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran

²- Department of Audiology, School of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran, Iran

Received: 13 Jan 2020, Revised: 24 Feb 2020, Accepted: 13 Mar 2020, Published: 15 Jul 2020

Abstract

Background and Aim: In noisy environments, two types of masking including energetic masking (EM) and informational masking (IM) occur. EM results from the spectral overlap of the target and maskers on the basilar membrane, while IM occurs at higher level. This paper aimed to review the concept of IM in terms of historical perspective and definitions, the important cues for releasing from it, age-related effects and its neural basis.

Recent Findings: The data from psychoacoustic, behavioral, and neuro-imaging studies were reviewed and discussed in order to provide an overall image of IM. According to these studies, it seems that perceptual segregation between the target and maskers is the most important cues for releasing from IM. This process takes place simply and without any effort in adults with normal hearing; however, it does not occur easily in children, elderly people and those with impaired hearing. Moreover, it seems that both top-down and bottom-up processing are involved in IM formation.

Conclusion: Since IM leads to failure in selection of auditory objects and prevents the individual from auditory scene analysis, understanding the IM concept leads to a better knowledge of

speech perception in noise.

Keywords: Informational masking; perceptual masking; energetic masking

Citation: Amiri M, Jarollahi F. An overview on informational masking. *Aud Vestib Res.* 2020;29(3):128-39.

Introduction

Understanding speech in the presence of background noise can be challenging. In today's society, people are constantly exposed to different types of background noises. The ability to communicate in the presence of these noises is essential for success in education and social interaction [1]. In a complicated acoustic environment comprising of various sound sources, selective attention to one source and ignoring others has been considered as one of the major challenges of the human auditory system. The acoustic waveforms resulted from different sound sources are linearly mixed with each other before arriving at the ears, and the brain is responsible for the auditory scene analysis and sound extraction [2]. In most of these conditions, competing speech materials originate from different sources, and listeners can spatially segregate the competing messages by using differences in inputs received by two ears. This listening situation is a classic cocktail party problem described by Cherry (1953) for the first time, and has been widely studied over the past 50 years [2]. This process takes place simply in

* **Corresponding author:** Department of Audiology, School of Rehabilitation Sciences, Iran University of Medical Sciences, Shahid Shahnazari St., Madar Square, Mirdamad Blvd., Tehran, 15459-13487, Iran. Tel: 009821-22228051, E-mail: jarollahi.f@iums.ac.ir

adults with normal hearing without making any effort, but not in a conversational environment that is very difficult to hear. Nevertheless, this phenomenon does not easily take place in children, older adults, and hearing-impaired people [3].

Masking is the main topic discussed in the cocktail party situation. The cocktail party scenario presents a perceptual challenge to the extent that the energy in the competing sound sources overlaps with that of the target signal in time and frequency. This type of interference is often referred to as energetic masking (EM) [3,4]. Wightman et al. [3] have indicated that, a different type of masking occurs qualitatively and quantitatively when the masker components randomly occur from one stimulus presentation to another presentation, or when there are many similarities between the target and competing stimuli. This type of masking which is resulted from the uncertainty of masker or the similarity between target and competing stimuli, is referred to as informational masking (IM).

The present study aimed to comprehensively investigate the concept of IM in terms of historical perspectives and definitions, the measurement methods, different cues of releasing from IM, age-related effects, IM in people with hearing loss, and the neural basis of IM.

Historical perspectives and definitions

Wegel and Lane were among the first scholars who classified masking. In one of their studies on masking, they mentioned that there are two types of masking: peripheral and central. The central masking which is generally small, is caused by conflict of sensations in the brain, and the peripheral masking is resulted from overlapping of stimuli in the end organ. The central masking may be always present to some extent, while the peripheral masking only occurs when two sounds stimulate a similar area of the basilar membrane. They finally stated that, all the large amounts of masking may be attributed to peripheral masking [5]. Later, in a study by Carhart et al. in the 1960s on speech recognition, it was concluded that perception of an oral message in the presence of multiple competing messages consisting of

meaningful speech is more difficult than the non-verbal maskers. They referred to this additional interference as perceptual masking, and stated that the term "cognitive interference" could be applied as well [6]. Later, Pollack used the terms IM and EM to describe peripheral and central masking. At a spring meeting of the Acoustic Society of America (AAA), Pollack claimed that IM is a form of masking that cannot be attributed to EM [7]. In 1987, Neff and Green examined uncertainty factor in competing sounds and its role in the formation of IM [8]. Since the introduction of IM, many studies have been conducted on it. Leek et al. argued that IM is defined as the degradation of a target detection embedded in a series of similar sounds, and it is not related to the physical interaction between the target signal and masker [9].

Informational masking measurement methods

The multitone masking paradigm has been used in the early studies on IM, and speech materials were then used to measure IM.

Multitone masking paradigm

The simultaneous multitone masking paradigm was introduced by Neff and Green. The masker tones which are randomly presented in terms of frequency are presented in two ways; once masker tones are presented alone, and once along with the target tones [8]. A protected frequency range around the target tone is defined and the frequencies of masker tones are not allowed to fall within this range. It is a critical band around the target tone. The purpose of this protected zone is to minimize the amount of EM. The initial question of Neff and Green was: how many masker tones are required to create noise? Different tones in a Gaussian noise that were consecutively occurring in a frequency bandwidth, were used in order to generate N random samples out of masker tones at each interval [8,10]. In all of the samples, the frequency of the target tone was fixed and determined through each block of trials. Their results obtained from three constant target frequencies of 250, 1000, and 4000 Hz was very interesting. Based on the critical-band energy-detector model, it was

expected to observe a very little masking in masker tones composed of small numbers of components whose frequency was outside the critical band of the target tones, since it is very unlikely that the masker components can fall near the frequency of the target tone to energetically mask it. There were significant amounts of masking (i.e. about 50 dB for both 1 and 4 kHz frequencies), where the components of masker tones were very low (about 10 components). In addition, the maximum amount of masking did not occur for maskers containing the most components (i.e. true Gaussian noise) [8].

When there are very few masker components (e.g. less than 10 components), IM is predominant; in contrary, with the increase in the number of masker components (e.g. more than 100 components), the observed masking is almost EM [8,10]. There is a plateau region between these conditions which means that 10–100 components are perceptually enough to produce significant uncertainty (which is an important factor in IM), while also can produce a significant amount of EM [10].

Kidd et al. invented another method by using the multitone masking paradigm. They used this method to demonstrate that perceptual segregation of target and competing stimulus can reduce IM. They also used a sequence of multitone bursts in each interval. There were two types of masker sequences in the proposed method including multiple-bursts same (MBS), and multiple-bursts different (MBD) [10]. The MBS is consisted of eight bursts with frequencies randomly selected from the frequency range of 200–5000 Hz (not within the critical band of the target tone) and one target tone with a constant frequency of 1000 Hz. The frequency characteristics of masker tones remain constant at all intervals [10,11]. The MBD mode includes eight bursts that are randomly selected from the mentioned frequency range, except that their sequences are not similar between the intervals and in each new burst, eight intervals are selected. In MBS, the behavior of the target tone frequency is also similar to that of the maskers and it fluctuates at each interval, however, the frequency of the target tone remains constant in MBD. In other words,

the MBS mode indicate the same behavior of the target and the masker tones, and the MBD mode shows the different behavior of these two tones [10]. Overall, it can be said that: 1) in MBS mode, the jitter produced at the target tone frequency leads to the formation of a new auditory stream and reduces the amount of masking to 20 dB; 2) in MBD mode, the jitter added to the target tone frequency leads to an increase in the amount of the masking by about 10 dB. Therefore, the amount of produced IM is highly dependent on the similarity between the target and masker tones; 3) in MBD, compared to MBS, changing the masker frequency and maintaining the target tone frequency form an auditory stream. Therefore, the amount of masking resulted from MBS mode is about 25 dB greater than that obtained by MBD mode. Hence, it can be stated that the masking caused by the MBS mode indicates the importance of perceptual grouping for releasing from IM [10,11].

Informational masking and speech recognition

Speech perception of a particular speaker in the presence of different speakers is a very complex process, and one's perception can be affected by several factors. Given the importance of speech perception in challenging noisy environments, the measurement of IM has been investigated in several studies on speech recognition [4,12–15]. Masking experiments consisting of multiple speakers, using a forced-choice speech recognition test in a closed-set, which was first used by Brungart [13], are the most effective experiments to separate the target signal from the competing maskers. Coordinate response measure (CRM) is one of these experiments which was first developed by Bolia et al. [12]. This corpus consists of fixed paradigm sentences in the format of "ready (call sign) go to (color, number) now". In these sentences, eight call signs, four colors and eight numbers [4] can be used. These sentences are spoken by four females and four males. Therefore, 256 sentences can be recorded by each speaker (8 call signs \times 4 colors \times 8 numbers). In this test, the listener is required to identify the color and number in a sentence with a given specific call sign. Competitive random masker

sentences use the different call signs, colors, and numbers from the target sentence [4,13]. Although sentences with the CRM paradigm had been initially designed to measure speech intelligibility in the presence of competitive noise and in military environments such as the US Air Force, the call-sign-based structure used in them is very appropriate for multi-talker listening tasks [4,13]. CRM sentences are context free which means that it is not possible to predict the color or number in these sentences, and this is an important factor in measuring the amount of IM [13].

In early studies on CRM, the live talkers have been used as CRM corpus. Due to the need for precise control on stimulus onset, the digital recordings of talkers is preferable to live talkers [12]. Currently, in addition to the American version of the CRM introduced by Bolia et al. [12], there are two other English versions using speakers with British accent. In the first British version, the American version of sentences has been used, but in the second one which was designed to develop a speech recognition test for evaluating the speech perception of UK Armed Forces by Semeraro et al. [14], a new version of CRM was developed and its validity and reliability were measured. They used 18 disyllabic call signs from the North Atlantic Treaty Organization (NATO) phonetic alphabet, 9 monosyllabic colors, and 9 monosyllabic numbers [14]. Brungart [13] was one of the first researchers who used CRM paradigm sentences for studying IM. He showed that, when using a competing speaker, there are large differences in performance based on the similarity between the target and competing speakers such that if using speakers with different genders, the least similarity will be observed, but in case of using two similar speakers (one as target speaker and other as competing speaker), the most similarity can be obtained. The method proposed by Brungart has been used in a wide range of studies to investigate the effect of the number of competing speakers and their similarities, binaural and spatial processing, and the differences between speech and non-speech maskers [13]. One of the interesting findings was presented by Brungart

and Simpson [16]. They measured individual's scores using the CRM test and under the conditions that target speech was presented to one ear and competing speech to the opposite ear. The contralateral competing speech did not affect the performance. In the next step, both target and competing speech materials were presented to the same ear. In this case, the detection of target speech became difficult depending on the ratio of target speech to competing speech. In the last step, a target speech was presented to one ear, one competing speech was presented to the contralateral ear, and one unrelated sentence was simultaneously presented to the same ear as target speech. In this case, the individual performance was worse than that using the two previous conditions. Later, Brungart and Simpson in another study indicated that the uncertainty resulted from the semantic content of ipsilateral competing speech is much more important than the content of the contralateral competing speech [17]. They concluded that this effect is related to the listener's limited ability to hold separate inputs arrived at the both ears while performing an ipsilateral separation task.

Different cues of releasing from informational masking

In a classic view, the auditory system can be considered as a set of tandem band-pass filter where one can select one or multiple filters and ignore the output of other irrelevant filters. If it was true and the listener could do it perfectly, then the masking observed in the multitone masking tests would be only EM and much less than the empirical observation. Is that possible that the listeners can be trained to do this task, or IM be completely eliminated through increasing the prior knowledge? It seems that the variability between individuals in terms of being susceptible to IM is much greater than that to EM [9,18].

Training

It seems that, in spite of hundreds of trials in one session of the multitone masking experiments and familiarizing the individuals with the test procedure, there is still no decrease in masked thresholds. There has been no systematic

evaluation on the extent of the multitone masking training over a long period of time [19]. Most studies have suggested a training period for the individuals to learn how to respond [18,20-23]. Some conclusions can be extracted from these studies. Neff and Callaghan used the multitone masking method on four people. They used multitone maskers comprised of 2 and 10 masking components. They plotted individual's thresholds as a function of the number of trials. In one out of the four evaluated subjects, there was a large decrease in masked thresholds in the method with two masking components during the first 100 trials. This subject also showed a slight decrease in threshold in maskers with 10 components. However, there was no significant reduction in threshold in other three subjects. They concluded that there are many individual differences in being susceptible to IM that cannot be overcome by repeating a simple training [20].

There is another similar finding reported by Neff and Dethlefs by examining the results of five individuals who had been identified to have a high susceptibility to IM. They stated that there is no relationship between susceptibility to IM and the amount of improvement over time. In some of the studied subjects, there was a significant improvement over time as a result of extensive training. They concluded that there is a potential for long-term training in some people to reduce the susceptibility to IM (especially in maskers with small number of stimulus components). However, for the most people, the performance is stable over time, regardless of whether their threshold is high or low [18]. Regarding this assumption, Oxenham et al. investigated the susceptibility to IM in experienced musicians. They concluded that the trained musicians are less susceptible to IM than the non-trained peers [21]. Swaminathan et al. also showed that the history of playing music can affect the release from IM [22]. In another study, Dai et al. measured the effect of training on susceptibility to IM among 24 normal hearing young people. The competing stimuli were vocoded by a computer and presented written on the computer screen, and then the speech recognition test was carried out. They found that after familiarizing

subjects with the content of competing noise, the more IM was occurred. So the important point that can be drawn from their study is that the more meaning of the competing speech may result in more IM [23].

Cuing

Compared to extensive training, the most important way for attracting the listener's attention to the location of the target stimulus is providing a copy of the target stimulus immediately before its presentation. By performing cuing, any variation is probably reduced and a strong sensory memory is formed from detecting the pitch of target stimulus for comparing with the sound that one hears later [24,25]. Richards and Neff conducted a comprehensive experiment on the effects of cuing. They used the multitone masking method. By presenting a fixed and specific target frequency in a set of maskers with random frequency and providing a cue (presenting a copy of the target), they observed that the amount of masking between subjects was decreased by about 5 dB in average. There were even greater effects (about 20 dB) when the frequency of both target and competing masker were randomized [24]. Richards et al. compared the effects of target, masker, and target plus masker cues when they were presented before and after the trial. The interesting result was that the individuals had better performance when presenting the masker alone and immediately before the trial. Although it seems irrational that the individuals had a better threshold by cuing the masker, it is a random change in the masker that causes uncertainty in the individual. This effect is not due to the peripheral processes (e.g. adaptation of auditory nerve) but instead it appears to have a central origin [25]. This effect may be because the listener is able to form a notch-rejection filter based on the masker faster than creating an acceptance filter for the target stimulus. Durlach et al. referred to this listener strategy as Listener Min (in contrast to the accepted filter strategy that was called Listener Max), where the Max indicates a listener who increases the target-to-masker ratio (TMR) by maximizing the target while the Min indicates a listener who attempts

to do it by minimizing the competing masker. This potential strategy is more general than the cuing and it has been used in many masking experiments. Certainly, there should be also other strategies with no available information [11].

Perceptual segregation of sounds

There are many studies that have used a variety of methods to illustrate the effects of perceptual segregation on reducing the IM [10,18,26,27]. In fact, it can be stated that IM inherently occurs due to failure in segregating the target from the masker. The logic behind this statement is that if the target is received as an auditory object separate from the competing stimulus, then one can definitely detect it [10]. Some studies conducted on supra-threshold tasks have argued on this issue. They believe that the target is not always required to be perceptually segregated from the competing maker for a detection process, and it often occurs only with the listener's awareness of the presence of the target. However, what is clear is that in situations where there is high IM, the detection process is often clearly improved by segregating the target from other competing maskers [10,27]. Following, we review some studies supporting the role of perceptual segregation phenomenon for overcoming IM.

Neff and Dethlefs, by manipulating stimulus properties in the multitone masking method, investigated their role in the formation of perceptual segregation. They used the following manipulations in order to modify the stimulus properties and consequently, the occurrence of perceptual segregation: a) reducing the duration of target compared to that of masker (10- and 100-ms presented in a 200-ms masker); b) dichotic presentation of target and masker; and c) creating a difference between qualitative characteristics of target (narrowband noise) and competing stimulus (multitone masker) [18]. In general, the amount of release from IM was reduced by increasing the number of masker components up to 8–10. In fact, except for the dichotic condition, there was no release from masking in maskers with more than 100 components. The benefit of perceptual segregation for IM is much greater than for EM.

Generally, the proportion of EM to IM is increased by increasing the number of masker components from very few to very high (e.g. 100 components), hence, there is little IM for 100-component maskers [18]. Durlach et al. also used other manipulations for the perceptual segregation, such as asynchronous initiation of target and competing stimuli, frequency sweep of target in the opposite direction to the frequency sweep of competing stimulus, dichotically presentation of target and competing stimulus, and two types of spectro-temporal paradigm that changed the relative coherence between target and competing stimulus. Despite the high difference between these conditions in the studied subjects, an average segregation of about 17 dB in all of the above conditions was reported. Finally the authors concluded that the factor of similarity between the target and competing stimulus plays an important role in releasing from IM [26].

Auditory spatial processing

Since the introduction of cocktail party problem, there have been some studies for detection, discrimination, and recognition of the target in the presence of competing sounds that are in the same direction with the target or spatially separated from it. Almost in all cases, subjects performed better by spatial separation of the target from competing stimuli compared to when the sounds are co-located [28-30]. This improvement in performance is referred to as spatial release from masking which is usually resulted from the important processes such as binaural processing and analyses [28]. The binaural processing enables the listener to spatially separate the target and competing sounds and, consequently, to analyze the auditory scene after localizing the source of target. Auditory streams are formed by analyzing the auditory scene, and eventually separating the target from the competing stimulus [29]. The binaural processing usually uses cues such as time difference in receiving the sounds by the ears or level difference to locate the sound source. In binaural analyses such as the Equalization-Cancellation process, either the representation of the target is amplified or the representation of competing stimulus is

weakened. This also improves the spatial release from IM [28,29]. It should be noted that, although auditory spatial processing leads to release from both EM and IM, the amount of release resulted from spatial separation of the target and competing stimulus in IM is much greater than in EM. Therefore, according to the results of Yost et al., it can be said that the most important cue in the release from IM is spatial separation of target and competing stimulus [28]. The next interesting point is that, in young adults with normal hearing, in case of a high similarity between the target and competing stimuli, there will be a greater spatial release from masking compared to the low-similarity case [30].

Other cues of release from informational masking

Brungart and Simpson in their study concluded that, in case of using speakers with different genders to utter target and competing sentences, there would be less IM compared to the case where talkers had same gender. This is probably related to the similarity or difference in speech frequency and rate between speakers [16]. Freyman et al. examined the amount of IM in talkers with the same or different nationality and concluded that, in case of using talkers and listeners with the same nationality, there would be more masking because of the greater similarity between the target and competing sentences. It was shown that, the accent of German speakers uttering English sentences was a sign of a decreased masking [31]. Carhart et al. found that perceptual masking is significantly dependent on the number of competing talkers. They found that, perceptual masking was increased as the number of competing talkers increased to three and decreased as their numbers increased more than three [6]. Yost et al. reported that, in case of using three talkers, there would be greater difficulty in dividing attention to the target in comparison with the two-talker case [32]. Moreover, in three-talker cases, spatial cuing is more helpful for releasing from masking. Brungart et al. found that, in diotic condition using two or three maskers, there are more masking with high TMR compared to the case with one masker, and this might be due to the increase in both EM and IM

[4]. Hall et al. in studying the ability of children and adults, also reported that there are much more masking in the presence of two competing talkers [33].

Effect of age on informational masking

A number of studies have examined the role of age in focusing attention on one auditory source and ignoring other sources.

Children

Spatial processing and paying attention to a particular sound source in the presence of other sound sources is crucial for children to success in educational and social settings. It has been shown that children are highly distracted by competing sound sources [26,27]. Allen and Wightman investigated the effects of uncertainty on target and masker frequency among children aged 3–5 years old and adults. They assumed that children were unable to focus attention on a specific frequency area. In their study, thresholds in children under all evaluated situations were higher than in adults. Moreover, they showed that the incidence of IM was higher in children when there was uncertainty in the frequency of masker. Overall, they concluded that children may be poor listeners probably due to the immaturity of central mechanisms in them which are responsible for focusing attention [34].

Oh et al. examined the masked thresholds using the multitone masking method for a group of children and adults. The thresholds in children varied due to the number of masker components and consequently, the type of available masking. This means that, in case of using more masker components (EM), there would be approximately similar masking amount in both groups; however, when there were few masker components, children had higher masking than adults. With the fewest masker components, there was a 50 dB difference between the two groups. These differences cannot be attributed to anatomical or physiological maturity or even the difference in auditory filter bandwidth of children. According to the aforementioned studies, the greater amount of IM in children may be due to their decreased selective attention abilities [35].

Whightman and Kistler, by studying the recognition thresholds of coordinate response measure (CRM) sentences in the presence of ipsilateral competing speech presented to both ears in a group of 4–16 years old children, concluded that children need higher TMR. There was about a 15 dB increase in IM amount of children. It is amount increased by about 5 dB in both groups by presenting contralateral competing speech. They concluded that selective attention abilities in children are poorer than in adults, and children are more susceptible to IM [36]. It is required to evaluate the performance of older groups in order to investigate the susceptibility to IM throughout life.

Elderly people

Most of the elderly people complain about speech perception in noisy situations. It is clear that their poor speech recognition can be due to different factors such as peripheral hearing loss and cognitive impairment or processing problems. Determination of the role of each of these factors in speech perception problems of the elderly listeners is difficult. Most of the elderly people with normal hearing have complaints about speech perception [37]. To investigate the effects of aging on IM regardless of hearing impairment, Rajan and Cainer used 20–69 year-old subjects with normal hearing in a frequency range of 250–4000 Hz. These participants were divided into five age groups. Speech in noise perception test was performed on the subjects using sentences and two types of babble noise and speech-shaped noise. Speech-shaped noise was used to create EM and bubble noise, consisting of eight speakers, was used to create both EM and IM. Sentence recognition scores using babble noise for the oldest age group (59–69 years) were significantly lower than those of other age groups. In case of using a masker causing more cognitive load, people over the age of 60 years needed higher SNR for detection of 50% of sentences. They concluded that people over 60 years of age had impairment in the separation of target speech from background noise due to a modality-specific decline in cognitive processing and, consequently, the

decreased ability to use acoustic and phonetic cues [38]. In a recent study conducted by Amiri et al. on two groups of young and elderly normal hearing listeners, found that the speech perception ability of elderly listeners is considerably reduced in the presence of meaningful background noise. Moreover, the decrease in SNR significantly reduces the perceptual abilities of elderly listeners [39].

Since IM causes higher cognitive load in comparison with EM, many scholars have concluded that older people have generally lower speech perception than younger people (e.g. Ben-David et al. [40] Füllgrabe et al. [41]). In all of these studies, the speech perception of the two groups of young (less than 30 years old) and older (older than 60 years) people has been compared. Therefore, to answer the question whether IM affects middle-aged people, Goossens et al. [42] investigated speech perception in people aged 20–80 years divided them into three age groups of young, middle-aged, and older cohorts. All of the subjects had normal hearing (thresholds up to 4000 Hz) with no cognitive impairment. Their main hypothesis was that, although there is similar amount of EM in all age groups, there should be a significant decrease in speech perception of the older group if using IM. They concluded that, in both EM and IM, older age group had poorer speech perception compared to the young group, but this increase in IM was significantly greater than in EM. Another important point was that in case of EM, the hearing impairment was the most important factor in speech perception, while in IM the decline in speech perception was independent of hearing impairment. They considered the effects of aging on poor speech perception in older adults was due to deficiency in temporal processing, and emphasized on the improvement of the quality of rehabilitation programs to prevent older adults from missing conversations [42]. In another study conducted in an investigating the speech recognition performance of young and elderly people with normal hearing, it was reported that, the more meaningful the competing noise (competing noise was selected in three ways: part of the speech text, a bubble noise, and single

person competing noise), the more the IM would occur. In this study, the poorer performance of the elderly people was attributed to the age-related decline in their cognitive abilities in word recognition [43].

Some studies have examined the role of age in detecting target speech from competing speech. According to these studies, the normal hearing elderly people usually have more problems in case of high similarity between target and competing speech [44,45]. In 2016, a group of young, middle-aged, and elderly participants were asked to answer the question whether the separation of target speech from competing speech that were simultaneously given to both ears can lead to more cognitive processing compared to when competing speech is presented to one ear, or whether aging affects this ability or not. The highest difference between the young and elderly listeners was observed in case of the presentation of bilateral symmetrical competing stimuli that were very similar to the target speech. Compared to the young group, elderly listeners needed higher SNR to recognize target speech. In case of presenting competing stimuli to only one ear, there was almost similar release from IM for all three age groups [45].

Informational masking and hearing loss

It is reasonable to assume that people with peripheral hearing impairment may show different susceptibility to IM or different use of cues to release from this type of masking, compared to those with normal hearing. Now the question is that to what extent these differences can be attributed to processes involved in peripheral mechanisms such as reduced sensitivity, compression, or broader auditory filters [46,47]. Micheyl et al., and Alexander and Lutfi used the multitone masking method in their studies on people with sensory-neural hearing impairment. There are differences between the two studies, but their general conclusion was that sensory-neural hearing impairment does not increase the susceptibility of individuals to IM and hearing-impaired people can even have less IM compared to people with normal hearing [46,47]. According to Alexander and Lutfi, the CoRE model can

predict this result due to the reduced dynamic range of hearing-impaired people [47]. Based on this model, it is suggested that the amount of IM will increase by increasing the output variability of the peripheral filters. Most of this variability is normally resulted from the difference between the output levels in the trials where the masker components are located in the filter and those where the masker components are not located in the filter. In the second mode, the filter output will be very low. If the individual's thresholds are increased as a result of hearing impairment and the range of filter output level is reduced, the estimated IM amount is also reduced [48]. Overall, it can be concluded that, due to the broader auditory filters in hearing-impaired people, the broadband spectrum stimuli may interfere with each other more at the peripheral auditory system and thus, they have more EM. On the other hand, in the mentioned studies it was illustrated that, the amount of IM is more likely decreased by increasing EM induced by a series of excitatory stimuli [46,47].

Neural basis of informational masking

There are limited studies on the neural basis of IM. The role of pre-attentive central auditory processing in susceptibility to IM was first addressed by Brungart and Simpson. They reported that, when presenting one ipsilateral competing stimulus and one speech-like contralateral competing stimulus at the same time, or by changing the amplitude of the contralateral competing stimulus, there was more intervention in speech perception compared to when only one ipsilateral competing stimulus is presented. This process considered as a bottom-up processing [16]. Scott et al. [49] used positron emission tomography (PET) to investigate the neural basis of speech perception in the presence of steady-state noise (EM) and a competing speaker (IM). By investigating brain activity at different SNRs, they concluded that, as Brungart et al. [4] predicted using behavioral tests, noise and speech act separately in case of being used as masker; hence, the neural basis of IM is different from that of EM. In summary, they concluded that the rostral and dorsolateral parts of the prefrontal

cortex and posterior part of the parietal cortex are applied in EM, while in the IM, the bilateral superior temporal gyrus is activated in addition to the above parts indicating the parallel information processing resulted from the two talkers. Using three types of competing noise, Szalárdy et al. [50] investigated the results of functional magnetic resonance imaging (fMRI) on 15 young French people with no hearing or lingual impairment. These participants were asked to listen to a list consisting of 30 words in four conditions including a) in silence, b) in the presence of broadband noise (EM), c) in the presence of a bubble noise consisting of 4 speakers (IM), and d) in the presence of the previous bubble noise presented in reverse (more IM than condition b and lower IM than condition c), and then choose the word they heard from the two words displayed on the screen. They concluded that, the left auditory cortex, the bilateral superior temporal gyrus, and the left supramarginal gyrus are activated in IM (at both phonological and lexico-semantic levels). By reducing the fMRI response under condition d from that of condition c (i.e. the condition containing the lexicosemantic level), more activity was observed in the 21st speech region while in case of reducing the fMRI response obtained under condition d from that of condition b (i.e. the condition containing the phonological level), there was more activity in the 22nd speech region. Therefore, it can be said that the most part of bilateral superior temporal gyrus is involved in IM [50].

Carlile and Corkhill provided a new insight into the processes involved in IM by investigating it in the presence of various types of competing noises (speech noise, speech-like noise, and amplitude-modulated noise). They argued that both top-down and bottom-up processes are involved in IM. This indicates that both endogenous and exogenous attentions are engaged in the release from IM. Endogenous attention is a top-down processing, while exogenous attention is described as a bottom-up processing. In exogenous attention, in case of using a stimulus with salient information, the listener pays attention to an object or a stream that is not appropriate for the task. This can be the basis of a phenomenon

called odd-sex distractor, where by the use of a female masker talker in the presence of a male target talker and another male masker talker, more IM is produced compared to when all talkers have the same gender. This phenomenon is caused particularly by a salient stimulus and is an example of the role of upstream processing in creating IM. Therefore, IM cannot be a pure product of downstream processing [51].

Conclusion

Masking is the main topic discussed in cocktail party problem and is divided into two types of EM and IM. Overall, IM resulted from limitations at the central level of auditory system unlike EM that is resulted from the limitations caused by frequency selection at the peripheral level. IM leads to a failure in selection of auditory objects and therefore, cause impairment in auditory scene analysis. Adult people with normal auditory system usually use different cues for releasing from IM and tracking speech in noisy environments. However, this phenomenon does not occur easily in children, elderly people and those with impaired hearing. Although there are several neural bases for IM, it seems that both top-down and bottom-up processing are involved in its formation.

Conflict of interest

The authors declared no conflicts of interest.

References

1. Anderson S, Kraus N. Objective neural indices of speech in noise perception. *Trends Amplif.* 2010;14(2):73-83. doi: [10.1177/1084713810380227](https://doi.org/10.1177/1084713810380227)
2. Cherry EC. Some experiments on the recognition of speech, with one and two ears. *J Acoust Soc Am.* 1953;25(5):975-9. doi: [10.1121/1.1907229](https://doi.org/10.1121/1.1907229)
3. Wightman FL, Kistler DJ, O'Bryan A. Individual differences and age effects in a dichotic informational masking paradigm. *J Acoust Soc Am.* 2010;128(1):270-9. doi: [10.1121/1.3436536](https://doi.org/10.1121/1.3436536)
4. Brungart DS, Simpson BD, Ericson MA, Scott KR. Informational and energetic masking effects in the perception of multiple simultaneous talkers. *J Acoust Soc Am.* 2001;110(5 Pt 1):2527-38. doi: [10.1121/1.1408946](https://doi.org/10.1121/1.1408946)
5. Wegel RL, Lane CE. The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. *Physics Rev.* 1924;23(2):266-85.
6. Carhart R, Tillman TW, Greetis ES. Perceptual masking in multiple sound backgrounds. *J Acoust Soc Am.* 1969;45(3):694-703. doi: [10.1121/1.1911445](https://doi.org/10.1121/1.1911445)

7. Pollack I. Auditory informational masking. *J Acoust Soc Am.* 1975;57(S1). doi: [10.1121/1.1995329](https://doi.org/10.1121/1.1995329)
8. Neff DL, Green DM. Masking produced by spectral uncertainty with multicomponent maskers. *Percept Psychophys.* 1987;41(5):409-15. doi: [10.3758/bf03203033](https://doi.org/10.3758/bf03203033)
9. Leek MR, Brown ME, Dorman MF. Informational masking and auditory attention. *Percept Psychophys.* 1991; 50(3):205-14. doi: [10.3758/bf03206743](https://doi.org/10.3758/bf03206743)
10. Kidd G Jr, Mason CR, Deliwala PS, Woods WS, Colburn HS. Reducing informational masking by sound segregation. *J Acoust Soc Am.* 1994;95(6):3475-80. doi: [10.1121/1.410023](https://doi.org/10.1121/1.410023)
11. Durlach NI, Mason CR, Kidd G Jr, Arbogast TL, Colburn HS, Shinn-Cunningham BG. Note on informational masking. *J Acoust Soc Am.* 2003;113(6):2984-7. doi: [10.1121/1.1570435](https://doi.org/10.1121/1.1570435)
12. Bolia RS, Nelson WT, Ericson MA, Simpson BD. A speech corpus for multitalker communications research. *J Acoust Soc Am.* 2000;107(2):1065-6. doi: [10.1121/1.428288](https://doi.org/10.1121/1.428288)
13. Brungart DS. Informational and energetic masking effects in the perception of two simultaneous talkers. *J Acoust Soc Am.* 2001;109(3):1101-9. doi: [10.1121/1.1345696](https://doi.org/10.1121/1.1345696)
14. Semeraro HD, Rowan D, van Besouw RM, Allsopp AA. Development and evaluation of the British English coordinate response measure speech-in-noise test as an occupational hearing assessment tool. *Int J Audiol.* 2017; 56(10):749-58. doi: [10.1080/14992027.2017.1317370](https://doi.org/10.1080/14992027.2017.1317370)
15. Shinn-Cunningham B. Models of plasticity in spatial auditory processing. *Audiol Neurootol.* 2001;6(4):187-91. doi: [10.1159/000046830](https://doi.org/10.1159/000046830)
16. Brungart DS, Simpson BD. Within-ear and across-ear interference in a cocktail-party listening task. *J Acoust Soc Am.* 2002;112(6):2985-95. doi: [10.1121/1.1512703](https://doi.org/10.1121/1.1512703)
17. Brungart DS, Simpson BD. Within-ear and across-ear interference in a dichotic cocktail party listening task: effects of masker uncertainty. *J Acoust Soc Am.* 2004; 115(1):301-10. doi: [10.1121/1.1628683](https://doi.org/10.1121/1.1628683)
18. Neff DL, Dethlefs TM. Individual differences in simultaneous masking with random-frequency, multicomponent maskers. *J Acoust Soc Am.* 1995;98(1):125-34. doi: [10.1121/1.413748](https://doi.org/10.1121/1.413748)
19. Kidd G Jr, Mason CR, Richards VM, Gallun FJ, Durlach NI. Informational masking. In: Yost WA, Popper AN, Fay RR, editors. *Auditory perception of sound source.* 2008. p. 143-89.
20. Neff DL, Callaghan BP. Effective properties of multicomponent simultaneous maskers under conditions of uncertainty. *J Acoust Soc Am.* 1988;83(5):1833-8. doi: [10.1121/1.396518](https://doi.org/10.1121/1.396518)
21. Oxenham AJ, Fligor BJ, Mason CR, Kidd G Jr. Informational masking and musical training. *J Acoust Soc Am.* 2003;114(3):1543-9. doi: [10.1121/1.1598197](https://doi.org/10.1121/1.1598197)
22. Swaminathan J, Mason CR, Streeter TM, Best V, Kidd G Jr, Patel AD. Erratum: Musical training, individual differences and the cocktail party problem. *Sci Rep.* 2015; 25(5):14401. doi: [10.1038/srep14401](https://doi.org/10.1038/srep14401)
23. Dai B, McQueen JM, Hagoort P, Kösem A. Pure linguistic interference during comprehension of competing speech signals. *J Acoust Soc Am.* 2017;141(3): EL249. doi: [10.1121/1.4977590](https://doi.org/10.1121/1.4977590)
24. Richards VM, Neff DN. Cuing effects for informational masking. *J Acoust Soc Am.* 2004;115(1):289-300. doi: [10.1121/1.1631942](https://doi.org/10.1121/1.1631942)
25. Richards VM, Haung R, Kidd G Jr. Masker-first advantage for cues in informational masking. *J Acoust Soc Am.* 2004;116(4 Pt 1):2278-88. doi: [10.1121/1.1784433](https://doi.org/10.1121/1.1784433)
26. Durlach NI, Mason CR, Shinn-Cunningham BG, Arbogast TL, Colburn HS, Kidd G Jr. Informational masking: counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *J Acoust Soc Am.* 2003;114(1):368-79. doi: [10.1121/1.1577562](https://doi.org/10.1121/1.1577562)
27. Kidd G Jr, Mason CR, Arbogast TL. Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns. *J Acoust Soc Am.* 2002;111(3):1367-76. doi: [10.1121/1.1448342](https://doi.org/10.1121/1.1448342)
28. Yost WA. Erratum: Spatial release from masking based on binaural processing for up to six maskers. *J Acoust Soc Am.* 2017;141(4):2473. doi: [10.1121/1.4979981](https://doi.org/10.1121/1.4979981)
29. Ahveninen J, Kopčo N, Jääskeläinen IP. Psychophysics and neuronal bases of sound localization in humans. *Hear Res.* 2014;307:86-97. doi: [10.1016/j.heares.2013.07.008](https://doi.org/10.1016/j.heares.2013.07.008)
30. Misurelli SM, Litovsky RY. Spatial release from masking in children with bilateral cochlear implants and with normal hearing: effect of target-interferer similarity. *J Acoust Soc Am.* 2015;138(1):319-31. doi: [10.1121/1.4922777](https://doi.org/10.1121/1.4922777)
31. Freyman RL, Balakrishnan U, Helfer KS. Spatial release from informational masking in speech recognition. *J Acoust Soc Am.* 2001;109(5 Pt 1):2112-22. doi: [10.1121/1.1354984](https://doi.org/10.1121/1.1354984)
32. Yost WA, Dye RH Jr, Sheft S. A simulated "cocktail party" with up to three sound sources. *Percept Psychophys.* 1996;58(7):1026-36. doi: [10.3758/bf03206830](https://doi.org/10.3758/bf03206830)
33. Hall JW 3rd, Buss E, Grose JH. Informational masking release in children and adults. *J Acoust Soc Am.* 2005; 118(3 Pt 1):1605-13. doi: [10.1121/1.1992675](https://doi.org/10.1121/1.1992675)
34. Allen P, Wightman F. Effects of signal and masker uncertainty on children's detection. *J Speech Hear Res.* 1995;38(2):503-11. doi: [10.1044/jshr.3802.503](https://doi.org/10.1044/jshr.3802.503)
35. Oh EL, Wightman F, Lutfi RA. Children's detection of pure-tone signals with random multitone maskers. *J Acoust Soc Am.* 2001;109(6):2888-95. doi: [10.1121/1.1371764](https://doi.org/10.1121/1.1371764)
36. Wightman FL, Kistler DJ. Informational masking of speech in children: Effects of ipsilateral and contralateral distractors. *J Acoust Soc Am.* 2005;118:3164-76. doi: [10.1121/1.2082567](https://doi.org/10.1121/1.2082567)
37. Jarollahi F, Amiri M, Jalaei S, Sameni SJ. The effects of auditory spatial training on informational masking release in elderly listeners: a study protocol for a randomized clinical trial. Version 2. *F1000Res.* 2019; 8:420. doi: [10.12688/f1000research.18602.2](https://doi.org/10.12688/f1000research.18602.2)
38. Rajan R, Cainer KE. Ageing without hearing loss or cognitive impairment causes a decrease in speech intelligibility only in informational maskers. *Neuroscience.* 2008;23;154(2):784-95. doi: [10.1016/j.neuroscience.2008.03.067](https://doi.org/10.1016/j.neuroscience.2008.03.067)
39. Amiri M, Jarollahi F, Jalaie S, Sameni SJ. A new speech-in-noise test for measuring informational masking in speech perception among elderly listeners. *Cureus.* 2020; 12(3). e7356. doi: [10.7759/cureus.7356](https://doi.org/10.7759/cureus.7356)
40. Ben-David BM, Tse VY, Schneider BA. Does it take older adults longer than younger adults to perceptually segregate a speech target from a background masker? *Hear Res.* 2012;290(1-2):55-63.

- doi: [10.1016/j.heares.2012.04.022](https://doi.org/10.1016/j.heares.2012.04.022)
41. Füllgrabe C, Moore BC, Stone MA. Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition. *Front Aging Neurosci.* 2015;6:347. doi: [10.3389/fnagi.2014.00347](https://doi.org/10.3389/fnagi.2014.00347)
 42. Goossens T, Vercammen C, Wouters J, van Wieringen A. Masked speech perception across the adult lifespan: Impact of age and hearing impairment. *Hear Res.* 2017; 344:109-24. doi: [10.1016/j.heares.2016.11.004](https://doi.org/10.1016/j.heares.2016.11.004)
 43. Lu Z, Daneman M, Schneider BA. Does increasing the intelligibility of a competing sound source interfere more with speech comprehension in older adults than it does in younger adults? *Atten Percept Psychophys.* 2016;78(8): 2655-77. doi: [10.3758/s13414-016-1193-5](https://doi.org/10.3758/s13414-016-1193-5)
 44. Helfer KS, Freyman RL. Aging and speech-on-speech masking. *Ear Hear.* 2008;29(1):87-98. doi: [10.1097/AUD.0b013e31815d638b](https://doi.org/10.1097/AUD.0b013e31815d638b)
 45. Helfer KS, Merchant GR, Freyman RL. Aging and the effect of target-masker alignment. *J Acoust Soc Am.* 2016;140(5):3844. doi: [10.1121/1.4967297](https://doi.org/10.1121/1.4967297)
 46. Micheyl C, Arthaud P, Reinhart C, Collet L. Informational masking in normal-hearing and hearing-impaired listeners. *Acta Otolaryngol.* 2000;120(2):242-6. doi: [10.1080/000164800750001017](https://doi.org/10.1080/000164800750001017)
 47. Alexander JM, Lutfi RA. Informational masking in hearing-impaired and normal-hearing listeners: sensation level and decision weights. *J Acoust Soc Am.* 2004;116(4 Pt 1):2234-47. doi: [10.1121/1.1784437](https://doi.org/10.1121/1.1784437)
 48. Lutfi RA. A model of auditory pattern-analysis based on component-relative entropy. *J Acoust Soc Am.* 1993; 94(2 Pt 1):748-58. doi: [10.1121/1.408204](https://doi.org/10.1121/1.408204)
 49. Scott SK, Rosen S, Wickham L, Wise RJ. A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *J Acoust Soc Am.* 2004;115(2):813-21. doi: [10.1121/1.1639336](https://doi.org/10.1121/1.1639336)
 50. Szalárdy O, Tóth B, Farkas D, György E, Winkler I. Neural correlates of informational masking in the human brain in the cocktail party situations. *Front Psychol.* 2019;10:786. doi: [10.3389/fpsyg.2019.00786](https://doi.org/10.3389/fpsyg.2019.00786)
 51. Carlile S, Corkhill C. Selective spatial attention modulate bottom-up informational masking of speech. *Sci Rep.* 2015;5:8662. doi: [10.1038/srep08662](https://doi.org/10.1038/srep08662)