EFFECTS OF MUSICAL EXPERIENCE ON THAI RATE-VARIED VOWEL LENGTH PERCEPTION

Angela Cooper[1], Yue Wang[2] & Richard Ashley[3]

(1) Department of Linguistics, Northwestern University;
(2) Department of Linguistics, Simon Fraser University; (3) Bienen School of Music, Northwestern University
akcooper@u.northwestern.edu; yuew@sfu.ca; r-ashley@northwestern.edu

ABSTRACT

Musical experience has been demonstrated to play a significant role in the perception of non-native speech contrasts. The present study examined whether or not musical experience facilitated the normalization of speaking rate in the perception of non-native vowel length contrasts. Musicians and non-musicians were first briefly familiarized with Thai vowel length distinctions before completing identification and AX discrimination tasks with items contrasting in vowel length at three speaking rates. Results revealed that musicians significantly outperformed non-musicians at identifying and discriminating non-native rate-varying length distinctions, suggesting that their attunement to rhythmic and temporal information in music transferred to facilitating their ability to perceive non-native temporal speech contrasts at varying speaking rates.

Keywords: Speaking rate, vowel length, Thai, musical experience, perception

1. INTRODUCTION

Adult learners face numerous difficulties when perceiving second language (L2) speech contrasts. L2 phonetic features that are not used or are not prominent in the first language (L1) are often difficult to perceive [8]. However, in addition to linguistic factors, these difficulties have been found to be mediated by a number of extralinguistic factors, such as musical experience. In particular, musicians have demonstrated superior performance relative to non-musicians at perceiving a variety of speech contrasts, including lexical tones (e.g. [7]), vowels and consonants (e.g. [9]) and temporal speech contrasts, such as voice onset time and vowel duration [3]. This influence of musical training on linguistic processing has been attributed to the enhancement, through domain-specific experience, of domain-general auditory mechanisms, which serve to enhance the processing of acoustic features shared by both music and speech [1].

Despite the abundance of research on music and linguistic processing, relatively little work has dealt with speech materials spanning temporal domains larger than single syllables or words. Given that speech rarely consists of only single word increments, the ability to incorporate (lower- and higher-level) contextual information is paramount for effective speech communication. One area that has not yet been investigated is the impact of musical experience on the ability to normalize speaking rate, particularly for non-native temporal contrasts that are affected by rate variations. This ability requires listeners to track the "tempo" of the preceding speech context in order to make judgments about the duration identity (e.g. long vs. short) of the target segment. Rate-varied temporal contrasts have been found to be particularly challenging for non-native listeners [5], where they have greater difficulty identifying vowel lengths at faster speaking rates relative to slower rates.

Musically-trained listeners' enhanced auditory acuity could be facilitative or inhibitory when confronted with rate variability. Their ability to detect fine-grained acoustic distinctions has been found to be beneficial in discriminating non-native durational contrasts with relatively minimal variation [9]; however, it is conceivable that this ability to perceive minute distinctions could make it more difficult for them to ignore and abstract over variation. Alternatively, their experience with normalizing for rate in musical phrases, such that they would be able to identify, for instance, quarter or half notes in musical pieces played at different tempos, might transfer to the linguistic domain and facilitate their ability to extract vowel-to-word duration ratios, as a cue to vowel length, that remain relatively stable across speaking rates.

The current study compared the performance of musically-trained and untrained listeners on perceiving naturally-produced rate-varied vowel length distinctions in Thai. Participants completed both identification and AX discrimination tasks with items embedded in a carrier sentence to provide listeners with cues to speaking rate. The inclusion of between- and within-category discrimination pairs allowed us to examine whether listeners were forming length categories or relying on lower-level acoustic differences. Based on prior work suggesting that musical training can enhance acuity for certain

auditory features such as duration and pitch [6], we predicted that musicians would outperform non-musicians in both tasks, such that their attunement to rhythmic and temporal distinctions in music would facilitate their ability to normalize for speaking rate in non-native speech perception. Alternatively, it is also conceivable that their superior auditory acuity would enhance their sensitivity to within-category differences, which would predict better discrimination of within-category distinctions relative to non-musicians.

2. METHODS

2.1. Participants

Twenty-six American English listeners, with no prior knowledge of Thai or any other language with phonemic length distinctions, were included in this study. They were divided in two groups of listeners: non-musicians and musicians (n=13 in each). Non-musicians ("NM") had less than 3 years of musical experience and no experience within the last 5 years (8 females; $mean_{age}$=21 years; $mean_{musexp}$=1 year). Musicians ("M") were defined as having at least 7 years of continuous musical training and the current ability to play an instrument, ranging from 7 to 20 years of experience (9 females; $mean_{age}$=22 years; $mean_{musexp}$=13 years).

2.2. Stimuli

All stimuli were recorded by a female native speaker of Thai in a sound-attenuated booth at a 44.1 kHz sampling rate. Stimuli included 6 monosyllabic Thai pseudoword minimal pairs, contrasting in vowel length (e.g., /sik/ vs. /si:k/). Each pair was matched for lexical tone and contained common Thai/English phonemes. Each item was produced in the carrier sentence /tʃʌn put kʌm wa ____ ik ti/ "I say the word ____ again" at slow, normal and fast rates of speech. Rate instructions were taken from [4]. The speaker was instructed that a "slow" rate was the slowest rate possible without any obvious pauses in the sentence, and the "fast" rate was described as the fastest rate possible without making speech errors. Table 1 provides mean vowel durations across target items for short and long vowels for 3 speech rates.

The identification and discrimination tasks used 3 minimal pairs each. "Same" and "different" AX discrimination pairs were created, whereby two sentences were placed in succession, separated by 500 ms. For the "same" pairs, two repetitions of the same sentence containing target items of the same length and speaking rate were used. The "different" pairs were constructed in 3 conditions: 1) Between Category-Same Rate, 2) Between Category-Different Rate, and 3) Within Category-Different Rate. For Condition 1, each trial contained a target item pair at the same speaking rate, differing only in vowel length (e.g., slow rate /sik/ and slow rate /si:k/). Condition 2 contained trials where target items differed in vowel length but also in speaking rate. Three rate patterns were created: 1) Slow (short) + Fast (long), 2) Norm (short) + Fast (long), and 3) Slow (short) + Norm (long). Finally, Condition 3 contained target item pairs of the same vowel length but at different speaking rates. The same three rate patterns were included for both lengths (e.g., Norm (short) + Fast (short)). All of the pairs were counterbalanced for order of presentation.

Table 1: Mean vowel durations for target stimuli for long and short vowels at each speaking rate.

Rate	Short (ms)	Long (ms)
Slow	130	265
Normal	98	169
Fast	72	115

2.3. Procedure

Participants completed both the identification and discrimination tasks in a sound-attenuated booth. Stimuli were played free-field over Alesis Point 7 speakers at a comfortable listening volume. Task order was counterbalanced across listeners.

The vowel length identification task was a two-alternative forced-choice task, where participants listened to target items presented in their carrier sentence and indicated whether they heard a long-vowel word or a short-vowel word by pressing a number on a computer keyboard. The screen displayed the carrier sentence in both Thai and English, along with its phonetic transcription, and a choice of two words (displayed in both Thai and English). Participants had 2 seconds to make a response. The task consisted of 72 randomized trials (3 syllables x 3 rates x 2 lengths x 4 repetitions), which were divided into two blocks of 36 trials each.

For the AX discrimination task, participants heard pairs of sentences containing target word pairs and were asked to indicate whether the target word in the second sentence was the same or different from the first target word. They had two seconds to make a response. On the screen, the carrier sentence was displayed along with a choice of "Same word" or "Different word". Participants completed 36 trials in Condition 1 (3 syllables x 3 rates x 2 order counterbalancing x 2 repetitions), 72 trials in Condition 2 (3 syllables x 3 rate patterns x 2 length counterbalancing x 2 rate counterbalancing x 2 repetitions), 72 trials in Condition 3 (3 syllables x 3

rate patterns x 2 lengths x 2 rate counterbalancing x 2 repetitions) and 36 trials in Condition 4 (3 syllables x 3 rates x 2 lengths x 2 repetitions) for a total of 216 trials. Trials from all 4 conditions were randomly presented over the course of 6 blocks of 36 trials each.

To acquaint listeners with task procedures, the identification and discrimination tasks were preceded by brief familiarization sessions, including task instructions and practice trials. The trials were identical to the main task trials, except they provided feedback on the accuracy of participants' responses as well as the correct answer after each trial.

3. RESULTS

For the identification task, the proportion of correct responses was tabulated for each length and speaking rate for musician and non-musician groups (Table 2). Overall accuracy across rates and lengths revealed a substantial difference between groups (musicians: 81% vs. non-musicians: 60%). However, there appeared to be a strong bias for both groups to respond "short", with overall lower accuracy rates for long vowels (61%) relative to short vowels (80%).

Table 2: Mean percent correct vowel length identification (standard error in parentheses) for each condition

Rate & length	Non-musician	Musician
Long		
Fast	26.9 (3.6)	49.4 (3.9)
Normal	51.9 (4.0)	76.2 (3.3)
Slow	75.0 (3.5)	89.3 (2.4)
Short		
Fast	84.0 (2.9)	95.2 (1.6)
Normal	60.9 (3.9)	85.1 (2.8)
Slow	60.9 (3.9)	91.1 (2.2)

To account for this potential bias, the proportions of hit rates (defined as the proportion of short vowels to which participants correctly responded "short") and of false alarms (defined as the proportion of long vowels to which participants incorrectly responded "short") were used to compute *d-prime* values.

These data were analysed using linear mixed-effects regression models (LMER; [2]) with *d'* scores as the dependent variable. A contrast-coded fixed effect of Group and Helmert contrast-coded fixed effects for Rate (A: Fast vs. Norm + Slow; B: Norm vs. Slow) along with their interactions with Group were included in the model. The model also contained a random intercept for participant as well as a random slope by participant for Rate. Model comparisons were performed to determine whether the inclusion of each of these fixed factors and their interactions made a significant contribution to the model.

As illustrated in Figure 1, there was a robustly significant effect of Group (β=1.4, SE β=0.26, $\chi^2(1)=19.906$, $p<0.001$), with musicians significantly outperforming non-musicians across speaking rates. There were also significant effects of Rate across groups, where listeners were significantly worse at identifying vowel lengths at a fast speaking rate relative to slower speaking rates (β=0.78 SE β=0.18, $\chi^2(1)=14.667$, $p<0.001$) and worse at a normal rate relative to a slow rate (β=0.79, SE β=0.14, $\chi^2(1)=21.603$, $p<0.001$). Neither of the group x rate interactions were significant ($\chi^2 < 3.01$, $p>0.083$).

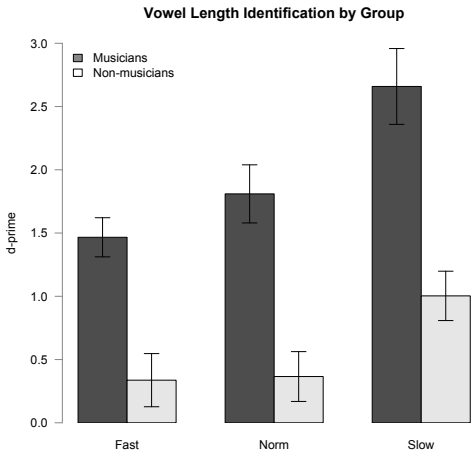

Figure 1: Mean *d'* scores (+/- 1 standard error) for each speaking rate by group

For the discrimination task, the proportions of hit rates (proportion of "different" trials that participants correctly indicated were "different") and of false alarms (proportion of Condition 4 "same" trials to which participants incorrectly indicated were "different") were used to calculate *d-prime* scores for each rate pattern (Table 3).

LMER models with the same fixed and random effects structure as for the identification task analyses was used to analyse Condition 1 (Between Category-Same Rate). A significant effect of Group was found (β=1.12, SE β=0.19, $\chi^2(1)=21.572$, $p<0.001$), with musicians better able to discriminate short and long vowels in Thai across rates relative to the non-musicians. A significant Rate A effect (Fast vs. Norm + Slow) was found ($\chi^2(1)=31.215$, $p<0.001$) along with a significant Group x Rate A interaction (β=1.07, SE β=0.32, $\chi^2(1)=9.2069$, $p=0.002$), with a smaller difference in accuracy between fast and slower rates for non-musicians as compared to musicians. None of the other effects or

interactions reached significance ($\chi^2(1)<3.27$, $p>0.07$).

Similar LMER models were constructed for Condition 2 (Between Category-Different Rate), with Group and Helmert contrast-coded fixed effects of Rate pattern (A: Slow-Norm vs. Slow-Fast + Norm-Fast; B: Slow-Fast vs. Norm-Fast). A significant Group effect was yielded ($\beta=0.56$, SE $\beta=0.24$, $\chi^2(1)=4.7028$, $p=0.030$) along with a significant Group x Rate A interaction ($\beta=-0.64$, SE $\beta=0.3$, $\chi^2(1)=4.3052$, $p=0.038$). Musicians demonstrated significantly higher discrimination accuracy for the Slow-Norm rate pattern relative to the other rate patterns; whereas, non-musicians did not show differences in discrimination as a function of rate pattern.

Condition 3 utilized LMER models similar to Condition 2, with Group and Helmert contrast-coded fixed effects of Rate pattern (A: Fast-Norm vs. Slow-Fast + Slow-Norm; B: Slow-Fast vs. Slow-Norm) and their interactions. Only one significant effect emerged (between Slow-Fast and Slow-Norm rate patterns; $\chi^2(1)=11.55$, $p<0.05$). All other effects and interactions did not reach significance ($\chi^2(1)<2.84$, $p>0.05$).

Table 3: Mean d' scores (standard error in parentheses) for each condition (1=Between-Category, Same Rate, 2=Between-Category, Different Rates, 3=Within-Category, Different Rates) by Group. Condition 2 also indicates the specific lengths used at which rate (S=short, L=long).

Condition	NM	M
1-Fast	0.4 (0.1)	1.0 (0.2)
1-Norm	0.9 (0.2)	2.1 (0.2)
1-Slow	1.0 (0.2)	2.5 (0.2)
2-Norm(S)-Fast(L)	1.5 (0.2)	1.9 (0.1)
2-Slow(S)-Norm(L)	1.3 (0.2)	2.2 (0.2)
2-Slow(S)-Fast(L)	1.5 (0.2)	1.8 (0.2)
3-Slow-Norm	1.3 (0.1)	1.0 (0.1)
3-Norm-Fast	1.1 (0.1)	1.2 (0.1)
3-Slow-Fast	1.5 (0.1)	1.4 (0.1)

To compare overall performance by group and condition, an LMER model was constructed containing a contrast-coded fixed effect of Group and Helmert contrast-coded fixed effects for Condition (A: Condition 1 + 2 vs. Condition 3; B: Condition 1 vs. 2). Random effects for participant and random slopes for Condition by participant were also included. Significant effects for Group and each Condition comparison were found ($\chi^2(1)>6.1$, $p<0.003$). Additionally, both interactions were significant ($\chi^2(1)>6.4$, $p<0.01$). Follow-up subset models performed on each group revealed that musicians had significantly lower d' scores in Condition 3 relative to Conditions 1 and 2 ($\chi^2(1)=10.518$, $p=0.001$) and no difference between Conditions 1 and 2. However, performance in Condition 1 for non-musicians was significantly worse than Conditions 2 and 3 ($\chi^2(1)>8.8$, $p<0.03$).

4. DISCUSSION

The findings of the present study are consistent with our initial prediction that musicians' experience with rhythmic and temporal distinctions in music would transfer to facilitate their ability to normalize for speaking rate when perceiving non-native vowel length distinctions. Both groups were affected by speaking rate, such that they displayed poorer identification accuracy at faster relative to slower speaking rates. However, musicians were found to be significantly more accurate than non-musicians at identifying whether a vowel was short or long at each rate of speech. Similarly, in the discrimination task, musicians were significantly more accurate at between-category discriminations, either at the same rate or at different rates, as compared to within-category discriminations. Non-musicians, on the other hand, appeared to be responding "different" when pairs were different rates rather than responding to differences in length categories, as evidenced by higher d' scores in Conditions 2 (between-category) and lower scores in 3 (within-category), which involved changes in speaking rate, relative to Condition 1, which contained between-category pairs with the same speaking rate.

Musicians demonstrated a rapid formation of non-native length categories that were relatively robust enough to withstand considerable speaking rate variability. They were capable of tracking the speech rate of a given carrier sentence and accounting for that rate, while abstracting over considerable acoustic variation, when considering the vowel length of the target item. These results highlight the enhancement of domain-general auditory abilities from musical experience, namely that experience with extracting sound units and tracking regularities within a complex auditory environment can enhance the ability to acquire regularities in a speech environment [6]. This current work extends prior research on musicianship and linguistic processing by situating the target L2 contrast in a more ecologically-valid context, namely in sentential contexts at multiple speaking rates. These findings reveal that the beneficial effects of musical training on speech perception that have been well-attested in the literature (e.g., [2-5]) remain even in larger speech contexts.

5. REFERENCES

[1] Asaridou, S., McQueen, J. 2013. Speech and music shape the listening brain: evidence for shared domain-general mechanisms. *Front. Psych.* 4, 321.

[2] Baayen, R., Davidson, D., Bates, D. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390-412.

[3] Chobert, J., François, C., Velay, J., Besson, M. 2012. Twelve Months of Active Musical Training in 8- to 10-Year-Old Children Enhances the Preattentive Processing of Syllabic Duration and Voice Onset Time. *Cerebal Cortex,* 24, 956-967.

[4] Hirata, Y. 2004. Effects of speaking rate on the vowel length distinction in Japanese *J. Phon*. 32, 565-589.

[5] Hirata, Y., Whitehurst, E., Cullings, E. 2007. Training native English speakers to identify Japanese vowel length contrast with sentences at varied speaking rates. *J. Acoust. Soc. Am*. 121, 3837-3845.

[6] Kraus, N., Chandrasekaran, B. 2010. Music training for the development of auditory skills. *Nat. Rev. Neurosci.* 11, 599-605.

[7] Marie, C., Delogu, F., Lampis, G., Olivetti Belardinelli, M., Besson, M. 2011. Influence of Musical Expertise on Segmental and Tonal Processing in Mandarin Chinese. *J. Cogn. Neurosci.* 23, 2701-2715.

[8] McAllister, R., Flege, J., Piske T. 2002. The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *J. Phon.* 30, 229-258.

[9] Sadakata, M., Sekiyama, K. 2011. Enhanced perception of various linguistic features by musicians: a cross-linguistic study. *Acta Psychologica*, 138, 1-10.