Preliminary Analysis of Japanese Pitch Accent through Perception of Native Speakers of Tonal Language*

○ T. Pongkittiphan, N. Minematsu, K. Hirose（The University of Tokyo）

1 Introduction

In second language learning, it is not a simple task for foreign students to master and be fluent in one target language within a few years. Linguists believe the differences and similarities between learners' first language (L1) and the languages that are being learned (L2) are important factors affecting foreign language learning [1].

When encountering an unknown foreign language, the beginning learners firstly tend to map what they perceived into the perspective of their mother language; for example, trying to map what they heard into the sounds in L1, or trying to transcribe it using the written script of L1. However, the number of kinds of letters or phonemes is different among languages. If L1 has a small set of phonemes and L2 has a larger set, L2-based representation of L1 sounds will inevitably results in information loss.

Japanese is known to have a comparatively smaller set of phonemes. Two English words, "bun" and "van", are written in the same Japanese word "バン", pronounced as /ban/. The sound /v/ has to be replaced by a sound /b/ because Japanese does not have phonemic variation of /b/ and /v/. Not only phonemic system but Japanese prosody also has its unique intonation. In Japanese phonology, a value of high or low is assigned to each mora and this assignment depends on features related to word, meaning, sentence, dialect, etc. [2]. To be able to speak Japanese fluently, it is necessary for learners to practice pronouncing Japanese words with correct pitch accents.

Although researchers have recently developed Online Japanese Accent Dictionary (OJAD) and the automatic prediction of Japanese pitch accent [3], the beginning learners do not know how to pronounce it correctly using their L1's prosodic pattern.

In this study, we are interested in how the beginning learners, whose L1 is a tonal language, perceive Japanese pitch accent in the view of their L1. In the other words, how high/low pitch levels in Japanese can be expressed by tones in a tonal language. This preliminary analysis focuses on the native Thai speakers whose language has five tones.

We also investigate the possibility to construct an automatic Japanese-pitch-to-Thai-tone transducer that can convert Japanese sentences annotated with pitch accent into its corresponding sequence of words written in Thai. This system can be used as a pronunciation guideline for Thai beginning learners to study about Japanese pitch accent and how to pronounce it correctly at the beginning level.

2 Accent characteristics of Japanese and Thai language

2.1 Japanese pitch accent

Japanese has a binary pitch level; high (H) or low (L), which is assigned to the smallest unit of speech production of Japanese, called "mora".

Type-n word accent is the accent type showing a rapid downfall of fundamental frequency (F_0) immediately after the n-th mora. Fig. 1 shows the four accent types of 3-mora words of Tokyo dialect and their accent nuclei indicated by filled black circles. In the case of type-0 accent, it means that there is no accent nucleus.

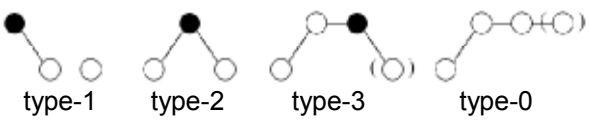

Fig. 1 Accent types observable in 3-mora words of Tokyo dialect of Japanese

* "声調言語を母語とする話者による日本語高低アクセントの知覚に関する予備的分析"，
ポンキッティパン　ティーラポン、峯松信明、広瀬啓吉　（東京大学）

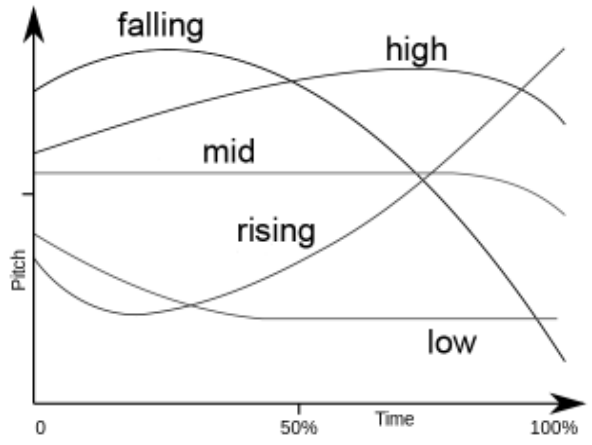

Fig. 2 Thai language tones chart

2.2 Thai tonal system

Thai is one of the tonal languages, which has five lexical tones; mid, low, falling, high and rising (henceforth, T0, T1, T2, T3 and T4 respectively) [4, 5]. Each syllable, which is the smallest unit in Thai, always has one of the five tones. These tones are characterized by unique patterns of F_0 which are illustrated in Fig. 2. Thai words, which have the same consonant and vowel but have different tones, mostly have different meanings e.g. the word "ma" with T0, T3 and T4 means "come", "horse" and "dog", respectively.

3 Transcription of Japanese accents to Thai tones

3.1 Design and tasks

Basically, the Japanese writing system does not include any visual symbols for pitch accent. Recently, Minematsu et al. [6] developed an accent-labeled Japanese text corpus. However, this corpus is constructed by observing only the text information. It does not provide any speech utterances which are the only accessible materials for the labelers who cannot even read Japanese.

To end this, one labeler was asked to listen to 503 Japanese utterances of the ATR database read by a male speaker of MMI, and then annotated the information of 1) accentual phrase boundary and 2) location of the accent nucleus, observed in each utterance. In the transcriptions, each sentence is divided into accentual phrases classified into one of the n accent types. In this study, we selected only 50 utterances of the A-set containing 239 phrases and 1,430 moras in total.

Next, six native Thai speakers, who are living in Thailand and have no knowledge of Japanese, were asked to listen to each utterance of the selected 50 Japanese utterances and to wrote down what they heard using Thai phonemic script. They had to pay attention especially to what kind of tones they perceived, and were allowed to repeat listening to the utterance until they were sure of their answers.

The transcriptions of two example phrases are shown as follow;

Jap-text : あらゆる / 現実を
Jap-accent : a ra yu* ru / geN ji tsu o
Thai-trans : a1 ra3 i3 ru1 / ken0 ci1 sw1 o1

where Jap-text, Jap-accent and Thai-trans are Japanese original text, its accent-labeled text and its annotation in Thai, respectively. Let note that "*" in Japanese phoneme indicates the position of accented mora, and the number located at the end of each Thai syllable indicates the tone type from T0 to T4. Therefore, high (H) and low (L) Japanese pitch levels are now defined based on the tonal language's perspective.

3.2 Tone occupancy

Table 1 shows the occupancy of each tone found in all 8,580 moras annotated by the six Thai subjects, mentioned in section 3.1. The interesting finding is that rising tone (T4) rarely exists in the reading utterances, in which its occupancy is less than 1%. Actually, T4 can be noticed in Japanese interrogative sentences where a rising pattern of pitch is found at the end of the sentences. Moreover, the first two majorities, which occupy about 75%, are low (T1) and mid tone (T0), respectively. The pitch patterns of T1 and T0 are more flat than those of the remaining falling (T2) and rising tone (T4), which have the rapid pitch change from high to low and vice versa. Considering these facts, the movement of Japanese pitch is quite constant or slightly changing in the perception of Thai listeners.

Table 1 Occupancy of each tone [%]

T0 mid	T1 Low	T2 falling	T3 high	T4 rising
27.18	48.74	11.35	12.42	0.28

Table 2 The reliability of agreement

	S1	S2	S3	S4	S5
S2	0.369				
S3	0.347	0.453			
S4	0.381	0.325	0.331		
S5	0.314	0.412	0.297	0.449	
S6	0.427	0.360	0.307	0.270	0.354

3.3 Inter-labeler agreement

We checked the inter-labeler agreement among six labelers. Using Cohen's kappa coefficient, Table 2 shows the inter-labeler agreement between each pair of labelers. From the result, the coefficients are ranging from 0.27 to 0.45, implying that the obtained annotations have very weak inter-labeler agreement. In the other words, the same Japanese utterance is annotated into a sequence of Thai tones by each labeler's perception which is poorly correlated to others' perception. Although we should have instructed labelers to control their labeling strategy, in this paper, we tentatively use the current labeling results to construct a Japanese-pitch-to-Thai-tone transducer.

4 An automatic Japanese-pitch-to-Thai -tone transducer

4.1 Goal and definition

Regarding Japanese-to-Thai transcriptions, described in section 3, in this study, we focus on how Japanese pitch should be mapped into Thai tone. So we ignore phonemic information in the Thai transcriptions and focuses only on tones.

The goal is to construct a language transducer that receives sequences of Japanese mora with pitch accents as input, and predicts their corresponding sequences of Thai tones.

4.2 Preparation of feature vectors

Regarding a binary Japanese pitch level, the information of the pitch level of the previous or the following mora might influence the pitch level of the current mora. To end this, we used Conditional Random Field (CRF), which is a powerful classifier that considers the neighboring samples, while most of the ordinary classifiers often ignore the neighboring information. The related work predicting of accent nucleus position [6] also used CRF, which is found to be effective.

Table 3 The features prepared for CRF

Features
▪ mora itself (mora ID) ▪ pitch level of the mora (H or L) ▪ pitch accent type of the phrase (Type-n) ▪ forward position of the mora in the phrase ▪ backward position of the mora in the phrase ▪ long vowel or not

Table 4 Precisions, recalls, and F1-scores [%]

	S1	S2	S3	S4	S5	S6	All
P	61.02	59.97	51.36	69.34	65.28	63.16	64.32
R	31.10	36.14	34.13	57.61	47.93	45.91	44.77
F1	41.20	45.10	41.01	62.93	55.28	53.17	52.79

Although inter-labeler agreement was not high, we used all of the labels as they were. At first, to select features that can optimize the predictor, we used all of 8,580 moras as a training dataset and 10-fold cross validation for the evaluation. After parameter tuning, the optimal features are shown in Table 3. We found the last feature, which indicates whether the mora is long vowel, is one of the effective features. Preliminary experiments showed that if we delete this feature, the F1-score dropped by 3.65 %. We carefully observed and found that the Japanese long vowel is annotated as one long syllable in Thai. As a result, the tones of the two consecutive moras of a long vowel are labeled as the same one.

Next, we constructed six different subject-dependent predictors using only each individual transcription containing 1430 moras in order to investigate whether the weak inter-labeler agreement, described in section 3.3, really affects the prediction performance or not.

4.3 Experimental results

Table 4 shows the results of precisions, recalls, and F1-scores of 10 cross-validation experiments.

1) From the results of six subject-dependent predictors, their F1-scores vary from 40% to 62%, implying that there are some transcriptions of labelers that seem to be difficult for CRF-based prediction and some that can be predicted slightly well.

2) Using all of 8,580 moras from six labelers, the predictor can get 52.79% of F1-score. This moderate performance is not good enough to construct a practical prediction. Regarding the

weak inter-rater agreement, it might not be the most critical factor affecting the predictor performance. Even constructing each predictor individually, its F1-score does not change much from 52.79%. However, for the preliminary study, it is quite acceptable because its performance is probably suppressed by the poor transcripts in which the performance of their corresponding subject-dependent predictors is worse.

The possible improvements are 1) to collect a larger number of transcriptions from all of the existing 503 Japanese utterances spoken by the same MMI speaker and 2) selecting native Thai subjects who have great efficiency in Japanese because their perception would be more reliable and can be used as one correct reference done by advance learners.

However, we do not claim that the transcriptions of the beginning learners are useless, but they can be used in different purposes e.g. to analysis the difference between the perception of beginning learners and that of advance learners, which would be helpful to study how learner's perspective of L2 is shifted from L1's world to L2's one.

5 Conclusions

The analysis of how Japanese pitch accent is perceived and expressed in the view of Thai tone is investigated. Six native Thai speakers listened to the same 50 Japanese utterances of ATR database and identified what kind of Thai tone they perceived. Considering the occupancy of five Thai tones in the transcriptions, about 75% of annotated tones are mid (T0) and low tone (T1), in which their F_0 patterns are considerably flat. This can be firstly implied that the movement of Japanese pitch is quite constant or slightly changing in the perception of Thai listeners.

For the preliminary experiment, Conditional Random Fields (CRF) is used to construct a Japanese-pitch-to-Thai-tone transducer, which can be used by the Thai beginning learners as a guideline showing Japanese pronunciation written in Thai, which is more understandable to Thai learners.

In future works, we are planning to annotate all of the existing 503 Japanese utterances to implement more practical transducer. Then, the resynthezied samples of Japanese utterance based on predicted Thai tones will be presented to native Japanese listeners to evaluate whether the predicted tones are sound natural and intelligible enough. This backward evaluation method can roughly simulate the situation when Japanese people are trying listen to Japanese utterances spoken by native Thais. The future study would be one of the important keys to understand how the perceptual adaption and prosodic transfer among tonal and non-tonal language.

References

[1] Murial Saville-Troike., "Introducing second language acquisition," Cambridge University Press, 2006.

[2] Timothy J. Vance., "The sounds of Japanese," Cambridge University Press, 2008.

[3] N. Minematsu et al., "日本語韻律教育の支援を目的としたオンラインアクセント辞書と読み上げチューターの開発″，電子情報通信学会音声研究会資料，SP2012-115, pp. 1-6, 2013-2.

[4] C. Hansakunbuntheund et al., "Thai tagged speech corpus for speech synthesis," Proc. O-COCOSDA, pp. 97-104, 2003.

[5] O. Krityakien, K. Hirose, N. Minematsu, "F0 contour generation of Thai speech using the tone nucleus model," Proc. NCSP, 2013-3.

[6] N. Minematsu, R. Kuroiwa, and K. Hirose, "CRF-based statistical learning of Japanese accent sandhi for developing Japanese text-to-speech synthesis systems," Proc. ISCA Workshop on Speech Synthesis, pp. 148–153, 2007.