The effect of backward noise on lexical tone discrimination in Mandarin-speaking amusics

Zixia Fan¹, Jing Shao², Weigong Pan¹, Min Xu³, Lan Wang¹

¹Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
²Department of English Language and Literature, Hong Kong Baptist University
³Institute of Corpus Studies and Applications, Shanghai International Studies University

fanxiaowan007@outlook.com, jingshao@hkbu.edu.hk, blockend@foxmail.com, xumin@shisu.edu.cn, lan.wang@siat.ac.cn

Abstract

Congenital amusia is a neurogenetic disorder, affecting music pitch processing. It also transfers to the language domain and negatively influences the perception of linguistic components relying on pitch, such as lexical tones. It has been well established that unfavorable listening conditions impact lexical tone perception in amusics. For instance, both Mandarin- and Cantonese-speaking amusics were impaired in tone processing under simultaneously noisy conditions. Backward noise is one of the adverse listening conditions, but its interference mechanism is distinct from the simultaneous masking. Therefore, it warrants more studies to explore whether and how backward noise affects tone processing in amusics. In the current study, eighteen Mandarin-speaking amusics and 18 controls were tested on discrimination of Mandarin tones under two conditions: a quiet condition involving relatively low-level processing and a backward masking condition involving high-level processing (e.g., tone categorization) where a native multi-talker babble noise was added to target tones. The results revealed that amusics performed similarly to controls in quiet conditions, whereas their performance was found poorer in backward masking conditions. These findings shed light on how adverse listening environments influence amusics’ lexical tone processing and provided further empirical evidence that amusics may be impaired in the high-level phonological processing of lexical tone.

Index Terms: congenital amusia, backward masking, lexical tone discrimination, Mandarin.

1. Introduction

Approximately 1.5% of the population [1] suffers from lifelong impairment of the ability to perceive and process musical pitch in the absence of brain injury [2]. This disorder is termed congenital amusia (henceforth, amusia) [3]. The core deficit of amusia concerns the processing of musical pitch [4], such as fine-grained pitch discrimination [5], pitch direction [6], as well as pitch memory [7].

Empirical evidence has shown that amusia should be viewed as a syndromic disorder, and it is also frequently accompanied by deficits in sound processing, not specific to music [8]. For instance, speech intonation processing [6], [9] and emotional status identification [10], in which pitch serves as a critical cue. In addition, individuals with amusia also have difficulties in lexical tone processing regardless of whether their native language is tonal or non-tonal. For instance, Nguyen et al. [11] found that native French speakers with amusia performed significantly worse than the typical controls on discrimination of Mandarin lexical tones, which indicates that the deficits in music have extended to lexical tone perception. Similarly, French-speaking amusics also showed impairments in the lexical tone discrimination tasks of Thai [12]. These studies revealed that non-tonal individuals with amusia have difficulties in lexical tone processing. For native tonal language speakers with amusia, many studies also suggested that they have behavioral difficulties with linguistic tone processing. For example, Nan et al. [13] found that nearly half of Mandarin-speaking amusics had impairments in the identification and discrimination of Mandarin lexical tones. In another study, Jiang et al. [14] found that native Mandarin speakers with amusia showed no improvement for discrimination pairs with crossing classification boundaries for either speech or nonlinguistic analogues, indicating a lack of categorical perception. Similar findings were reported in Cantonese-speaking amusics [15], [16] [17], [18], suggesting that amusics had a significant impairment in lexical tone perception in different scenarios.

Some studies have confirmed that amusics are also impaired in high-level phonological processing of lexical tones. Among them, a line of research focused on the effects of acoustic variations on lexical tone perception, which involved the process of extracting the tone category from the incoming acoustic signal and mapping it to the stored phonological representations [17], [18]. For instance, it was found that Mandarin-speaking amusics discriminated tone pairs which were carried by the same syllable similarly well to controls, while their performance was impoverished in the condition where the tone pairs were carried by different syllables [13]. Furthermore, Cantonese-speaking amusics were also impaired in higher-level phonological processing of lexical tones in native tone identification and discrimination under high syllable variation condition [17], [18]. According to these findings, it is clear that amusics experienced a deficit when the demand on phonological processing was increased.

The other line of the research examined the effect of noise on lexical tone perception. It is well documented that unfavorable listening conditions impact lexical tone perception [19] in a significant way, as it increases the demand for tone sensitivity, phonological representations, phonological memory, and attentive processing. For instance, Shao et al. [15] found that Cantonese speakers with amusia showed degraded performance compared with the controls in tone identification and discrimination tasks when Cantonese tones were embedded in noise. Similar results were found in Mandarin-speaking amusics’ lexical tone identification by
Tang et al. [20]. These findings suggested that tone perception of amusics was impoverished under the simultaneous noise condition.

Yet, it is still unclear whether backward noise would affect lexical tone perception in amusics. Research has shown that the masking mechanisms underlying the simultaneous noise and the backward noise to probe stimuli were completely different. The simultaneous noise runs throughout the presentation of the target stimuli [21], that is, it influences the whole brain processing of the stimulus. The magnitude of its masking effect is determined by the association between the acoustic representations (e.g., frequency and intensity) of noise and stimuli [21]. On the other hand, the backward noise inhibits the brain’s post-processing of the stimulus since the processing would be terminated upon hearing the followed noise [22]. An important factor contributing to the magnitude of this masking effect is the interval between the stimulus and the noise [23]. To acquire further insight into lexical tone processing in amusics, it is worthwhile to investigate how backward noise affects processing in amusics.

It has been documented that backward noise has a negative impact on speech perception in individuals with disorders. For example, children with language impairment are particularly prone to the influence of backward noise. They had significantly higher signal thresholds (dB) than their peers with normal speech abilities [24]. Children with reading disabilities also had deficit in speech processing under backward masking condition [25]. This is because backward noise exerts a significant masking effect on the probe stimuli which increases the difficulty in extracting abstract speech categories, short-term memory store, as well as attentive processing [27], [28]. It is predicted that this more demanding task might be particularly challenging for amusics. To test this hypothesis, we investigated whether and to what extent the backward noise influences lexical tone processing in Mandarin-speaking amusics.

2. Method

2.1 Participants

In the experiment, 18 amusics and 18 matched controls were recruited. All participants were native Mandarin speakers. They were all right-handed, had normal hearing as measured with an audiometer (GSI 18), and reported no history of neurological illness and background in musical training. Amusics and controls were selected according to the Montreal Battery of Evaluation of Amusia (MBEA) [26]. It consists of six parts: three of them are pitch-based tests (Scale, Contour, Interval), two of them are duration-based tests (Rhythmic and Metric Tasks), and the last one is Memory Task. In the MBEA test, none of the amusics scored higher than 70 in the global score, which is the mean of all six subtests, whereas all controls scored 89 or higher. Amusics’ global scores were significantly lower than controls’, $t(34) = 18.03$, $p < 0.001$, according to independent-samples t-tests. Furthermore, amusics also scored significantly lower than controls in each subtest (all $p < 0.001$). Participants’ characteristics are summarized in Table 1. All the participants provided their written consent forms for the study which was approved by the Human Subjects Ethics Sub-committee of Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences.

2.2 Stimuli

In this experiment, the stimuli were 12 words contrasting four Mandarin tones (high level tone-T1 /55/, mid rising tone-T2 /35/, low dipping tone-T3 /214/, and high falling tone-T4 /51/). On three syllables (/tuæ/, /leɪ/ and /tæʊ/). A female Mandarin speaker was recorded reading aloud the words for six times in a soundproof booth. One clear token was selected for each word, and it was normalized in duration to 400 ms and in mean intensity to 60 dB using Praat [27], respectively. The Fo trajectory was preserved after the normalization. The stimuli included two types of conditions: one type was followed by multi-talker babble noise lasting 500 ms at 75 dB (backward masking condition), while the other type was unaccompanied by any noise (quiet condition).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Amusics</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male / Female (Total)</td>
<td>11 / 7 (18)</td>
<td>12 / 6 (18)</td>
</tr>
<tr>
<td>Age (Range)</td>
<td>23.3 ± 3.3 yr</td>
<td>20.5 ± 2.5 yr</td>
</tr>
<tr>
<td>MBEA, Mean (SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scale</td>
<td>59.9 (7.7)</td>
<td>96.2 (3.4)</td>
</tr>
<tr>
<td>Contour</td>
<td>61.7 (10)</td>
<td>95.9 (3.5)</td>
</tr>
<tr>
<td>Interval</td>
<td>59 (8)</td>
<td>95.2 (4.5)</td>
</tr>
<tr>
<td>Rhythm</td>
<td>71.5 (13.5)</td>
<td>96.4 (3.4)</td>
</tr>
<tr>
<td>Meter</td>
<td>50.7 (17)</td>
<td>84.2 (14.3)</td>
</tr>
<tr>
<td>Memory</td>
<td>76.4 (12.3)</td>
<td>96.3 (5.7)</td>
</tr>
<tr>
<td>Global</td>
<td>63.1 (6.6)</td>
<td>94.1 (3.1)</td>
</tr>
</tbody>
</table>

2.3 Procedure

The experiment generated by E-Prime 2.0 consisted of two blocks corresponding to the two conditions, and the presentation order of them was counterbalanced across the participants. Prior to the experiment, participants were given a few practice trials (four tones carried by /zha/) to familiarize them with the procedure.

During the procedure, 4 words in each set were divided into 6 different tone pairs and 4 same tone pairs. Each set was presented in a sub-block, generating 3 sub-blocks in total, so that confusion could be avoided. In each sub-block, the different tone pairs were repeated twice, and the same tone pairs were repeated 3 times, resulting in an equal number of different and same tone pairs, which were intermixed and presented randomly. Each tone pair began with an 800-ms blue fixation on the computer screen, followed by the presentation of the first stimulus via the headphones. Then, an 800-ms red fixation appeared, followed by the second stimulus. There was no interval between the end of the first stimulus and the beginning of the red fixation. The participants were instructed to determine whether the two words carried the same tone or different tones by pressing "left arrow" (same) or "right arrow" (different) on a keyboard within 5 seconds. Results were collected in terms of accuracy and response time (RT).

2.4 Data analysis

For the results, accuracy and RT were analyzed. Accuracy was the percentage of correctly identified trials for each condition per participants. RT was determined by measuring the offset of the second stimulus in a tone pair until a response was
made. In addition, incorrect trials were excluded from RT analysis, as well as trials exceeding three SDs of each condition (4%).

For statistical analysis of accuracy, general linear mixed-effects models were fitted. In our model selection process, we used backwards elimination of non-significant effects. In the full model, group (amusics and controls), condition (backward masking condition and quiet condition) and tone pair (T1 & T1, T1 & T2, T1 & T3, T1 & T4, T2 & T2, T2 & T3, T2 & T4, T3 & T3, T3 & T4, T4 & T4) were fixed factors, as well as interaction between them, and random effects were subject and syllable. Based on Bayesian Information Criterion (BIC) values [28] and Akaike Information Criterion (AIC) values [29], the optimal model was estimated, that is the model included group, condition and two-way interaction as fixed effects, and subject as the random effect. For statistical analysis of RT, linear mixed-effects models were computed. The procedures were the same as those described above. Additionally, in the appropriate model, group, condition, and two-way interaction were all included as fixed effects, with subject intercept as a random effect. The analysis above were performed with R (R Core Team, 2014), using the lme4 package [30], emmeans package [31], and the tidyverse package [32].

3. Results

![Figure 1: The discrimination accuracy of the amusic and control groups. Dotted lines indicate the chance level accuracy (0.5). BMC = Backward masking condition; QC = Quiet condition.](image_url)

Figure 1 shows the lexical tone discrimination accuracy for amusics and controls under the backward masking condition and the quiet condition. For accuracy, there were significant main effects of condition, $\chi^2(1) = 56.049, p < 0.001$, and group, $\chi^2(1) = 4.901, p < 0.05$, as well as significant two-way interactions between them, $\chi^2(1) = 8.8611, p < 0.01$. Furthermore, post hoc analysis of the interaction between group and condition were performed using emmeans package [29] in R. The results showed that amusics performed significantly worse than controls in the backward masking condition ($p < 0.01$), while no statistical difference was found between performance of amusics and controls in the quiet condition ($p = 0.8914$). For the amusic group, the backward masking had a significant effect on their performance compared to the quiet condition ($p < 0.0001$). But for the control group, there was no significant difference in their responses neither in the backward masking condition nor the quiet condition ($p = 0.3669$), that is they performed equally well in the two conditions. No other effects were significant.

Figure 2 shows RT of amusics and controls under the two conditions. There was a significant main effect of condition, $\chi^2(2) = 74.789, p < 0.001$, and significant two-way interactions between group and condition, $\chi^2(1) = 17.459, p < 0.001$. The results of post hoc analysis on the interaction effect showed that RTs elicited by the backward masking condition were significantly shorter than those in the quiet condition, both within the amusic group ($p < 0.01$) and the control group ($p < 0.0001$). No other effects were significant.

![Figure 2: The discrimination RT of the amusic and control groups. BMC = Backward masking condition; QC = Quiet condition.](image_url)

4. Discussion

The present study examined the effect of backward noise on lexical tone discrimination in Mandarin-speaking amusics and matched controls. Overall, the results showed that Mandarin-speaking amusics were impaired in lexical tone discrimination under the backward masking condition, but not in the quiet condition.

For the lexical tone processing, Shuai and Gong (2014) [33] proposed a three-stage model based on the experiments in which auditory stimuli evoke ERP components in the brain. In the first stage, acoustic signals of tone and syllable are received by the brain. In the second stage, further processing takes place, which involves the prediction of pitch contour, normalization and categorization of the lexical tone, as well as syllable structure. In the third stage, linguistic representations are integrated, helping to confirm the recognized tone and syllable. According to this model, a high-level phonological processing is indeed required in stage two and three. As mentioned, the backward noise interferes with the brain’s processing of probe stimuli upon hearing the followed noise [22]. Consequently, backward noise may inhibit lexical tone post-processing, which consists of a series of high-level phonological processing. Based on these results, amusics’ poor performances in lexical tone discrimination under the backward masking condition suggests that they have the difficulty in high-level phonological processing.

We found that amusics performed significantly worse than controls in accuracy under the backward masking condition when they discriminated Mandarin tone pairs. This means that their discrimination sensitivity was decreased compared to the
controls. Additionally, these results are consistent with previous findings on Cantonese- [15] and Mandarin-speaking amusics [20] to a large extent, namely that amusics were less accurate on lexical tone perception in noisy environments. However, the mechanisms of lexical tone processing underlying the noise condition and the backward masking condition are different. Primarily, the fundamental difference between the two types of noises is their masking approaches to the probe stimulus. In backward masking, a target sound is presented immediately followed by another sound which masks the effect of the first sound [21], [22], [23]. Simultaneous masking, on the other hand, refers to a sound that is made inaudible by a noise or unwanted sound that lasts the same duration as the original sound [23]. Furthermore, due to the overlap of tones and noise in simultaneous noise conditions, the brain is challenged to extract acoustic cues of the impaired tonal signals in the preliminary processing, but the tonal signals are intact under the backward masking condition. Apart from this, the backward masking condition only prevents the post-processing of lexical tone related to high-level phonological processing, while the simultaneous masking affects the entire lexical tone processing including low-level pitch processing and high-level phonological processing.

Therefore, the deficits of amusics under the two adverse conditions reflect two different types of processing, which may further shed light on the mechanism of the tone perception impairment in amusics. For the backward masking condition, it is possible that Mandarin-speaking amusics have difficulties in high-level phonological processing or certain parts of it, such as pitch patterns perception, lexical tone normalization and categorization, as well as linguistic information integration when they conduct the lexical tone discrimination. For the simultaneous noisy condition, however, they might show difficulty in low-level pitch processing related to perceiving and extracting the impaired acoustic cues, as well as in high-level phonological processing.

Furthermore, the results also echo with the findings from studies of language-impaired children and reading-disabled children. It should be noted that although all three groups demonstrated deficits in speech perception under backward masking conditions, it might be different kinds of deficiencies. Language-impaired children are more susceptible to the influence of noise because they lack both the optimal language representations and the auditory processing mechanisms unlike speakers without language impairment, and thus the deficit is more pronounced in adverse conditions [21]. Children with reading disabilities were more likely to experience a non-language auditory impairment than children without reading impairments, which was mainly exhibited in the spectral component and specific temporal to the auditory processing [25]. This deficit appears to be due to problems with written language processing that cause auditory processing deficits or overlap between the two processing within the central nervous system. Amusics, on the other hand, possess essentially complete language processing functions similar to normal musical controls [15]. Instead, it is more likely that high-level phonological awareness, such as phonological representation sensitivity, phonological memory and attentive processing, is impaired in amusics, which results in degraded performance on tone discrimination under the backward masking condition.

For the quiet condition, Mandarin speakers with amusia performed equally well as controls in Mandarin tone discrimination. Similar results were obtained in the study of Nan et al. in 2010 [13]. These findings suggested that Mandarin-speaking amusics have no flaws in tone discrimination tasks in the quiet condition when the tones were constantly carried by the same segments. In essence, if the base segments are identical in the tone discrimination, the critical processing is the tone comparison and it is not necessary to consider high-level tonal processing, such as tone normalization and categorization [17]. But Cantonese-speaking amusics were impaired in tone discrimination under the same conditions [17]. Although it has been documented that tone discrimination in both Mandarin and Cantonese is determined by the dimensions of height and direction on F0 [34], six tonal categories in Cantonese result in a much denser tonal space, while, Mandarin has four tones with a distinct direction on the frequency scale [17], leading to a much wider tonal space. Therefore, compared to the Cantonese system, the Mandarin tonal system appears to have a more distinguishable structure resulting in a greater distance between categories in a certain space, which may facilitate tone discrimination [35].

It is notable that RTs in amusics group and control group were shorter in the backward masking condition compared to the quiet condition and RTs elicited by controls were significantly shorter. As participants’ RTs were recorded from the offset of each trial regardless of in the quiet condition or the backward masking condition, it is not difficult to find that there were 500-ms of noise before recording participants’ responses in the backward masking condition. Therefore, the fact that RTs for amusics and controls were shortened is not surprising under this condition. A common finding is that correct responses took shorter RT than incorrect responses in complex attentional cognitive tasks [36]. Combining the performance of two groups on the accuracy of tone discrimination, it confirmed that the effect of backward noise was greater on Mandarin-speaking amusics than on Mandarin-speaking controls, indicating that Mandarin-speaking amusics appear to lack adequate anti-interference skills under this situation.

5. Conclusion

In the current study, lexical tone processing abilities of amusics and controls were tested in two conditions: quiet and backward masking. The results indicated that Mandarin-speaking amusics have deficits in high-level processing in lexical tone discrimination under the backward masking condition. It is possible that their abilities in the prediction of pitch contour, normalization, categorization of the lexical tone, or the integration of linguistic representations might be degraded. These findings deepened our understanding about the mechanism of deficient lexical tone processing in congenital amusia.

6. Acknowledgement

This work was supported by grants from the National Key R&D Program of China (2020YFC0204100), the National Natural Science Foundation of China (NSFC 11904381), the National Natural Science Foundation of China (NSFC U1736202) and Shenzhen Science and Technology Program (KQTD2020082013106007).
References


