ISCA Archive
http://www.isca-speech.org/archive

4th International Symposium on
Tonal Aspects of Languages (TAL-2014)
Nijmegen, The Netherlands
May 13-16 2014

Within-Category Variance and Lexical Tone Discrimination in Native and Non-Native Speakers

Christian W.G. Hoffmann [1], *Makiko Sadakata*[1], *Ao Chen*[4], *Peter Desain*[1] &

James M. McQueen[1,2,3]

[1] Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition, Radboud University Nijmegen, Nijmegen, Netherlands
[2] Behavioural Science Institute, Radboud University Nijmegen, Nijmegen, Netherlands
[3] Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands
[4]Utrecht Institute of Linguistics OTS, Utrecht University, Utrecht, The Netherlands

c.hoffmann@donders.ru.nl, m.sadakata@donders.ru.nl, a.chen@uu.nl, p.desain@donders.ru.nl, j.mcqueen@pwo.ru.nl

Abstract

In this paper, we show how acoustic variance within lexical tones in disyllabic Mandarin Chinese pseudowords affects discrimination abilities in both native and non-native speakers of Mandarin Chinese. Within-category acoustic variance did not hinder native speakers in discriminating between lexical tones, whereas it precludes Dutch native speakers from reaching native level performance. Furthermore, the influence of acoustic variance was not uniform but asymmetric, dependent on the presentation order of the lexical tones to be discriminated. An exploratory analysis using an active adaptive oddball paradigm was used to quantify the extent of the perceptual asymmetry. We discuss two possible mechanisms underlying this asymmetry and propose possible paradigms to investigate these mechanisms.

Index Terms: tone, L2 perception, Mandarin, Dutch, perceptual asymmetry

1. Introduction

According to the multi-store model of categorical perception [1], different memory processes are involved in categorical decisions: sensory, short-term categorical, and long-term categorical. Linguistic experience shapes long-term categorical memory and thereby allows for bottom-up matching and top-down expectations. A host of prior studies has shown that tonal categories are difficult to acquire for non-tonal language speakers, mostly because of the differences in the linguistic function of pitch between tonal and non-tonal languages (lexical vs. prosodic). Conversely, the processes involved in pitch perception have been shown to differ between tone- and non-tone language speakers, both in precision [2] and in relative weighting of acoustic cues used for identification of pitch patterns [3]. That is, not only do tone- and non-tone-language speakers differ in their long-term categorical memory, but also in the specificity of their short-term categorical memory. Prolonged training with novel pitch stimuli, however, can lead to re-tuning of pitch perception towards a more native pattern [4].

The acquisition of novel auditory categories is enhanced if the learner is exposed to a certain amount of within-category variation, such as different phonological contexts and/or multiple speakers [5]. However, the exact amount of variation needed for optimal learning differs across individuals [6]. What remains unclear is to what extent speakers of a non-tone language are already able to deal with within-category variance before any training (i.e. using their short-term categorical memory). Here, we focus on investigating the influence of within-speaker variance on the perception of Mandarin lexical tone categories by Dutch native speakers.

Previous studies have focused on perceptual learning of monosyllabic materials [2, 7, 8]. However, the true challenges in learning lexical tone may lie in perception of multisyllabic materials: identification and discrimination accuracy depend on the tonal context and position within a given word. For example, English native speakers struggle to identify Mandarin Tone 4 in non-final position, frequently confusing it with Tone 1, most likely due to interference from the English intonation system (Broselow, Hurtig & Ringen 1987 quoted in [8]). The present study therefore investigates effect of variance on perception of multi-syllabic words.

Experiment 1 compares perception accuracy of disyllabic materials (a Tone 1-Tone 4 continuum) by native speakers of Mandarin and Dutch. We predict both identification and discrimination results to show clearer signs of categorical perception in native speakers of Mandarin than in native speakers of Dutch. Experiments 2 and 3 further test the influence of within-category acoustic variance on the perception of the Tone 1 – Tone 4 continuum for Dutch and Mandarin native speakers. We predict that native speakers of Mandarin should be able to accommodate within-category variance in discrimination judgments, whereas even a modest amount of variance should impair the judgments made by Dutch speakers.

2. Experiment 1

2.1. Participants

10 native speakers of Mandarin Chinese (8 female, 2 male, mean age = 24) and 8 native speakers of Dutch (6 female, 2 male, mean age = 25) participated in the experiment.

2.2. Stimuli

One female native speaker of Mandarin Chinese produced utterances of the disyllabic pseudoword “asa” with Tone 1 (high level) or Tone 4 (high falling) on the first syllable. The tonal target of the second syllable was always Tone 3 (i.e. there were 1-3 and 4-3 sequences). Each recorded 1-3 utterance was paired with a 4-3 utterance, equalized in duration and number of pitch points and linear point-wise 9-step continua were constructed using the PSOLA method in PRAAT [9].

2.3. Procedure

2.3.1. Tests for Mandarin participants

For the identification task, the Mandarin speakers were presented with a pseudo-randomized sequence of stimuli from one of the constructed continua. Each individual stimulus was repeated 10 times within the sequence without being presented twice in a row, leading to a total of 90 trials. For each trial, participants had to indicate the lexical tone of the first syllable (i.e. Tone 1 or Tone 4) via button-press.

In the AX discrimination task, each trial consisted of stimulus pairs spanning two steps in the continuum, for a total of 7 pairs (1-3, 2-4 … 6-8, 7-9). The presentation order of the individual pairs was counter-balanced across the experiment. Each pair was presented 20 times. Filler pairs using the same stimulus (e.g. 1-1) were added, resulted in 174 trials in total (140 trials + 34 fillers). Participants had to indicate whether they thought the two stimuli sounded the same or different via button-press.

2.3.2. Tests for Dutch participants

As the Dutch native speakers were unfamiliar with the lexical tone categories, we opted to use an AXB labeling task instead of an identification task. For each trial, stimuli A and B were representatives of the endpoints of the Tone 1 – Tone 4 continuum (counter-balanced for order) with stimulus X being a random stimulus from the 9-step continuum. Each AXB triplet was presented 10 times, leading to a total of 90 triplets. Participants had to indicate via button-press whether stimulus X sounded more like stimulus A or stimulus B.

The AX task was similar to the AX task of the Mandarin participants but with 113 trials (7 pairs * 2 orders * 6 repetitions + 29 filler items).

	Group	Mean	Std.-Dev.
Slope	Mandarin	1.88	1.12
	Dutch	0.91	1.06
Threshold	Mandarin	4.49	0.83
	Dutch	3.57	0.89

Table 1: Descriptive Statistics for Experiment 1

2.4. Analysis

We fitted a cumulative normal curve through each participant's identification function, extracting the 50% threshold and corresponding slope using the PROBIT analysis implemented in SPSS [10]. Both slopes and thresholds were used for statistical analysis using one-way ANOVAs.

2.5. Results

Means and standard deviations for both slope and threshold can be found in Table 1. The difference in steepness of the slopes of the identification functions was marginally significant: $F(1,16)=3.49$, $p=.08$, Cohen's d=0.89. The location of the threshold also differed between groups: $F(1,16)=5.17$, $p<.05$, Cohen's d=1.07. The overall identification data with fitted identification functions can be found in Figure 1.

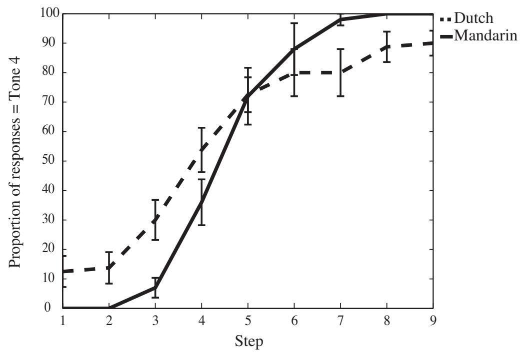

Figure 1: *Identification functions in Experiment 1*

Concerning the discrimination task, the repeated-measures 2 (Groups) x 7 (Stimulus Pairs) ANOVA revealed a main effect of Stimulus Pair: $F(2.889,46.217)=15.29$, $p<.001$ and a significant interaction of Group and Stimulus Pair: $F(2.889,46.217)=19.95$, $p<.001$. Analyzing the effects within the language groups, we find that the effect of Pair only becomes significant for the Mandarin group: $F(2.191,19.721)=36.314$, $p<.001$ (all Greenhouse-Geisser corrected), not for the Dutch group: $F(6,42)=1.40$, $p=0.24$. Figure 2 reveals the effect more closely. Whereas the Mandarin group shows a clear discrimination peak around the pair 3-5, the discrimination function of the Dutch participants remains relatively flat.

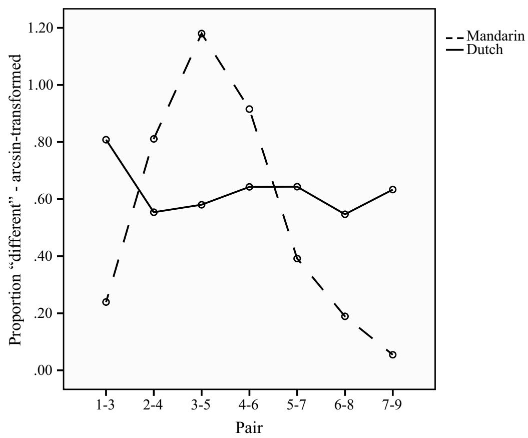

Figure 2: *Discrimination functions in Experiment 1*

2.6. Discussion

Our findings are in line with previous literature and our predictions. Dutch native speakers can perceive the acoustic differences that are underlying lexical tone discrimination, indicated by the steepness of the identification function derived from the AXB task. However, they do not perceive pitch differences categorically; their discrimination function remains mostly flat. Mandarin native speakers on the other hand show categorical perception for the stimuli used in this experiment.

3. Experiment 2

Experiment 2 was designed to probe the influence of within-category variance on lexical tone discrimination for both Mandarin and Dutch native speakers. If the perception of lexical tone is indeed categorical for Mandarin native speakers, then their discrimination abilities should not be influenced in any way by within-category variance.

Conversely, if the discrimination between lexical tone categories takes place on a purely acoustic level for Dutch native speakers, then within-category variance should have a strong effect on their discrimination scores.

3.1. Participants

Another 6 native Dutch speakers (3 male, 3 female, mean age = 24.2) and 3 native Mandarin (2 male, 1 female, mean age = 29) speakers participated in Experiment 2.

3.2. Stimuli

This experiment used three "asa" sequences with Tone1 – Tone 4 continua on the first syllable, and Tone 3 on the second syllable. The continua were constructed in the way described in section 2.2.

We used steps 1 through 3 and 7 through 9 of each continuum in this experiment because Experiment 1 indicated that native speakers of Mandarin perceived them as good exemplars of the two categories. We artificially increased our number of category exemplars by changing the average pitch of the initial syllable in 8 Hz steps from -32 to + 32 Hz. We created three levels of within-category variance: the first one contained a single stimulus from each category (closest to the category center, calculated by the arithmetic mean of the dimensions "pitch slope" and "average pitch height" for the originally recorded stimuli), the second one contained the 5 stimuli closest to each category center, the third one the 33 stimuli closest to each category center. A schematic overview of the average pitch height and slope of the stimuli used can be found in Figure 3.

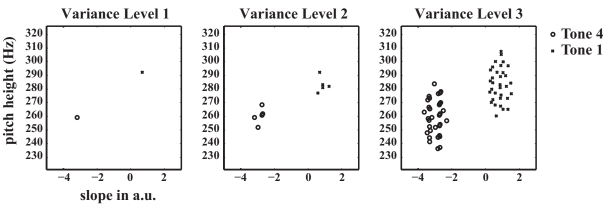

Figure 3: *Pitch properties of the stimuli used in Experiment 2*

3.3. Procedure:

We used an active oddball task to investigate the category discrimination accuracy. This type of task allowed us to investigate event-related potentials related to lexical tone discrimination in naïve subjects (EEG responses were collected but not reported here). Each oddball run consisted of 480 stimuli (duration: 600 ms, SOA: 1000 ms) and had a deviant probability of 15% (72 deviants). Deviants were either Tone 1 – Tone 3 (13) sequences ("D1") with 43 standards, or 43 sequences ("D4") with 13 standards, randomly sampled from the three levels of variance. A total of 12 blocks (3 levels of variance x 2 deviants x 2 repetitions) was presented to each participant. For 7.5% of the trials (36 trials per run), participants had to indicate whether they thought the last two stimuli they heard were examples of the same category or not. Question trials were counter-balanced to occur with a probability of 0.5 after two standards, and 0.5 after either a standard-deviant or deviant-standard pair.

3.4. Analysis

We extracted the proportion of correct responses for each individual block and participant. The data was then arcsin-transformed. As an initial analysis yielded no significant differences between the first and second presentation of any given block, we collapsed the data across individual blocks. We then entered it into a 2 (Groups) x 2 (Presentation Orders) x 3 (Levels of Variance) repeated-measures ANOVA.

3.5. Results

An overview of the proportion of correct responses for each level of variance and each group can be found in Figure 4. Each bar represents an individual block. Blocks with a Tone 4 deviant are marked "D4", blocks with a Tone 1 deviant are marked "D1." The repeated-measures ANOVA revealed a main effect of Variance, $F(1.68,11.7)=39.79$, $p<.001$, $\eta_p^2=0.85$, a main effect of Presentation Order, $F(1,7)=18.34$, $p< .01$, partial $\eta_p^2=0.72$, and a main effect of Group, $F(1,7)= 12.46$, $p<.01$, $\eta_p^2=0.64$. Furthermore, two interactions were significant: Order x Group, $F(1,7)=19.83$, $p<.01$, $\eta_p^2=0.74$ and Variance x Group, $F(1.68,11.7)=15.26$, $p<.001$, $\eta_p^2=0.68$. Neither the interactions Order x Variance nor Order x Variance x Group were significant.

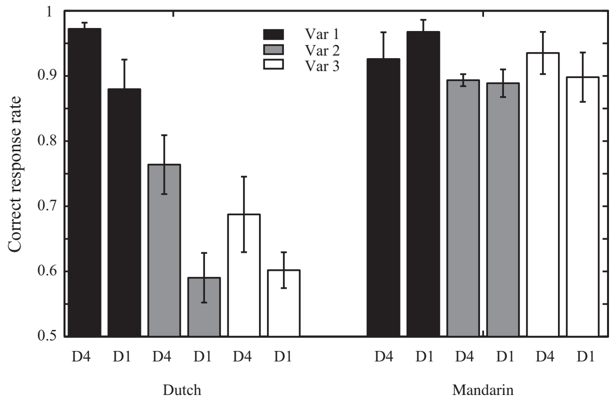

Figure 4: *Correct response rate in Experiment 2. D4: Tone 4 deviant. D1: Tone 1 deviant.*

3.6 Discussion

Within-category acoustic variance does indeed affect discrimination scores of non-native speakers. The Dutch native speakers are able to identify the acoustic difference between Tone 1 and Tone 4 for isolated tokens. However, the introduction of acoustic variance hinders them in their discrimination efforts. This is most likely due to the fact that the Dutch speakers cannot yet distinguish between within- and across category variance. On the other hand, the discrimination accuracy of Mandarin native speakers is not at all influenced by the addition of within-category variance, again showing the categorical nature of their perception of lexical tones.

An interesting finding was the main effect of Presentation Order, which is exclusive to the Dutch participants as shown by the Order x Group interaction effect. Dutch speakers were better able to cope with within-category variance if the standard category of the oddball run was the level tone (Tone 1) than if it was the falling tone (Tone 4).

4. Experiment 3

Experiment 3 was designed to further probe the order effect found in Experiment 2. We used an active adaptive oddball paradigm based on a staircase procedure to quantify the amount of within-category lexical tone variance that naïve Dutch speakers are able to ignore during auditory discrimination.

4.1. Participants

8 native Dutch speakers participated in this part (4 male, 4 female, mean age = 29.75).

4.2. Stimuli

We used the same stimuli as in Experiment 2.

4.3. Procedure

We constructed a total of 8 levels of variance in the same manner as described in Experiment 2, using (1,3,5,8,13,21,33, 54) stimuli per level and category. Instead of presenting 12 blocks of oddball runs with a fixed level of variance, we presented 4 blocks of adaptive oddball runs (alternating which category served as the standard between blocks). Each run consisted of 360 stimuli with a deviant probability of 15% (54 deviants). The corresponding number of questions was 27 (7.5%). The actual task remained the same as in Experiment 2. We used a 3-up, 2-down procedure to adjust the level of variance within the oddball run online. Three correct answers meant increasing the variance by one level (e.g. from 8 to 13 possible stimuli), whereas 2 incorrect answers led to a decrease in variance by one level.

4.4. Analysis

As the proportion of correct responses was necessarily similar across participants, due to the nature of an adaptive staircase task, we used the average level of the last two reversals as the dependent measure. If there were no reversals present, we used the highest level reached as the score for that particular run. The data for each participant and block was entered into a 2(Presentation Order) x 2(Block) repeated-measures ANOVA.

4.5. Results

The repeated measures ANOVA revealed a marginal effect of presentation order: $F(1,7)=4.27$, $p=.078$, $\eta_p^2 = 0.379$. Neither the main effect of Block nor the interaction between block and order reached significance. On average, participants were able to cope with 8 different stimuli (variance level 4) for the blocks containing Tone 1 as the deviant, and with 21 different stimuli (variance level 6) if Tone 4 was the deviant. The mean progress through the individual levels of variance can be seen in Figure 5. It seems that an asymptote is not being reached yet.

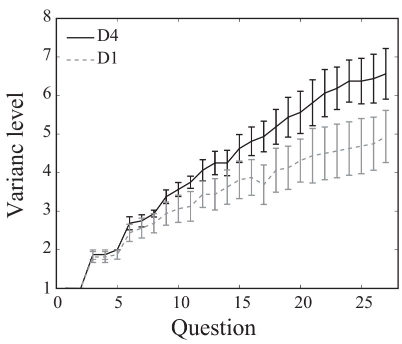

Figure 5: *Average progression through the different levels of variance in Experiment 3*

4.6. Discussion

The results of Experiment 3 provided further evidence for the perceptual asymmetry in Dutch native speakers between level and falling pitch contours. As in Experiment 2, participants showed lower discrimination scores if Tone 1 was the deviant within a train of Tone 4 standards than in the reverse case. In fact, none of the participants showed a reversal of this effect. The marginal effect might speak for the transient nature of the asymmetry, that is, some compensatory effects might already be at work within the 4 oddball runs. However, it might also be due to individual differences in the ability to perceive pitch.

5. General Discussion

Dutch native speakers seem to acoustically differentiate between level (Tone 1) and falling (Tone 4) contours in a disyllabic context, whereas Mandarin native speakers seem to do so categorically. This leads to the observed differences in the slope of the identification contours and relative variance around the category boundary, as well as the absence of a clear peak in the discrimination task for native Dutch speakers.

As expected, acoustic variance overlying level and falling contours affects behavioral performance of non-native speakers more than native speakers, due to the acquired lexical tone categories of the native speakers. Interestingly, acoustic variance influences discrimination scores asymmetrically in non-native speakers. The identification of a level contour within a context of falling contours is influenced more by acoustic variance than in the reverse case. This is similar to perceptual asymmetries found in for example in vowel discrimination, both in adults and infants [11].

This effect might be due to interference from the native Dutch intonation system: both ends of the continuum can be interpreted as a Dutch word with a pitch accent on the first syllable, i.e. "H*L L%", yet the pitch contour of a T4T3 combination might be less acceptable than that of a T1T3 combination. Less prototypical stimuli can lead to a bigger pop-out effect within prototypical stimuli, as shown in [11]. In the converse case, the non-prototypical standard might get perceptually pushed towards the prototypical representation, making the appearance of a prototypical deviant less salient and thereby leading to the observed asymmetry.

Another possible explanation of the observed asymmetry might be that the falling pitch contour of T4 as a deviant within a stream of level contours could lead to the impression of a prosodic boundary, essentially grouping all the T1 standards into one declination unit. In the converse case, the T1 deviant by itself does not serve as a prosodic boundary, leading to less of a pop-out effect.

We have shown the existence of such a perceptual asymmetry in discrimination in only a single phonological context. Further research is needed to investigate the stability and reproducibility of the effect. Should this be a stable effect, the next logical step would be to investigate the influence of perceptual learning on such an asymmetry. Our prediction would be that the acquisition of stable tonal categories should serve to diminish the observed effect. Furthermore, it would be interesting to investigate individual differences in this asymmetry in a manner similar to the authors of [6], who have shown that an individual's initial sensitivity to pitch contours greatly predicts learning success in monosyllabic items.

6. Conclusions

Dutch native speakers do not judge lexical tone stimuli in a categorical way. However, we have found evidence that Dutch speakers are not completely unprepared when confronted with lexical tone categories. In general, asymmetries as the one outlined above can help shed more light on the relative influence of one's native language on the perceptual learning of novel categories.

7. Acknowledgements

The present research was funded by the Technology Foundation STW and the Donders Centre for Cognition. Furthermore we would like to thank Anne Cutler, Carlos Gussenhoven and all other conference attendees for their valuable remarks on this paper.

8. References

[1] Xu, Y., Gandour, J.T., and Francis, A.L., "Effects of language experience and stimulus complexity on the categorical perception of pitch direction", J. Acoust. Soc. Am., 120(2):1063-1074, 2006.

[2] Leather, J. H. "Recognition of Chinese Word Tone from F0, with and Without Amplitude and Speaker Information." European Conference on Speech Technology, 1987.

[3] Gandour, J. T. and Harshman, R. A., "Crosslanguage differences in tone perception: A multidimensional scaling investigation," Language and speech, 21(1):1–33, 1978.

[4] Chandrasekaran, B., Sampath, P. D. and Wong, P. C. M., "Individual variability in cue-weighting and lexical tone learning," J. Acoust. Soc. Am., 128(1):456-465, 2010.

[5] Sadakata, M. and McQueen, J. M., "High stimulus variability in nonnative speech learning supports formation of abstract categories: Evidence from Japanese geminates," J. Acoust. Soc. Am., 134(2):1324-1335, 2013.

[6] Perrachione, T. K., Lee, J., Ha, L. Y. Y., and Wong, P. C. M., "Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design," J. Acoust. Soc. Am., 130(1):461-472, 2011.

[7] Wang, Y., Spence, M. M., Jongman, A., and Sereno, J. A., "Training American listeners to perceive Mandarin tones," J. Acoust. Soc. Am., 106(6):3649–3658, 1999.

[8] Wang, Y., Sereno, J. A., and Jongman, A., "L2 acquisition and processing of Mandarin tone," The Handbook of East Asian Psycholinguistics, 1:250–257, 2006

[9] Boersma, P. and Weenink, D. (2014). "Praat: doing phonetics by computer" [Computer program]. Version 5.3.64, from http://www.praat.org/, accessed on 12 Feb 2014

[10] Best, C. T. and Strange, W., "Effects of phonological and phonetic factors on cross-language perception of approximants," Journal of Phonetics, 20(3): 305–330, 1992.

[11] Polka, L. and Bohn, O.-S., "Asymmetries in vowel perception," Speech Communication, 41(1):221–231, 2003.