The Journal of International Advanced Otology

J Int Adv Otol 2016; 12(3): 282-9 • DOI: 10.5152/iao.2016.2540

Original Article

Recognition of Deep Band Modulated Consonants in Quiet and Noise in Older Individuals with and Without Hearing Loss

Hemanth Narayan Shetty, Vishal Kooknoor

Department of Audiology, All India Institute of Speech and Hearing, Karnataka, India (HNS)
Department of Speech and Hearing, Samvaad Institute of Speech and Hearing, Karnataka, India (VK)

OBJECTIVE: The purpose of the present study was to investigate the effects of temporal envelope enhancement using deep band modulation (DBM) on consonant identification scores (CIS) and transfer of features such as place of articulation, manner of articulation, and voicing.

MATERIALS and METHODS: Data were collected from four groups of ten participants each. These were grouped into younger (YNH) and older adult (ONH) individuals with normal hearing and younger (YHI) and older adult (OHI) individuals with hearing impairment who had mild to moderate sensorineural hearing loss. The CIS and transfer features for 21 vowel–consonant–vowel (VCV) syllables in unprocessed and DBM conditions with quiet and noise were obtained from each participant.

RESULTS: The results showed that consonant identification scores (CIS) in DBM conditions was significantly better than in unprocessed conditions in quiet and at 5 dB and 10 dB signal-to-noise ratios (SNRs). This was true in each group. Voicing was transmitted better than other features in each condition for all the groups except for the YHI group, for whom place of articulation was conveyed best in unprocessed conditions and manner of articulation was transmitted best in DBM conditions. Furthermore, in the YHI, ONH, and OHI groups, manner of articulation was conveyed better than place of articulation in both unprocessed and DBM conditions.

CONCLUSION: At reduced SNRs, cues from DBM enabled the listener to repeat the heard VCV syllables. The effect of aging and the combined effects of aging and hearing loss are partly lessened by DBM through enhancement of the manner of articulation in VCV syllables.

KEYWORDS: Older adults, speech perception, envelope enhancement

INTRODUCTION

Speech is a complex dynamic signal that fluctuates in both amplitude and frequency over time [1]. To perceive these inherent fluctuations in a signal, the auditory system encodes information directly from the peripheral to the central auditory level to infer a message. Normal perception depends on intact peripheral and central auditory processing mechanisms. However, in a realistic environment, noise always accompanies speech, resulting in a deleterious effect on spectral parameters [2]; also, noise obscures temporal modulation depth [3]. It is well established that temporal asynchrony is observed in older adults [4]; the deleterious effect of noise degrades the temporal envelope and spectral content of speech, resulting in poor perception. Because of this effect, even in the absence of hearing loss, older adults often complain that they can hear speech but find it difficult to follow information in the presence of noise [5]. This grievance is even more common in older adults who suffer from hearing loss [4]. The most common type of hearing loss seen in older adults is sensorineural hearing loss, in which dysfunction is noted in the cochlear system and/or in the neural system [6]. Hearing loss in these systems reduces vibration of the basilar membrane due to loss of sensory receptors [4], which leads to wider auditory filters [7, 8]. Thus, high intensity is required to reach the threshold of audibility where adjacent frequencies are stimulated. Additionally, there is a lack of synapse between sensory receptors and neural ganglions; this results in asynchronous neural firing. In certain circumstances, if noise dominates speech, it becomes more difficult for older adults to understand speech. This is because noise reduces temporal dips in speech and masks weak consonants [9]. Furthermore, noise obscures the spectral component of speech [2]. Distortion created by noise reduces available cues in speech; thus, these cues are not sufficient for the retrieval of information by older adults with hearing impairment.

To resolve the impaired perception faced by older adults, the temporal envelope enhancement strategy shows promise [10]. The temporal envelope is a low-frequency modulation that plays an instrumental role in the perception of speech when spectral and temporal cues are degraded by noise. Thus, enhancing the temporal envelope provides cues for the perception of speech. One such algorithm is deep band modulation (DBM), which enhances temporal modulation depth in speech.

Corresponding Address: Hemanth Narayan Shetty E-mail: hemanthn.shetty@gmail.com
Submitted: 16.04.2016 **Accepted:** 08.11.2016

In deep band modulation, the temporal envelope is enhanced, resulting in a significant increase in the modulation depth. The envelope is filtered and gains are provided differently such that for consonants, the gain is increased and for vowels, the gain is compressed. This algorithm increases the consonant-to-vowel ratio and reduces the upward spread of masking. In addition, at lower signal-to-noise ratios (SNRs), listeners perceive speech by higher amplitude modulation, which is less energetically masked by noise. Furthermore, rescaling the entire length of the timescale enables the auditory system to access correct lexical items. Nagarajan et al. [11] recruited learning disabled individuals with temporal asynchrony [10]; their results showed improved perception. In another study, Hemanth and Akshay [12] found that DBM improved the perception of speech in older adults. However, there is a dearth of literature on how DBM improves speech perception in individuals with age-related hearing loss. Thus, in the present study, it was hypothesized that DBM may compensate for temporal asynchrony in older adults with hearing loss and may also alleviate impaired perception of speech degraded by noise. The aim of the study was to investigate consonant recognition in unprocessed and DBM conditions by study participants in quiet and at different SNRs. The specific objectives of the study were a) to compare consonant identification scores between unprocessed and DBM conditions obtained at different SNRs in each group of study participants, b) to compare CIS between groups in unprocessed and DBM conditions obtained at each SNR, and c) to investigate sequential transfer of information in unprocessed and DBM conditions at each SNR from the study participants in each group.

MATERIALS and METHODS

Participants

A total of 40 participants were involved in the study. They were classified into four groups; each group comprised ten participants. Group I included younger normal hearing (YNH) adults (>18 years to <25 years); their pure tone thresholds were better than or equal to 15 dB HL at frequencies from 250 Hz to 8 kHz (in octaves) in both ears. Group III included younger hearing impaired (YHI) participants who had moderate sensorineural hearing impairment; they were required to have air conduction thresholds of 40 or 55 dB HL at frequencies from 250 Hz to 8 kHz in both ears. This ensured that the participants in Group III matched the participants in Group I in age. Group II comprised older normal hearing (ONH) participants (>60 to <75 years) whose hearing thresholds were better than or equal to 25 dB HL at frequencies from 250 Hz to 8 kHz (in octaves) in both ears. Group IV included older adults with hearing impairment (OHI) whose thresholds were similar to those of Group III and who were matched for age with Group II. The thresholds at frequencies from 250 Hz to 8 kHz (in octaves) in both ears from each group are depicted in Figure 1. All participants had word recognition scores of 75% or higher in quiet conditions in both ears. Additionally, the middle ear status of all participants was required to be normal, as indicated by type A tympanogram. None of the participants had any other otological or neurological problems.

In the present study, all testing procedures used non-invasive techniques, and all the procedures were explained to the participants before testing. Informed consent was obtained from all the participants. The study was approved by the ethical committee of the institution.

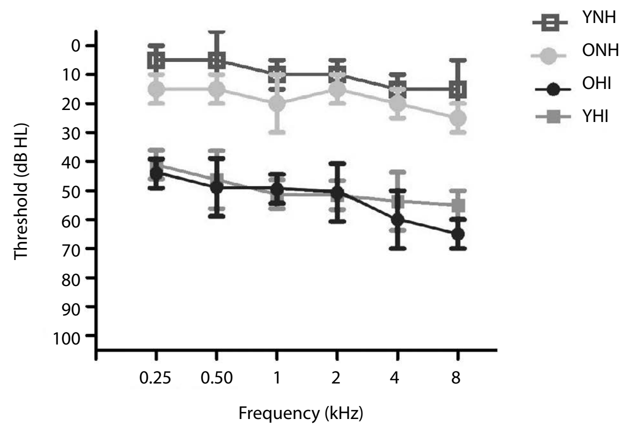

Figure 1. Audiometric threshold at each frequency from 250 Hz to 8 kHz in octaves for the four groups of study participants.

Stimulus Preparation

The phonemes in Kannada with frequencies of occurrence greater than 0.5% [13] were selected (/k/, /g/, /tʃ/, /t/, /d/, /ŋ/, /t/, /d/, /n/, /p/, /b/, /m/, /j/, /r/, /l/, /v/, /ʃ/, /s/, /h/, /l. /, and /dʒ/) and were paired with a low short central vowel, /i/, at the initial and final positions. The vowel–consonant–vowel (VCV) syllables were used to obtain consonant recognition and also to estimate the sequential transfer function. Three female speakers whose mother tongue was Kannada were chosen to utter the 21 VCV syllables with normal vocal effort. These VCV syllables were recorded using Adobe Audition software (Adobe Version 3; Adobe Systems Incorporated, Park Avenue, San Jose) via a recording microphone placed at a distance of 10 cm from the lips of the speaker [14]. The recorded stimuli were digitized using a 32-bit processor at 44,100 Hz sampling frequency. A goodness test was performed to verify the test stimuli. That is, the 21 stimuli uttered by the three speakers were presented to ten normal hearing individuals. The stimuli uttered by the speaker who was judged to have the most naturalness on a 3-point rating scale (3=natural, 2=less natural, and 1=unnatural) were selected to conduct the experiments.

Each VCV syllable was digitally mixed with speech babble noise[15] at a +10 dB SNR using the SNR MATLAB code (MATLAB version 2009B; Natick, Massachusetts, USA). The noise onset preceded the onset of a VCV syllable by 100 ms and continued until 100 ms after the end of each syllable. The noise was ramped using the cosine square function with a ramp duration of 30 ms. The onset of noise before the onset of each syllable is believed to prevent unintended onset effects. A similar procedure was conducted to digitally mix the VCV syllables with the speech babble noise [16] at a +5 dB SNR.

Deep band modulation versions of the VCV stimuli were prepared using Praat software (version 5.3.56, developed by the Institute of Phonetic Science, University of Amsterdam, Netherlands) using an algorithm adopted by Nagarajan et al. [11] Each VCV syllable was passed through 20 second-order Butterworth filters by the filter bank method. The center frequencies from these 20 filters were logarithmically spaced between 100 Hz and 10 kHz. From the output of each narrow band channel, an envelope was extracted (i.e., a Hilbert transform was computed using fast Fourier transform (FFT)). The envelope in each narrow band channel was filtered using second-order Butter-

Table 1. Classification of consonants by phonetic features

Features	/k/	/g/	/m/	/tʃ/	/l/,	/s/,	/ʃ/,	/l./	/b/,	/d/,	/ḍ/	/dʒ/	/t/,	/e/	/v/	/j/	/ŋ	/p/	r	n	h
Voicing	-	+	+	-	+	-	-	+	+	+	+	+	-	-	+	+	+	-	+	+	-
Place	vel	Vel	Bil	paa	alv	alv	paa	lal	Bil	alv	den	pla	alv	den	lab	ron	ret	bil	alv	alv	glo
Manner	plo	Plo	Nas	aff	lat	fri	fri	liq	Plo	plo	plo	aff	Plo	plo	fri	gli	nas	plo	lat	nas	fri

bil: bilabial; lad: labiodental; alv: alveolar; paa: palatoalveolar; den: dental; ret: retretroflex; pal: palatal; glo: glotal; lal: linguaalveolar; nas: nasal; plo: plosive; fri: fricative; lat: lateral; aff: affricative; liq: liquid; gli: glide

Voice: "+" Voiced; "-" Unvoiced

Table 2. Consonant identification scores obtained in quiet and at different SNRs from the study participants

SNR	Quiet		+10 dB SNR		+5 dB SNR	
Condition Group	UP Mean (SD)	DBM Mean (SD)	UP Mean (SD)	DBM Mean (SD)	UP Mean (SD)	DBM Mean (SD)
YNH	61.50 (1.87)	60.15 (4.26)	59.40 (2.25)	58.35 (2.12)	60.15 (3.12)	56.40 (2.14)
YHI	33.75 (3.88)	36.90 (3.69)	31.25 (2.49)	34.80 (2.62)	27.45 (5.91)	29.40 (4.01)
ONH	40.95 (7.17)	43.65 (7.49)	41.25 (5.92)	43.50 (6.35)	33.75 (3.38)	42 (8.39)
OHI	26.65 (5.26)	29.85 (8.35)	25.95 (2.24)	25.20 (3.79)	20.70 (3.22)	21.06 (7.86)

UP: unprocessed; DBM: deep band modulated; YNH: younger normal hearing; YHI: younger hearing impaired; ONH: older adult normal hearing; OHI: older adult hearing impaired; SNR: signal-to-noise ratios

worth filters with cutoff frequencies between 3 and 30 Hz [15]. Furthermore, the envelope was rectified in each channel. Before summating the envelope from each channel, a gain of 15 dB [15] was provided for the channels within the frequency range of 1–4 kHz to obtain the DBM VCV syllables. Each DBM VCV syllable was digitally mixed with speech babble noise [16] at SNRs of +10 dB and +5 dB, respectively, using the SNR MATLAB code.

Procedure

Each participant was seated comfortably in a soundproof room, and a closed set of target test VCV stimuli was displayed to the participant on a monitor screen. The unprocessed VCV stimuli were played through the GUI of MATLAB software (version 2009B) loaded on a personal laptop. The output of the laptop was connected to the auxiliary input of the audiometer. The output of the dual channel diagnostic audiometer (Inventis Piano, Padova, Italy) was delivered through TDH 39 headphones at the participant's most comfortable level. Before the presentation of the stimuli, the level of the presentation was monitored with the calibration tone. During the presentation of the stimuli, the average deflection on the VU meter was measured at 0 dB. Each VCV speech syllable was presented three times. The participant was instructed to click the heard syllable on the monitor screen. A successive stimulus was presented 2 s after the previous response. Similar procedures were carried out at +10 dB and +5 dB SNR conditions. The order of the stimuli was randomized across the participants. After a sufficient pause requested by each participant, the same procedure was conducted in DBM conditions at different SNRs. This ensured that the order of testing in different conditions and SNRs was counterbalanced across the study participants.

Data Analyses

In the consonant identification task, a score of one mark was assigned for each correct response. For an incorrect response, a score of zero was assigned. The maximum total of obtained marks was 63 (i.e., 21 consonants * three repetitions) under each of the experimental conditions. In sequential transfer of information analysis (SINFA; Linguistics and Phonetics, London, United Kingdom), consonants are classified based on three phonetic features (i.e., voicing, place of articulation, and manner of articulation). The phonetic features of these 21 consonants, classified under voicing, place, and manner, is tabulated in Table 1. Furthermore, the information transmitted for each feature of articulation was established. Three trials of identification data from all consonants were united across the participants in each group. The result is a single confusion matrix, which was created for six different conditions. The confusion matrices of all six conditions were then analyzed using SINFA to investigate the percentages of the features of information transmitted. The sequential information analysis procedure utilized in the present study was adopted from Wang and Bilger [17]. This analysis was performed using the Xfer feature information software package (developed by the Department of Linguistics and Phonetics, London, United Kingdom). In the first iteration, each phonetic feature was calculated. In subsequent iterations, the features with the highest percentages of information transmitted in the previous iteration were held constant and partialled out.

Statistical Analyses

The study results were statistically analyzed using the Statistical Package for the Social Sciences for Windows (SPSS Version 20.0; IBM Corp, Armonk, New York). The data of consonant identification scores (CIS) collected from study participants in different experimental conditions were subjected to repeated measures analysis of variance (ANOVA). A p value of <0.05 was considered statistically significant. The SINFA results were descriptively analyzed to determine the feature information transferred in each experimental condition.

RESULTS

Consonant identification scores and SINFA data were collected in unprocessed and DBM conditions in quiet and at +5 dB SNR and +10 dB SNR from each group of participants.

Consonant Identification Scores

The mean (M) and standard deviation (SD) of the CIS in unprocessed and DBM conditions at different SNRs from each group are tabulated in Table 2. Two-way repeated measures ANOVA was performed on the CIS, comparing conditions (unprocessed and DBM) and SNRs (quiet, +10 dB, and +5 dB) between subject factors as groups (YNH, ONH, YHI, and OHI). The results revealed a significant main effect for the condition (F (1, 36)=12.81, p=0001), such that CIS was better in DBM conditions than in unprocessed conditions. A significant main

Table 3. F ratio, p value, and df (3, 36) of MANOVA obtained from groups under each experimental condition

Condition	UP		DBM	
SNR	F ratio	p	F ratio	p
Quiet	96.35	0.000***	43.71	0.000***
+10 dB SNR	168.13	0.000***	53.70	0.000***
+5 dB SNR	108.42	0.000***	61.72	0.000***

UP: unprocessed; DBM: deep band modulated; SNR: signal-to-noise ratios. ***p>0.001

effect was also observed for the SNRs (F (2, 72)=36. 62, p=0000), in which the CIS decreased with decreasing SNR. However, there was no significant interaction effect for the condition (F (2, 72)=36.60, p=0.140). In addition, a significant difference was noted in subject factors between groups (F (1, 36)=190.45, p=0.000), suggesting that CIS was better in the YNH group than in the ONH group, followed by the YHI group and the OHI group.

Although there was no significant difference in the interaction effects, a significant difference was observed in the main effects of the conditions, SNRs, and groups. Thus, in each group, three post hoc paired sample t-tests were conducted to determine the SNR at which the CIS were significantly different between conditions. The results revealed that a significantly better CIS was observed in DBM conditions than in unprocessed conditions at +5 dB SNR in the ONH group (t (9)=−3.45, p=0.007), the YHI group (t (9)=−2.93, p=0.000), and the OHI group (t (9)=−3.18, p=0.011).

Additionally, to determine which experimental conditions caused significant differences between the groups, multivariate analysis of variance (MANOVA) was performed. The MANOVA results are displayed in Table 3. The results revealed that irrespective of conditions and SNR, the mean CIS were significantly different between groups. Furthermore, to determine the groups between which the differences were significant, a Duncan post hoc test was performed on each experimental condition. The results of the Duncan test revealed that all groups were significantly different from each other; for example, better CIS was noted in the YNH group than in the ONH group, followed by the YHI group and the OHI group. This was true in each condition.

Sequential Transfer of Information Analysis

Twenty-one stimuli were used; each stimulus was diverse in its phonetic features. Each of the 21 stimuli was presented three times, so a total of 63 stimuli were presented. This was true for each experimental condition. Sequential Transfer Of Information Analysis (SINFA) was applied for both unprocessed and DBM conditions in quiet and noise (+10 and +5 dB SNR); the data were obtained from each group of participants to assess the amount of information transferred from stimulus to response for a set of the most relevant phonetic features. The information transmitted from each phonetic feature in different conditions from all the groups is shown in Figure 2. The maximum information in bits that can be transmitted for the 21 stimuli, which were presented three times, is 4.39.

Except for the YHI group, participants from all groups conveyed voicing information better than other phonetic features in different experimental conditions. Specifically, in the YNH group, voicing information was conveyed relatively better than place of articulation,

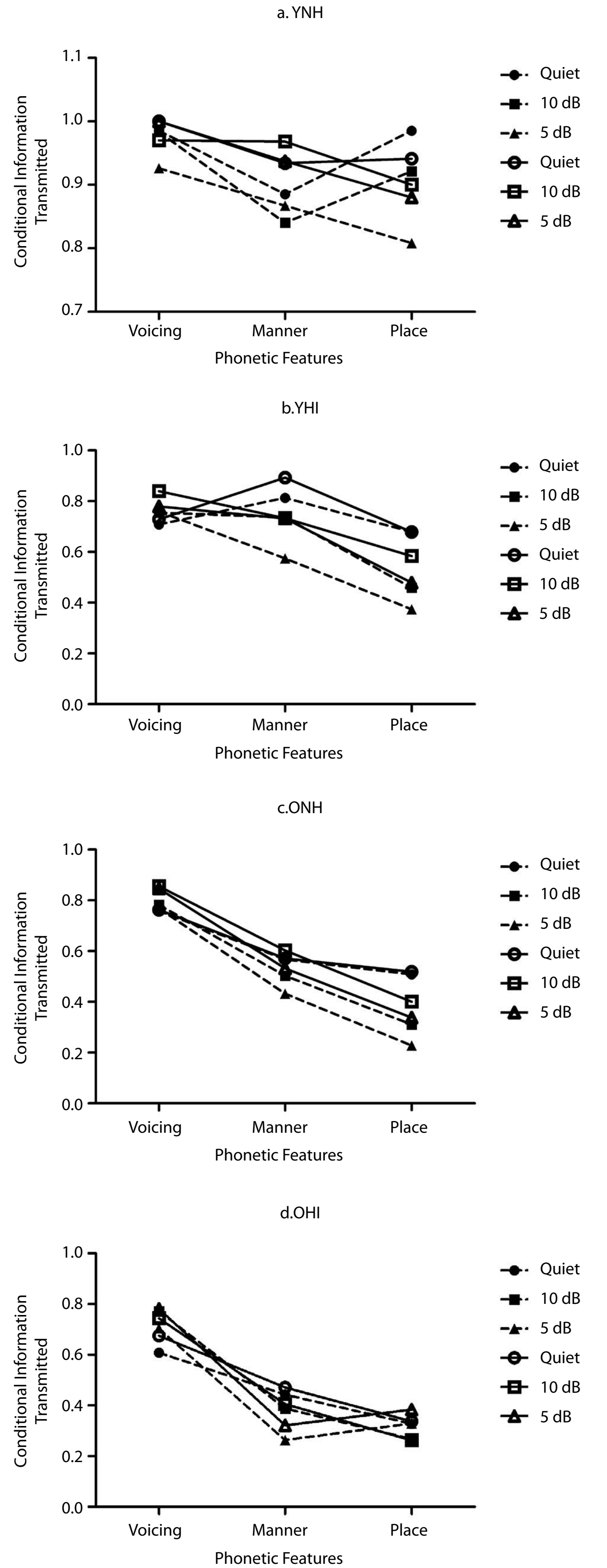

Figure 2. Information transmitted for each phonetic feature under different experimental conditions from each group (a. YNH; b. YHI; c. ONH; d. OHI). Filled shapes with dotted lines represent unprocessed conditions. Unfilled shapes with solid lines correspond to DBM conditions.

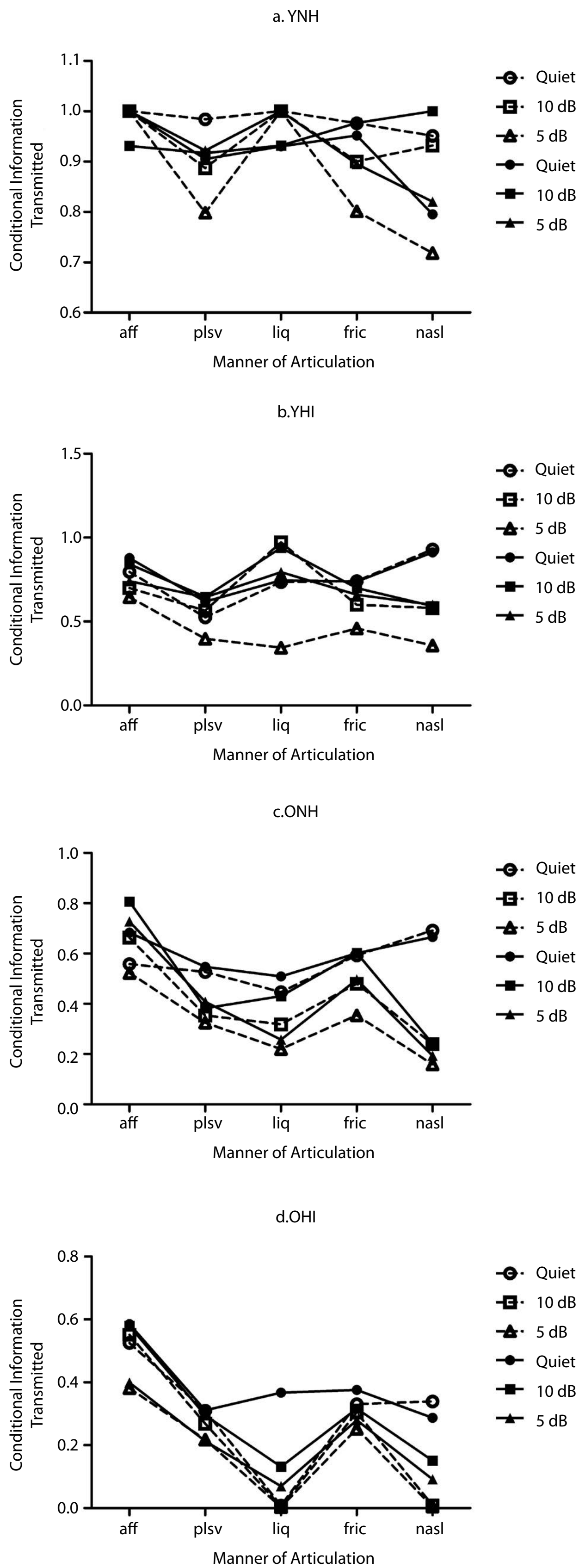

Figure 3. Information transmitted for each manner of articulation under different experimental conditions for each group (a. YNH; b. YHI; c. ONH; d. OHI). Unfilled shapes with dotted lines symbolize unprocessed conditions. Filled shapes with solid lines represent DBM conditions.

followed by manner of articulation, in quiet and at +10 dB SNR in unprocessed conditions. However, at +5 dB SNR, the transmitted information was better for the voicing feature than for the manner of articulation, followed by the place of articulation. In DBM conditions, irrespective of SNR, the voicing information was carried best, and the manner of articulation was conveyed relatively better than the place of articulation. For the YHI group, except at +5 dB SNR in unprocessed conditions, the information conveyed for the manner of articulation was relatively better than that conveyed for voicing, followed by place of articulation, under each experimental condition. In unprocessed conditions, at +5 dB SNR, the voicing feature transmitted best, and the manner of articulation was conveyed relatively better than the place of articulation. For the ONH group, voicing information was conveyed best, and the manner of articulation was conveyed better than the place of articulation under different experimental conditions. For the OHI group, except at +5 dB SNR under each condition, voicing information was transferred best, and the manner of articulation was transferred relatively better than the place of articulation. At + 5 dB SNR under each condition, voicing information was conveyed best, and the information of place of articulation was transferred better than the manner of articulation. Furthermore, it was observed that each of the phonetic features was transmitted better in DBM conditions than in unprocessed conditions at different SNRs. This was true for each group. It was also observed that each feature was conveyed better in the YNH group than in the ONH group, followed by the YHI group and the OHI group. This was true for each experimental condition.

Manner of Articulation

Figure 3 shows the conditional information transmitted for the experimental conditions as a function of manner of articulation. It is noted that for the YNH and YHI groups, there was no trend in the transmission of information in different manners of articulation for each condition in quiet and at different SNRs. In the ONH group, for both unprocessed and DBM conditions, affricatives and fricatives conveyed greater amounts of information than plosives, liquids, and nasals. In the OHI group, it was observed that in unprocessed conditions, the information transmitted was less than 0.010 for liquids in quiet and for liquids and nasals at different SNRs. Furthermore, in unprocessed conditions, in quiet and at two SNRs, more information was contributed by affricatives and fricatives than by plosives. It was also observed that in DBM conditions at +10 dB SNR and +5 dB SNR, affricatives, fricatives, and plosives were conveyed better than liquids and nasals. However, in quiet conditions, affricative information transmitted better than other manners of articulation. In addition, each manner of articulation was processed better in DBM conditions than in unprocessed conditions in quiet and at the two different SNRs. Furthermore, the information conveyed for each manner of articulation was more readily accessed by the YNH group than by the ONH group, followed by the YHI group and the OHI group.

Place of Articulation

Figure 4 shows the conditional information transmitted for the experimental conditions as a function of place of articulation. It was noted that in each experimental condition, except in the YNH group, the information transmitted threshold was less than 0.010 for the bilabial and alveolar places of articulation. Additionally, palate-alveolar conveyed greater amounts of information than velar.

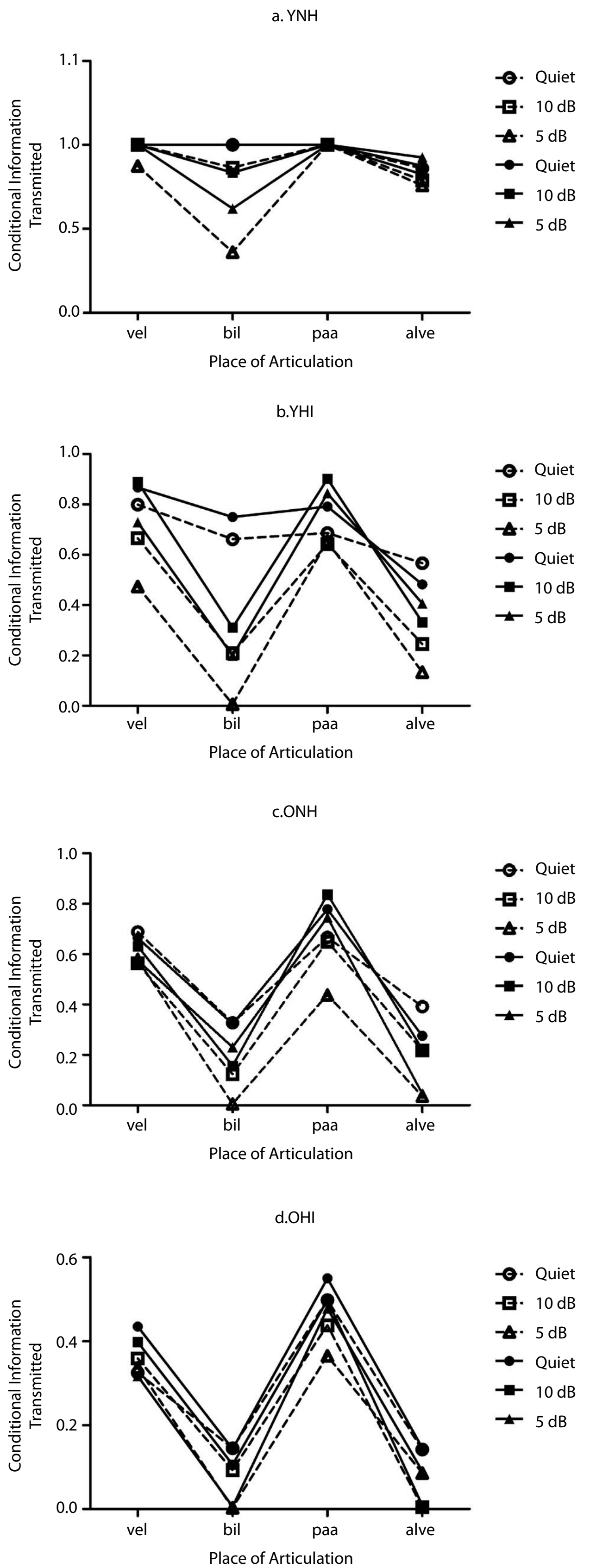

Figure 4. Information transmitted for each place of articulation under different experimental conditions for each group (a. YNH; b. YHI; c. ONH; d. OHI). Unfilled shapes with dotted lines symbolize unprocessed conditions. Filled shapes with solid lines represent DBM conditions.

Furthermore, more place information was conveyed in DBM conditions than in unprocessed conditions under each experimental condition. It was also observed that the information conveyed for each place of articulation was more readily accessed by the YNH group than by the ONH group, followed by the YHI group and the OHI group.

DISCUSSION

CIS and Sequential Information Transmitted Under Each Condition

As expected, in each group, CIS decreased with decreasing SNR in both unprocessed and DBM conditions. The consonant identification scores at different SNRs were better in DBM conditions than in unprocessed conditions for all groups of participants. In addition, at +5 dB SNR, CIS was significantly better in DBM conditions than in unprocessed conditions in the YHI, ONH, and OHI groups. In DBM conditions, although the noise of +5 dB SNR modifies the spectral parameters, the higher amplitudes of modulation depth in each VCV stimulus are less energetically masked; this provides cues for the repetition of heard stimulus. Although the YHI, ONH, and OHI groups all have temporal resolution impairment, the participants were able to detect subtle changes in the amplitudes of modulation depth embedded in noise. It appears that DBM lessens temporal impairment by increasing the modulation depth.

In the YHI group, due to damage in the cochlea, the stiffness of the basilar membrane is decreased, which leads to wider auditory filters. Thus, frequency resolution is impaired in this group. In addition, noise alters the spectral content of speech. The option to utilize the temporal cue to repeat the heard stimulus was made available to YHI participants. This is clearly reflected in the study results, which showed improved perception of DBM in +5 dB SNR. From the sequential transfer function, it was noted that in both conditions, the voicing and manner of articulation transmitted better than the place of articulation. However, more information was transmitted in DBM conditions than in unprocessed conditions. Specifically, in unprocessed conditions, all manners of articulation were equally transmitted. However, in DBM, other than plosives, all manners of articulation were conveyed relatively better. This is probably because consonantal information is carried by a slow envelope (<8 Hz) of the speech signal [18]. This provides evidence that a slow envelope of speech is processed better than a faster envelope of speech. Regarding place of articulation, for both conditions, velar and palato-alveolar information were conveyed better than bilabial and alveolar information. In UP conditions, most participants perceived bilabials as alveolar and palate-alveolar. It was observed that in DBM conditions, perception of the bilabial and alveolar places of articulation was improved. It is likely that enhancing the speed of the modulations of the speech signals helped participants to process salient cues, such as formant transition and burst amplitude. It is reasonable that enhancing the consonant portion of speech and compressing the vocalic portion may have improved the listeners' ability to process the slow modulation envelope of each VCV (3–30 Hz) stimulus. Enhancement of the modulation depth by 15 dB was found to be beneficial and increased the resistance of the cues to the obscuring of temporal dips [7] by +5 dB SNR. This indicates that a 15 dB modulation gain in each syllable enables processing of the available amplitudes of temporal dips at a +5 dB SNR.

Temporal impairment is generally present in the ONH group. These participants rely more on spectral cues than temporal cues for the perception of speech. However, in noise, the spectral content of speech is obliterated. Thus, in the present study, an experiment was conducted in which spectral cues were removed by adding noise to the speech cues. In addition, the temporal envelope was enhanced in one set of stimuli (DBM), and the temporal envelope remained the same in the other set of stimuli (unprocessed). At +5 dB SNR, although the noise disrupted the spectral content of the speech, the older adult participants with normal hearing could use temporal cues to extract available modulation depth in the noise, which resulted in better scores in DBM conditions. This was supported by a study by Lorenzi et al. [19]. The improved perception in DBM conditions can be explained by the increased distinction of consonant and vowel segments in the stimuli and the enhancement of the modulation depth by 15 dB; this improved the listeners' ability to process the slow modulation envelope in each stimulus. This was found to be beneficial and made the cues more resistant to the obscuring of temporal dips. The results of the present study are in agreement with a research report by Hemanth and Akshay [12], who found better perception in DBM conditions than in unprocessed conditions. This infers that the reduction in the modulation depth of the speech envelope caused by noise was compensated by expanding the low-frequency temporal modulation by 3–30 Hz. In addition, at +5 dB SNR, it was noted that voicing and manner of articulation features are conveyed better than voicing for both conditions. The amount of information transmitted is relatively higher in DBM conditions than in unprocessed conditions. Specifically, in unprocessed conditions, all the manners of articulation were transmitted to the same extent. However, affricatives and fricatives were transmitted relatively better than the other articulations. This indicates that after enhancement, listeners can capture cues that have slow envelope modulation across a stimulus. Regarding the place of articulation, for both conditions, bilabial and alveolar features were less transferred than velar and palate-alveolar features. In unprocessed conditions, bilabial syllables are substituted by alveolar syllables. In addition, alveolar sounds are replaced by palato-alveolar syllables. However, in enhanced conditions, the errors are still present but are relatively fewer than in unprocessed conditions. It is speculated that the faster modulations of the speech signal enabled participants to processes salient cues, such as formant transition and burst amplitude.

The study participants in the OHI group have cochlear pathology. It is known from the literature that older adults with cochlear hearing loss have broadened auditory filters [20] and temporal asynchrony [6]. Due to these wider auditory filters, spectral peaks and valleys in the stimuli are smoothed out in both unprocessed and DBM conditions. In addition, noise degrades the spectral component of speech [2] and also masks the lower amplitude of the temporal envelope [21]. Therefore, it is likely that listeners will rely on higher amplitudes of temporal cues in the presence of noise. In addition, few auditory filters are available for analysis, especially at lower frequency bands; however, noise accompanied with VCV syllables taxes the available filters so that noise accumulates in functioning filters, leading to reduced recognition at lower SNRs. This occurs for both unprocessed and DBM syllables. However, the adverse effect in unprocessed conditions is greater than that in DBM conditions. Thus, listeners in this group required higher SNR levels to achieve good performance. From the sequential transfer function, it was noted that voicing and place of articulation were conveyed better than place of articulation. In manner of articulation, liquids and nasals are conveyed less than other manners of articulation for both conditions at +5 dB SNR. Liquids are differentiated from nasals and other consonants by formant onset frequency and transitions [22]. Because differential sensitivity for low-frequency signals is poor in individuals in the OHI group, they probably perceived nasals (/m/) instead of liquids and vice versa. In the alveolar and bilabial places of articulation, less information was transmitted compared to other places of articulation for both conditions. Because the members of this group have frequency resolution impairment, they were unable to identify rapidly changing formant bursts and transitions. Even after enhancement, the probability of error in repeating the heard bilabial and alveolar sounds was reduced but did not improve significantly. Thus, the results of the present study for the OHI participants clearly revealed improvement in VCV syllables processed by DBM conditions, in which enhancement of the modulation depth even at lesser SNRs enabled the participants to use available temporal cues. This suggests that the listeners can perceive syllables in temporal dips in the masker. Another reason would be that the available modulation depth in the region where the envelope modulation is undermasked could have enabled older adults to easily understand the speech syllables i.e., release from masking [23].

Consonant Identification Scores and Sequential Information Transmitted Between Groups

Overall, the results revealed that CIS was better in the younger adult group compared to the older adult group, followed by the younger and older adult groups with hearing loss, in both unprocessed and DBM conditions at each SNR.

a. Does Aging Affect Consonant Recognition?

In unprocessed conditions at each SNR, CIS in the YHI group was found to be significantly better than in the OHI group. In addition, irrespective of SNR, all the features were well conveyed in the YNH group. However, in the ONH group, voicing information was transmitted better than the manner and place of articulation. Perception of voicing mainly depends on the voice onset time and the F1 (first formant) transition [24]. In the present study, individuals with ONH perceived voiced sound as voiced. The lesser transmission of the manner and place of articulation suggests temporal asynchrony in older adults, who might have failed to extract the available temporal dips in noise. In fact, these temporal dips across VCV syllables are more relevant cues in conditions where noise degrades temporal fine structure cues. Furthermore, it is speculated that due to temporal asynchrony, the neurons of older adults have difficulty processing important cues against background noise. Due to neuronal time lag in processing the present cues, the subsequent cues arise, which leads to incorrect access and results in misperception.

Thus, older adults find it difficult to follow speech, especially in noise conditions.

Interestingly, when the amplitude of the envelope was increased, the phrase perception at each SNR appeared to improve in older adults. As the results of the present study demonstrate, phrase perception in the YAG group was not significantly different from that of the OAG group in DMB conditions. Older adults could accesses the VCV syllables through voicing and manner of articulation. However, relatively less information was transmitted from the place of articulation. This clearly indicates that envelope enhancement lessens the asynchronous system and makes the inherent cues in the VCV syllables available although the noise partly obscured the temporal dips and spectral content of the sound. Furthermore, the auditory masking effect lessened interruption when processing the target VCV syllables when the envelope was increased by 15 dB. Thus, the negative impact of aging on perception seen in unprocessed conditions is partly reduced in conditions where the speech envelope is enhanced.

b. Does Age-Related Hearing Loss Have an Effect on Consonant Recognition?

The CIS in older adults were found to be significantly better than those of older adults with hearing loss. This infers that the combined effects of aging and hearing loss in OHI have a more negative impact on perception than hearing loss alone. Similar results were obtained even in DBM conditions at lesser SNRs. However, the perception of VCV syllables in DBM conditions was found to be significantly different between YHI and OHI participants at higher SNRs. From the sequential information, it was revealed that the voicing feature was transmitted best for both conditions in the YHI and OHI groups. Furthermore, it is clear that the place of articulation was less conveyed for both the YHI and OHI groups; this is due to their wider auditory filters. However, the manner of articulation was transmitted relatively better in DBM conditions than in unprocessed conditions. This result appeared to reduce the impact of the combined effects of aging and hearing loss in conditions where the envelope was extracted within a bandwidth of 3–30 Hz in each channel from the broad frequency band of the phrase. The envelope from this narrow bandwidth channel with a gain of 15 dB might be well above the noise level, allowing the available functioning filters to extract import cues for perception (listening in dips).

To conclude, at reduced SNR, the cues from DBM enable listeners to repeat heard VCV syllables. The effect of aging and the combined effects of aging and hearing loss are partly lessened by DBM through enhancing the manner of articulation in VCV syllables. In noise conditions, DBM enabled patients to listen through temporal dips where the amplitude of the envelope was enhanced. The use of a DBM strategy in individuals with sensorineural hearing loss has the potential to improve speech perception in adverse listening conditions. Hence, this strategy has scope to be utilized as a rehabilitation technique.

Ethics Committee Approval: Ethics committee approval was received for this study from the ethics committee of All India Institute of Speech and Hearing.

Informed Consent: Written informed consent was obtained from patients who participated in this study.

Peer-review: Externally peer-reviewed.

Author Contributions: Concept - H.N.S.; Design - H.N.S.; Supervision - V.K.; Resources -H.N.S., V.K.; Materials - H.N.S., V.K.; Data Collection and/or Processing - V.K.; Analysis and/or Interpretation - H.N.S.; Literature Search - V.K.; Writing Manuscript - H.N.S., V.K.; Critical Review - H.N.S., V.K.

Acknowledgements: The authors acknowledge Dr. S.R. Savithri, Director, All India Institute of Speech and Hearing, Mysore for permitting us to do the study. The authors also thank all the participants for their co-operation.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study has received no financial support.

REFERENCES

1. Jenstad LM, Souza PE. Temporal envelope changes of compression and speech rate: combined effects on recognition for older adults. J Speech Lang Hear Res 2007; 50: 1123-38. [CrossRef]
2. Cohen I. Speech enhancement using super-Gaussian speech models and noncausal a priori SNR estimation. Speech Communication 2005; 47: 336-50. [CrossRef]
3. Steeneken HJ, Houtgast T. A physical method for measuring speech-transmission quality. J Acoust Soc Am 1980; 67: 318-26. [CrossRef]
4. Moore BCJ. Frequency analysis and masking, in Hearing. Orlando, Florida Academic Press; 1995. [CrossRef]
5. Dubno JR, Dirks DD, Morgan DE. Effects of age and mild hearing loss on speech recognition in noise. J Acoust Soc Am 1984; 76: 87-96. [CrossRef]
6. Plomp R. Auditory handicap of hearing impairment and the limited benefit of hearing aids. J Acoust Soc Am 1978; 63: 533-49. [CrossRef]
7. Glasberg BR, Moore BC. Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. J Acoust Soc Am 1986; 79: 1020-33. [CrossRef]
8. Tyler RS, Wood EJ, Fernandes M. Frequency resolution and hearing loss. Br J Audiol 1982; 16: 45-63. [CrossRef]
9. Drullman R. Temporal envelope and fine structure cues for speech intelligibility. J Acoust Soc Am 1995; 97: 585-92. [CrossRef]
10. Johnson KL, Nicol TG, Zecker SG, Kraus N. Auditory brainstem correlates of perceptual timing deficits. J Cogn Neurosci 2007; 19: 376-85. [CrossRef]
11. Nagarajan SS, Wang X, Merzenich MM, Schreiner CE, Johnston P, Jenkins WM, et al. Speech modifications algorithms used for training language learning-impaired children. IEEE Trans Rehabil Eng 1998; 6: 257-68. [CrossRef]
12. Hemanth NS, Akshay M. Deep band modulation and noise effects: perception of phrases in adult. Hearing, Balance and Communication 2015; 13: 1-7.
13. Ramkrishna BS, Nair KK, Chiplunkar VN, Atal BS, Ramachandran V, Subramanian R. Relative Efficiencies of Indian Languages. Nature 1961; 189: 614-7. [CrossRef]
14. Winholtz WS, Titze IR. Conversion of a head-mounted microphone signal into calibrated SPL units. J Voice 1997; 11: 417-21. [CrossRef]
15. Narne VK, Vanaja CS. Effect of envelope enhancement on speech perception in individuals with auditory neuropathy. Ear Hear 2008; 29: 45-53.
16. Glasberg BR, Moore BC. Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech. Scand Audiol Suppl 1989; 32: 1-25.
17. Wang MD, Bilger RC. Consonant confusions in noise: a study of perceptual features. J Acoust Soc Am 1973; 54: 1248-66. [CrossRef]
18. Drullman R, Festen JM, Plomp R. Effect of temporal envelope smearing on speech reception. J Acoust Soc Am 1994; 95: 1053-64. [CrossRef]
19. Lorenzi C, Berthommier F, Apoux F, Bacri N. Effects of envelope expansion on speech recognition. Hear Res 1999; 136: 131-8. [CrossRef]
20. Van Tasell DJ. Hearing loss, speech, and hearing aids. J Speech Hear Res 1993; 36: 228-44. [CrossRef]
21. Fullgrabe C, Meyer B, Lorenzi C. Effect of cochlear damage on the detection of complex temporal envelopes. Hear Res 2003; 178: 35-43. [CrossRef]
22. Johnson K. Acoustic and auditory phonetics. 2 ed. Oxford: Blackwell: Wiley; 2003.
23. Fullgrabe C, Berthommier F, Lorenzi C. Masking release for consonant features in temporally fluctuating background noise. Hear Res 2006; 211: 74-84. [CrossRef]
24. Summerfield Q, Haggard M. On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. J Acoust Soc Am 1977; 62: 435-48.[CrossRef]