

Article

A Computational Model for Evaluating Transient Auditory Storage of Acoustic Features in Normal Listeners

Nannan Zong and Meihong Wu * 

School of Informatics, Xiamen University, Xiamen 361005, China; 31520190154747@stu.xmu.edu.cn

* Correspondence: wmh@xmu.edu.cn

Abstract: Humans are able to detect an instantaneous change in correlation, demonstrating an ability to temporally process extremely rapid changes in interaural configurations. This temporal dynamic is correlated with human listeners' ability to store acoustic features in a transient auditory manner. The present study investigated whether the ability of transient auditory storage of acoustic features was affected by the interaural delay, which was assessed by measuring the sensitivity for detecting the instantaneous change in correlation for both wideband and narrowband correlated noise with various interaural delays. Furthermore, whether an instantaneous change in correlation between correlated interaural narrowband or wideband noise was detectable when introducing the longest interaural delay was investigated. Then, an auditory computational description model was applied to explore the relationship between wideband and narrowband simulation noise with various center frequencies in the auditory processes of lower-level transient memory of acoustic features. The computing results indicate that low-frequency information dominated perception and was more distinguishable in length than the high-frequency components, and the longest interaural delay for narrowband noise signals was highly correlated with that for wideband noise signals in the dynamic process of auditory perception.

Keywords: sensitivity; acoustic features; instantaneous change in correlation; interaural delay; auditory perception; computational model; interaural coherence



Citation: Zong, N.; Wu, M. A Computational Model for Evaluating Transient Auditory Storage of Acoustic Features in Normal Listeners. *Sensors* **2022**, *22*, 5033. <https://doi.org/10.3390/s22135033>

Academic Editor: Iren E. Kuznetsova

Received: 24 May 2022

Accepted: 29 June 2022

Published: 4 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The concept of auditory perception seems so simple. People can easily hear the music of the neighborhood piano, detect the sound of nearby passing cars, and perceive the sound of waves in the distance [1]. On the surface, auditory experience is simple and effective, but there are still many difficulties and questions that have not been adequately addressed [2]. In particular, the information that reaches our hearing is usually incomplete and fuzzy, distributed in space and time, and is not sorted neatly according to its source [3]. Obviously, in order to support effective interaction with the world, human listeners need to extract a lot of information about the world from sound [4].

Notably, the auditory system of humans confronts the natural pressure to manage sound wave reflections in ordinary living conditions on a daily basis. In a noisy, reverberant environment, listeners receive not only sound waves directly originating from sound sources but also reflections of the sound sources [5]. That is, there are not only direct sound waves from the sound source but also reflected sound waves from the surface of surrounding objects, and the reflected sound always lags behind the direct sound in reaching the human ear [6]. In order to segregate the target signal perceptually from other interference stimuli in the reverberation environment, the auditory system must simultaneously distinguish the sound wave directly from the signal source with the reflection of the signal source in addition to the sound wave from the interference source with the reflection of the interference source [7]. Otherwise, the auditory scene will become chaotic. Thus, under

adverse hearing conditions, the high perception fusion tendency of acoustic signals is conducive to speech recognition.

Since auditory information is processed in a time sequential mode, the auditory transient “memory” and sequential auditory information readout from the auditory transient storage are very important for organizing sound stimuli components into auditory image units [5]. Therefore, the auditory system must make use of the acoustic features of direct sound retained in the initial auditory “memory” to carry out correlation “calculation” with the acoustic features of the reflected sound, and then conduct perceptual integration of the direct and reflected sound information to form the perception of a single auditory event [7]. Thus, the transient “memory” of the acoustic features is crucial for accurate localization of sound incidence in an everyday environment [5,7].

It has been well confirmed that due to the ability to transiently store raw acoustic features of the leading wave, the auditory system of human listeners can calculate the similarity between the leading and lagging sound waves and represent the calculation result at the perceptual level [8,9]. Previous studies have discovered that the ability to transiently store acoustic features contributes to the perceptual integration of the direct sound wave and its time-delayed reflections [10–12]. It has also been found that the acoustic features of waves coming directly from the sound source must be transiently held for a short period to achieve integration of the direct acoustic wave with its time-delayed reflection [13,14]. The remarkable ability to transiently store acoustic features enables the auditory system to recognize whether acoustic features of one wave are highly correlated with that of another wave. This is beneficial for speech recognition in noisy environments. There is evidence showing that speech performance under a noisy environment was strongly correlated with the ability of temporal storage capacity of acoustic features [10,11]. The ability to transiently store acoustic features is critical for perceptual segregation of target speech and an informational speech masker in a simulated noisy and reverberant environment [11]. This interaural correlation processing ability plays an important role in improving speech recognition by inducing a perceived spatial separation between target source and the other masking source. Thus, detecting a change in correlation is an essential component of auditory scene analysis. For correlation comparison, transient auditory storage of acoustic features is necessary [14]. Therefore, the ability to temporally store acoustical features is critical for later speech perception in noisy environments [11].

There has been increasing interest in discussing how to assess this remarkable ability to transiently store acoustic features and elucidating methods to provide a logical description of the calculation process of interaural correlated acoustical information in auditory scene analysis. Therefore, this study aims to explore methods to evaluate the ability of transient “memory” of acoustic features qualitatively.

There are many ways to represent auditory perception information, such as classical psychophysical methods for estimation (e.g., estimation of difference threshold [15–17] or only significant difference [8]), non-parametric Bayesian clustering algorithms for representation of auditory cues [18], and deep learning methods [19,20] for single and multiple sound source localization. These studies indicate that the minimal difference between two stimuli leads to a mathematical approach to relating to the external physical world based on experimental data and the internal psychic response. Delving further, researchers have found that the processing of interaural correlation could be investigated by measuring the sensitivity to an instantaneous change in interaural correlated steady-state noise [8]. This method, in which an instantaneous change in interaural correlation is created by placing an uncorrelated noise burst between two noise stimuli with identical waveforms, has been widely used for binaural gap detection. In other words, incoherence is introduced in this kind of method; for example, when an unrelated noise is added to the acoustic signal appearing in one of the ears. It can be seen that this instantaneous change sensitivity helps to assess the influence of the listener on auditory perception.

It has been well documented that humans can detect the binaural gap in interaural correlated steady-state noise when the interaural correlated noise arrives simultaneously [8,12,21].

Interestingly, when an interaural time difference [22] (i.e., interaural delay) was introduced, the binaural gap in interaural correlated steady-state noise was still detectable [23,24]. Since the acoustic signals in the left and right ears differ in some way, such as a simple time delay, the interaural incoherence occurs when measuring the similarity between the binaural sound inputs. The preservation of the sensitivity to the binaural gap even when an interaural delay was introduced indicated that acoustic temporal fine structure features of noise were maintained for the duration of the interaural interval [11,25,26] and allowed similarity computation between the binaural sound inputs. Thus, measuring the impact of interaural delay when the binaural gap is detected [27–29] can provide a way of investigating the transient memory of acoustic features.

Since a binaural gap is detected when the noise sections flanking the binaural “gap” are sufficiently large [27,28], the shortest detectable binaural gap could be one measure of the temporal response of the binaural system to instantaneous changes in interaural correlation. In the meantime, the transient memory of acoustic features of the noise signal entering the leading ear continues to attenuate and the interaural correlation of flanking the break-in correlation in the central representation gradually decreases. If the interaural integration of acoustic features (i.e., transient memory of acoustic features) is degraded by introducing an interaural delay, then the detectability of the binaural gap should be reduced, resulting in decreased sensitivity for binaural gap detection. Thus, the dynamic characteristic of transient memory of acoustic features can be reflected by the variability in the duration threshold for the binaural gap, detectable because the interaural delay is varied. However, to the best of our knowledge, whether the sensitivity to the binaural gap affects the interaural delay when the binaural gap is detected has not been well discussed in the previous literature.

In the present study, whether the ability to transiently store acoustic features is affected by the interaural delay was assessed by measuring the sensitivity for detecting the binaural gap at various interaural delays. Moreover, the present study also contributes to a discussion about whether the sensitivity to the binaural gap is related to transient memory of acoustic features. It is well known that the auditory system is extraordinarily capable of processing acoustic features [21]. There is evidence in a previous study that the jitter of interaural time differences in uncorrelated signals was determined by the fine structure acoustic feature and envelope modulations inherent to the stimulus and the bandwidth of the stimulus, and the absolute bandwidth of low-frequency noise bands is lower than that of high-frequency noise bands when the relative bandwidth is fixed [11]. However, to our knowledge, whether the change in absolute bandwidth affects the ability to transiently store acoustic features has not been well documented. In addition, whether this binaural ability is maintained when there is a break in a correlation between the interaural correlated narrowband and wideband noise, or whether the ability to temporally store high-frequency acoustic features is associated with the ability to transiently store low-frequency acoustic features, is still unclear.

Research has shown that narrowband noise might have a much higher perceptual fusion tendency compared to wideband noise [30]. Although wideband noise is closer to natural noise, regular narrowband noise more easily attracts attention. Moreover, compared with narrowband noise, wideband noise has a greater impact on the hearing and on the intelligibility of speech. Especially under a low signal-to-noise ratio, wideband noise will cause serious interference and damage to speech perception in noise. Thus, it is also important to investigate the influence of the components (i.e., which attribute) of wideband noise or narrowband noise on the ability of auditory transient memory of acoustic features of a random noise in the sound source. Whether maintaining information of auditory stimuli is dependent on interaural time delay or the frequency component also needs clarification.

Therefore, this study examined the detectability of a binaural gap between interaural correlated narrowband or wideband noise when the longest interaural delay was introduced and measured the ability to temporally store acoustic details associated with the

relationship between the transient storage of fine acoustic features of interaural wideband noise and narrowband correlated noise at different center frequencies. Moreover, the present study further builds a paradigm of a computational description model to explore the relationship between wideband and narrowband correlated noise with various center frequencies on the binaural ability to dynamically process acoustic details. Modeling the important functions of human hearing can provide an important opening for the development of a class of promising techniques. The main purpose of this modelling is to describe the evidence for the nature of human auditory organization and to explore the computational models which have been motivated by such evidence. Traditionally, behavioral tests are used to determine the sensitivity of the binaural gap detection in interaural correlation change. Although the behavioral temporal processing test is well established, the underlying mechanisms are not completely understood. These psychoacoustic characteristics of temporal storage capacity of acoustic features are helpful for understanding acoustic performance in reverberant environments. In addition, behavioral tests rely on the subjective reports of participants. An objective approach is useful for testing the sensitivity to the binaural gap detection in interaural correlation change. Hence, an intuitive approach is needed to evaluate the listener's ability to temporally store acoustic features. The computational models will be much more amenable to objective evaluation, and such evaluations will contribute to a more effective synergy between hearing research and speech processing research.

2. Materials and Methods

2.1. Participants

A total of 14 young students (9 females and 5 males, mean age = 22.786 years, from 20 to 25 years old) participated in this study. The participants all had normal and bilaterally balanced pure-tone hearing thresholds. All the participants gave their written informed consent to participate in the study. The experimental procedures of the present study involving human subjects were approved by the ethics review committee in Xiamen University.

2.2. Apparatus and Materials

The participants were tested individually in a quiet room. Participants listened to stimuli sequences via headphones (HD 265 linear, Sennheiser, Germany). Stimuli were generated with MATLAB (R2019b, the MathWorks Inc., Natick, MA, USA) software running on a PC (processor Intel(R) Core(TM) i7-8700 CPU @ 3.20ghz 3.19ghz, RAM 8 GB) with Windows 10 system. Gaussian wideband noise signal bursts were generated at a sampling rate of 48 kHz for the wideband noise stimulation condition. In the narrowband noise stimulation condition, signal stimuli had a fixed bandwidth of 1/3 octave and center frequencies of 0.2, 0.4, 0.8, 1.6, or 3.2 kHz. Each noise signal with a duration of 2000 ms was created, and presented over two headphones to listeners. The sound level was set at 60 dB sound pressure level, which was calibrated by the Larson Davis Audiometer Calibration and Electroacoustic Testing System (AUDit and System 824, Larson Davis, Depew, NY, USA) with "A" weighting.

2.3. Design and Procedure

Each trial in a testing block included two 2000 ms intervals of correlated noise (Figure 1a). In one interval, the right-headphone noise was the same as that of the left-headphone noise. In the other interval, the right-headphone noise was also identical to the left-headphone noise, except that the temporal middle of the 2000 ms noise was substituted with an interaural uncorrelated noise segment, in which this substituted noise segment introduced an instantaneous change in correlation (CIC) (Figure 1b) [10]. An interaural delay (ID) was introduced in the form of an interaural delay between the two intervals, which determined the interaural coherence of the interaural correlated noise (Figure 1c).

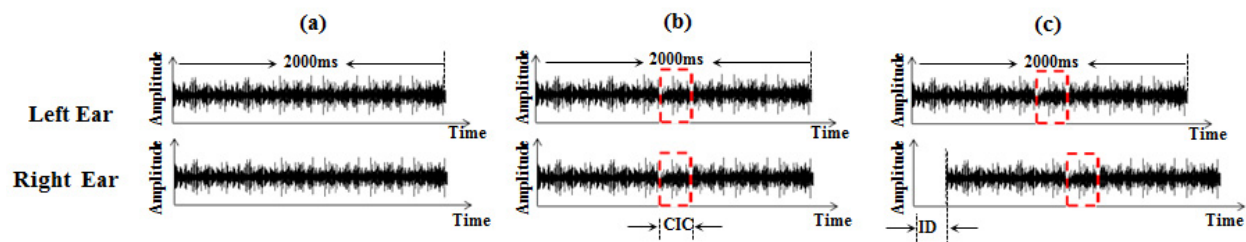


Figure 1. Illustration of the concept of CIC and ID in an example of steady-state noise. (a) Shows the two asynchronous but interaural correlated steady-state noise signals. (b) Shows an instantaneous change in correlation (CIC) inserted into the two asynchronous but interaural correlated steady-state noise signals. (c) Shows an interaural delay that was introduced on the basis of (b) for the measurement of the ability to transiently store acoustic features.

Participants performed an adaptive two-interval, two-alternative, forced-choice task track with visual indications of the listening intervals. The task of the participants was to identify which of the two intervals contained CIC (the binaural “gap”) immediately after the stimuli ended. During the experiment of this study, the duration of the CIC was fixed at 2000 ms according to the experiment explored in Huang et al., (2009) [10].

In each test, the probability that CIC was assigned to one of the two intervals was equal. The time interval separating the two intervals was set as 1000 ms. For each interval, the noise signals from the left earphone and those from the right earphone started simultaneously, and the duration of ID was systematically adjusted. New generated noise signals were used for each trial. The order of noise signal types was counterbalanced across participants using a Latin square order. The participant started a trial by pressing a command button on the keypad. The longest ID for CIC detection was measured and the duration length of ID for CIC detection was adjusted in a 3-down, 1-up procedure [31].

Screening procedures required the listener to identify each of the two 2000 ms intervals and they were asked to respond as soon as possible after the stimuli ended. The order name for each interval was presented on a computer monitor immediately prior to and during each trial.

The feedback on participants’ accuracy was displayed on a computer monitor immediately after each trial and responses were collected through a standard computer keyboard.

3. Computational Model and Logical Paradigm

A computational description model for initializing the ability to transiently store the acoustic details is described in the following.

- (1) Let the 2000 ms interval of correlated noise signals be denoted as $h_1(t)$, and the independent 200 ms noise segment that was interaural uncorrelated with $h_1(t)$ be denoted as $h_2(t)$. The noise signals presented to the left ear and right ear are indicated as $h_L(t)$ and $h_R(t)$, respectively, and were constructed by mixing $h_1(t)$ and $h_2(t)$ according to the following equations:

$$\text{Interval 1 : } h_L(t) = h_1(t); h_R(t) = h_1(t). \quad (1)$$

$$\text{Interval 2 : } h_L(t) = h_1(t); h_R(t) = \alpha h_1(t) + h_2(t) + \left(1 - \alpha - \frac{|h_2(t)|}{|h_1(t)|}\right) h_1(t). \quad (2)$$

- (2) Using the three-up, one-down procedure to obtain the longest ID for each session.

After three consecutive correct identification intervals containing CIC, the duration of ID was increased, and following one false identification, the duration of ID was decreased. According to this principle, let p be the probability of a positive response during a test, then the probability of obtaining an increased response sequence is denoted as p^3 , and

the probability of obtaining a decreased response sequence is indicated as $p^2(1-p) + p(1-p) + (1-p)$.

The procedure that systematically manipulates the interaural interval and computes the mean value over several sessions of the ID for each participant was developed to realize the function Φ , which was described in Levitt (1971) [31]. Let $\bar{t}_{IAI}$ denote the mean value over several sessions of the ID measures for each participant, which was calculated as follows:

$$\bar{t}_{IAI} = \Phi(H(t)) \quad (3)$$

- (3) Correlation coefficients [32] between the longest ID for wideband noise and those for various center frequencies of narrowband noise.

There are two stimulation conditions in this study: in narrowband noise stimulation conditions, for participant i ($i = 1, 2, \dots, n$), the longest ID for which each participant can detect 2000 ms CIC in narrowband noise with different center frequencies is indicated as $\bar{t}_i$ ($i = 1, 2, \dots, n$), and $\bar{t}_{ij}$ denotes the longest ID at which the CIC could be detected in the narrowband noise with the j th ($j = 1, 2, \dots, m$) kind of center frequency for participant i . In wideband noise stimulation conditions, T_i ($i = 1, 2, \dots, n$) denotes the longest ID at which the CIC in the wideband noise could be detected for participant i .

Let $\bar{\omega}_j$ denote the weight degree of contribution that the j th ($j = 1, 2, \dots, m$) kind of center frequency component makes to the auditory storage of the temporal storage of wideband noise features. Then, the calculation between the variability of the longest ID for wideband noise across participants and the longest ID for narrowband noise is indicated as follows:

$$\sum_{j=1}^m \bar{t}_{ij} \times \bar{\omega}_j = T_i (i = 1 \dots n) \quad (4)$$

where $\sum_{j=1}^m \bar{\omega}_j = 1$; $\bar{\omega}_j \gg 0$.

- (4) Find the best fit of $\bar{\omega}_j$ from the participants' data. This could be solved by a linear least-squares problem with linear constraints [33]. If given the data matrix t_{ij} , $\bar{\omega}_j$, then we could deduce the variability of the longest ID for wideband noise only from the narrowband noise across participants, and the mean longest ID at which the CIC in the wideband noise could be detected in wideband noise stimulation conditions. Let T' denote the mean longest ID at which the CIC in the wideband noise could be detected in wideband noise stimulation conditions, which was calculated as a function of each center frequency $\bar{t}_{ij}$ as follows:

$$T' = \frac{\sum_{i=1}^n T_i}{n}, T' = \frac{\sum_{i=1}^n \sum_{j=1}^m \bar{t}_{ij} \times \bar{\omega}_j}{n} \quad (5)$$

- (5) As T_i denotes the longest ID for participant i , let T_δ denote the ID at which the CIC in the wideband noise could be detected in wideband noise stimulation conditions, and let $\rho(T_\delta)$ denote the interaural correlation coefficient between the signals at the two ears, and let ϑ denote the maximum correlation coefficient. Thus, the maximum correlation coefficient ϑ was calculated as follows:

$$\vartheta = \max_{T_\delta}(\rho(T_\delta)) = \max_{T_\delta} \left(\frac{\int_{-\infty}^{\infty} h_L(t) h_R(t - T_\delta) dt}{\sqrt{\int_{-\infty}^{\infty} h_L(t)^2 dt} \sqrt{\int_{-\infty}^{\infty} h_R(t)^2 dt}} \right) \quad (6)$$

Let T_{μ_i} denote the duration threshold for binaural gap detected of participant i ; thus, the duration threshold for the binaural gap detected at various T_δ can be calculated as follows:

$$T_{\mu_i} = T_\lambda + T_\beta \times \rho(T_\delta) + T_{\gamma_i} + N(0, \sigma) \quad (7)$$

Thus, threshold T_{μ_i} includes a fixed intercept T_{λ} and an interaction term T_{β} . The decision variable β in multiple observation tasks is a weighted sum of the observation $\beta = \sum \xi_k x_k$, where $x_k (k = 1, 2, \dots, z)$ denotes the k th observation and ξ_k denotes the associated decision weight. Thus, the set of weights could represent a listener's decision strategy, and decision weights were estimated for each participant across conditions. Moreover, the differences in individuals were modeled by including random intercept γ_i , and $N(0, \sigma)$ denotes the measurement error, which is modeled as a normal distribution with standard deviation σ and zero mean.

In the present study, auditory trials recorded the quantitative measures of detected intervals and characterized the longest detectable instantaneous change in correlation for both wideband and narrowband correlated noise in the presence of the longest interaural delay.

4. Results

A one-way within-participant analysis of variance showed that there was a significant difference in the main effect of noise type ($F_{5,65} = 107.770, p < 0.001$). Post hoc analyses with the significance level set at 0.00333 indicated that there was a significant difference between the longest ID for wideband simulation noise and for narrowband simulation noise with center frequencies of 0.2, 0.4, 0.8, 1.6, and 3.2 kHz (all $p \leq 0.002$). As shown in Figure 2, the longest ID for center frequencies of 0.2, 0.4, or 0.8 kHz was significantly longer than that for narrowband simulation noise with center frequencies of 1.6 or 3.2 kHz. The results indicate that the performance of participants is generally better when the type of center frequency of narrowband simulation noise is in a low stage than that when the center frequency is in a high stage. This result also suggests that the duration time storage of low-frequency acoustic features lasts longer than that of high-frequency acoustic features.

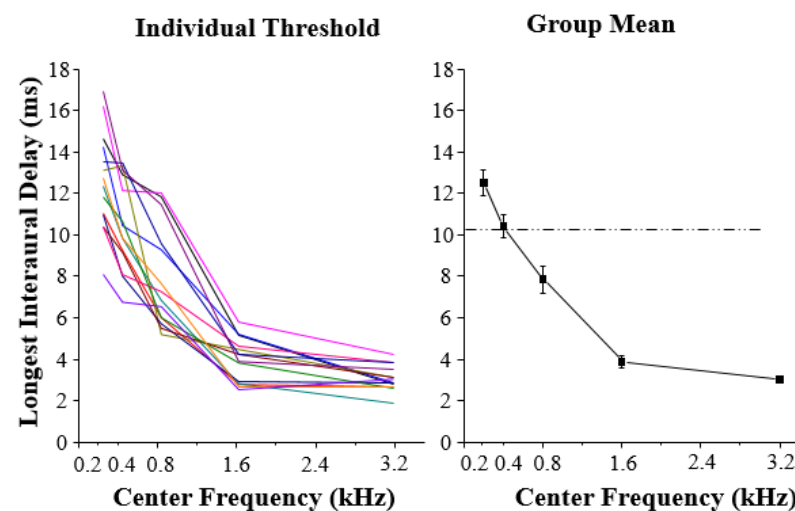


Figure 2. Individual thresholds (left panel) and group mean threshold (right panel) of the longest ID at which the CIC in the narrowband simulation noise could be detected as a function of the center frequency. The dash-dotted line represents the longest ID when the noise type was simulation wideband noise. The different color line represents the data of each individual.

To further investigate the relationship between the contribution that low-frequency components and high-frequency components make to the auditory storage of acoustical details in the dynamic process of auditory perception, a mathematical linear model with statistical computer packages in Matlab was used to generate the best fitting curve for participants' data. The linear model was adopted because it involves the simplest and seemingly most restrictive statistical properties: independence, normality, constancy of variance, and linearity. The group-mean longest ID at which the CIC could be detected in the narrowband noise as a function of center frequencies was examined. The best fitting

psychometric function was used for describing the relationship between the longest ID and the center frequencies, and the obtained value of quantitative goodness-of-fit of models to data was 0.9987. The simulation model can reveal that in the auditory process of low-level transient storage of acoustic details, low-frequency components and high-frequency components contribute differently to the auditory storage of broadband details.

Figure 3 shows the correlation between the longest ID at which a CIC could be detected for various center frequencies of narrowband simulation noise and that for wideband simulation noise. The results suggest that the temporal storing of low-frequency acoustic features lasted longer than that of high-frequency acoustic features. Furthermore, the longest ID for the wideband simulation noise was significantly correlated with that of each type of narrowband simulation noise (0.2 kHz, $r = 0.7267^{**}$; 0.4 kHz, $r = 0.8036^{**}$; 0.8 kHz, $r = 0.6407^{**}$; 1.6 kHz, $r = 0.5685^{**}$; 3.2 kHz, $r = 0.4616^{*}$), and the correlation coefficient increased as the center frequency of narrowband noise decreased.

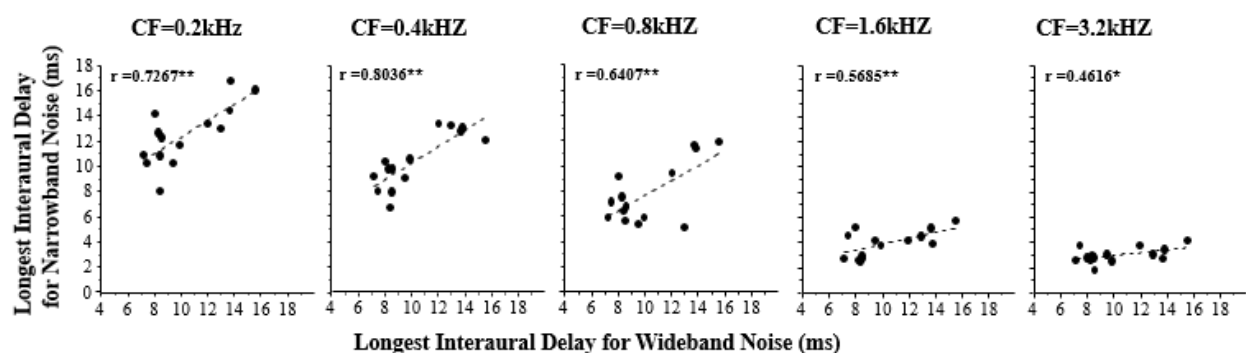


Figure 3. The longest ID at which a temporal break in correlation could be detected for each of the five center frequencies (CFs) of narrowband simulation noise as a function of that for wideband simulation noise. *, indicates significant at the level of 0.05; **, indicates significance at the level of 0.01.

Therefore, high-frequency components contribute less to the ability to temporally store wideband features than low-frequency components.

5. Discussion

The results in the present study show that when the duration of CIC is long enough, human listeners are easily able to detect the CIC embedded in wideband noise signals even when an ID was introduced. These results provide support to the deduction in Huang et al., (2009) [10] and Li et al., (2013) [11], showing that the duration threshold for detecting a sudden drop in band-limited noise labels becomes gradually larger as the center frequency increases from 0.2 to 3.2 kHz.

Since phase-locking tended to break down as the frequency increased [21,25], the longest IDs for detecting the CIC in high-frequency noise were shorter than those in low-frequency noise. As there was a reduction in the simultaneous correlation coefficient, the loss of phase-locking with the increase in frequency could partially explain the decrease in perceptual responses to the CIC.

The most interesting aspect of this result is that the longest interaural delay for narrowband noise signals was highly correlated with the longest interaural delay for wideband noise signals, and low-frequency components and high-frequency components all contribute to the auditory storage of wideband features. When the longest binaural delay is introduced, the binaural delay of the low-frequency component lasts longer than that of the high-frequency component. When a very short binaural delay is introduced, high-frequency components may not be detected. However, the duration of different frequency components has different benefits for auditory cues. Our results are also consistent with the binary theory of sound localization [34], suggesting that the time difference between sounds is an important cue only at low frequencies. It also supports the results in

Wightman et al., (1992) [35] and demonstrates the dominance of low-frequency temporal cues. Since the temporal dynamic of the acoustic signal recognition function was significantly correlated with the listeners' auditory ability to temporally store acoustic features [11], the relationship between the ability of auditory storage of wideband features and the contribution of each frequency component in the auditory processes of lower-level transient storage of acoustic fine features was important to discriminate what contributes to the sound localization and detection in the auditory scene analysis.

Furthermore, in a daily conversation environment with target stimuli and disruptive stimuli, which are full of high-frequency and low-frequency complex waveforms, the ability to transiently store acoustic features is crucial and functional in noisy conditions. Therefore, the auditory system must be tolerant of disruptive stimuli. Hence, the results of the present study may provide useful insights for further research on speech recognition in noise reverberant environments. The study may also be extended into investigation into the ability of auditory transient storage of acoustic feature interactions in both acoustic modeling studies and neurobiological studies.

6. Limitations and Future Work

Notably, some complex auditory signals (e.g., speech) contain both frequency and amplitude modulations. Moreover, the sensitivity to changes in interaural correlation in frequency-gliding sound has not been investigated yet. As shown, the auditory system is capable of integrating binaural information across different frequency channels, thus, we hypothesized that human listeners could hear a dynamic change in interaural correlation in noises with center frequency varying unidirectionally when both the spectral and temporal integrations are involved. However, whether the results and the calculation model will change stably within the same scope require further verification.

Although the results showed non-negligible tendencies, there was still a large part of the variability that could not be well explained. Future research will further refine the influencing factors and expand the number of participants to reduce this variability.

In addition, we only included normal hearing younger adults in the present study. The preliminary findings of the current exploratory study have implications for future studies on aging people, and/or hearing impairment populations. For example, the ability of transient auditory storage of acoustic features for older adults with normal hearing can also be estimated using the proposed "binaural integration paradigm" in the present study, and the computational model and logical paradigm in the current study can be adjusted to account for the influences of aging-related decline in binaural processing. Moreover, this paradigm may be used to explore the clinical implications, individuals with various degrees of hearing loss might be considered to be tested and investigate how the computational model can be modified to account for the hearing loss-related change.

7. Conclusions

The present study attempted to advance our understanding of how human listeners are able to temporally recognize rapid changes in interaural configurations. Experiments were conducted to examine whether the auditory ability to transiently store acoustic fine features could be maintained when an interaural delay was introduced.

It was found that a binaural gap CIC between interaural correlated narrowband or wideband noise was detectable even when introducing the longest interaural delay. The results indicate that the ability to transiently store information on acoustic features is frequency-dependent and this persistence of low-frequency features lasts longer than that of high-frequency features. The results of the present study also show that low-frequency information seems to dominate perception and is more distinguishable in length than high-frequency components. Moreover, the longest interaural delay for narrowband noise signals was highly correlated with the longest interaural delay for wideband noise signals in this dynamic process of auditory perception. Furthermore, the present study established a computational model to initialize the relationship between the contribution

of low-frequency and high-frequency components to the auditory storage of wideband features. The computational description models will provide a much more intuitive understanding of partial research of auditory processing, which will contribute to improving our understanding of how acoustic features are processed and how the sensitivity of the transient memory of acoustic features is represented.

Author Contributions: Conceptualization, M.W.; methodology, M.W. and N.Z.; validation, M.W. and N.Z.; formal analysis, M.W.; data curation, M.W. and N.Z.; writing—original draft preparation, M.W.; writing—review and editing, M.W. and N.Z.; funding acquisition, M.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Natural Science Foundation of Fujian Province (No. 2019J01002) and National Nature Science Foundation of China (No. 32071057; No. 61673322), and was partly supported by the Key project of National Key R&D project (No. 2017YFC1703303).

Institutional Review Board Statement: The experimental procedures of the present study involving human subjects were approved by the ethics review committee in Xiamen University.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The raw data supporting the conclusions of the current study will be made available by the corresponding author on reasonable request, without undue reservation.

Acknowledgments: We would like to express our sincere thanks to Liang Li at Peking University and the two anonymous reviewers for their helpful comments and suggestions for improving this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Popham, S.; Boebinger, D.; Ellis, D.P.; Kawahara, H.; McDermott, J.H. Inharmonic speech reveals the role of harmonicity in the cocktail party problem. *Nat. Commun.* **2018**, *9*, 2122. [\[CrossRef\]](#)
2. Qian, Y.-M.; Weng, C.; Chang, X.-K.; Wang, S.; Yu, D. Past review, current progress, and challenges ahead on the cocktail party problem. *Front. Inf. Technol. Electron. Eng.* **2018**, *19*, 40–63. [\[CrossRef\]](#)
3. Massaro, D.W. Preperceptual images, processing time, and perceptual units in auditory perception. *Psychol. Rev.* **1972**, *79*, 124–145. [\[CrossRef\]](#)
4. Kell, A.J.E.; Yamins, D.L.K.; Shook, E.N.; Norman-Haignere, S.V.; McDermott, J.H. A Task-Optimized Neural Network Replicates Human Auditory Behavior, Predicts Brain Responses, and Reveals a Cortical Processing Hierarchy. *Neuron* **2018**, *98*, 630–644. [\[CrossRef\]](#)
5. Lei, M.; Ding, Y. Interaural delay modulates the prepulse inhibition of the startle reflex induced by binaural gap in humans. *JASA Express Lett.* **2021**, *1*, 064401. [\[CrossRef\]](#)
6. Hartmann, W.M. Localization and lateralization of sound. In *Binaural Hearing*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 9–45.
7. Huang, Y.; Lu, H.; Li, L. Human scalp evoked potentials related to the fusion between a sound source and its simulated reflection. *PLoS ONE* **2019**, *14*, e0209173. [\[CrossRef\]](#)
8. Akeroyd, M.A.; Summerfield, A.Q. A binaural analog of gap detection. *J. Acoust. Soc. Am.* **1999**, *105*, 2807–2820. [\[CrossRef\]](#)
9. Culling, J.F.; Colburn, H.S.; Spurchise, M. Interaural correlation sensitivity. *J. Acoust. Soc. Am.* **2001**, *110*, 1020–1029. [\[CrossRef\]](#)
10. Huang, Y.; Huang, Q.; Chen, X.; Wu, X.; Li, L. Transient auditory storage of acoustic details is associated with release of speech from informational masking in reverberant conditions. *J. Exp. Psychol. Hum. Percept. Perform.* **2009**, *35*, 1618. [\[CrossRef\]](#)
11. Li, H.H.; Kong, L.Z.; Wu, X.H.; Li, L. Primitive Auditory Memory Is Correlated with Spatial Unmasking That Is Based on Direct-Reflection Integration. *PLoS ONE* **2013**, *8*, e63106. [\[CrossRef\]](#)
12. Boehnke, S.E.; Hall, S.E.; Marquardt, T. Detection of static and dynamic changes in interaural correlation. *J. Acoust. Soc. Am.* **2002**, *112*, 1617–1626. [\[CrossRef\]](#)
13. Naatanen, R.; Winkler, I. The concept of auditory stimulus representation in cognitive neuroscience. *Psychol. Bull.* **1999**, *125*, 826–859. [\[CrossRef\]](#)
14. Fan, L.; Kong, L.; Li, L.; Qu, T. Sensitivity to a Break in Interaural Correlation in Frequency-Gliding Noises. *Front. Psychol.* **2021**, *12*, 692785. [\[CrossRef\]](#)
15. Kolarik, A.J.; Culling, J.F. Measurement of the binaural auditory filter using a detection task. *J. Acoust. Soc. Am.* **2010**, *127*, 3009–3017. [\[CrossRef\]](#)
16. Verhey, J.L.; van de Par, S. Binaural frequency selectivity in humans. *Eur. J. Neurosci.* **2020**, *51*, 1179–1190. [\[CrossRef\]](#)

17. Harper, N.S.; McAlpine, D. Optimal neural population coding of an auditory spatial cue. *Nature* **2004**, *430*, 682–686. [[CrossRef](#)]
18. Larigaldie, N.; Christophe, R. A Nonparametric Bayesian Clustering Approach to Auditory Perception. Ph.D. Dissertation, Durham University, Durham, UK, 2021.
19. Salvati, D.; Drioli, C.; Foresti, G.L. Exploiting CNNs for improving acoustic source localization in noisy and reverberant conditions. *IEEE Trans. Emerg. Top. Comput. Intell.* **2018**, *2*, 103–116. [[CrossRef](#)]
20. Denham, S.L.; Winkler, I. Predictive coding in auditory perception: Challenges and unresolved questions. *Eur. J. Neurosci.* **2020**, *51*, 1151–1160. [[CrossRef](#)]
21. Goupell, M.J.; Hartmann, W.M. Interaural fluctuations and the detection of interaural incoherence: Bandwidth effects. *J. Acoust. Soc. Am.* **2006**, *119*, 3971–3986. [[CrossRef](#)]
22. Srinivasan, S.; Laback, B.; Majdak, P.; Arnoldner, C. Improving interaural time difference sensitivity using short inter-pulse intervals with amplitude-modulated pulse trains in bilateral cochlear implants. *J. Assoc. Res. Otolaryngol.* **2020**, *21*, 105–120. [[CrossRef](#)]
23. Baltzell, L.S.; Cho, A.Y.; Swaminathan, J.; Best, V. Spectro-temporal weighting of interaural time differences in speech. *J. Acoust. Soc. Am.* **2020**, *147*, 3883. [[CrossRef](#)]
24. Huang, Y.; Kong, L.Z.; Fan, S.L.; Wu, X.H.; Li, L. Both frequency and interaural delay affect event-related potential responses to binaural gap. *Neuroreport* **2008**, *19*, 1673–1678. [[CrossRef](#)]
25. Mossop, J.E.; Culling, J.F. Lateralization of large interaural delays. *J. Acoust. Soc. Am.* **1998**, *104*, 1574–1579. [[CrossRef](#)]
26. Dietz, M.; Wang, L.; Greenberg, D.; McAlpine, D. Sensitivity to Interaural Time Differences Conveyed in the Stimulus Envelope: Estimating Inputs of Binaural Neurons Through the Temporal Analysis of Spike Trains. *J. Assoc. Res. Otolaryngol.* **2016**, *17*, 313–330. [[CrossRef](#)]
27. Kong, L.; Xie, Z.; Lu, L.; Qu, T.; Wu, X.; Yan, J.; Li, L. Similar impacts of the interaural delay and interaural correlation on binaural gap detection. *PLoS ONE* **2015**, *10*, e0126342.
28. Liu, Z.; Wang, Q.; You, Y.; Yin, P.; Ding, H.; Bao, X.; Yang, P.; Lu, H.; Gao, Y.; Li, L. The role of the temporal pole in modulating primitive auditory memory. *Neurosci. Lett.* **2016**, *619*, 196–202. [[CrossRef](#)]
29. Körtje, M.; Baumann, U.; Stöver, T.; Weissgerber, T. Sensitivity to interaural time differences and localization accuracy in cochlear implant users with combined electric-acoustic stimulation. *PLoS ONE* **2020**, *15*, e0241015. [[CrossRef](#)]
30. Gelfand, S.A. *Hearing: An Introduction to Psychological and Physiological Acoustics*; CRC Press: Boca Raton, FL, USA, 2017.
31. Levitt, H. Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* **1971**, *49* (Suppl. S2), 467–477. [[CrossRef](#)]
32. Lee Rodgers, J.; Nicewander, W.A. Thirteen ways to look at the correlation coefficient. *Am. Stat.* **1988**, *42*, 59–66. [[CrossRef](#)]
33. Björck, Å. Least squares methods. *Handb. Numer. Anal.* **1990**, *1*, 465–652.
34. Strutt, J.W. On our perception of sound direction. *Philos. Mag.* **1907**, *13*, 214–232.
35. Wightman, F.L.; Kistler, D.J. The dominant role of low-frequency interaural time differences in sound localization. *J. Acoust. Soc. Am.* **1992**, *91*, 1648–1661.