An Analysis of Lombard Effect on Thai Lexical Tones: The Role of Communicative Aspect

Chariya Boontham[1], Chutamanee Onsuwan[1,2], Tanawan Saimai[3], Charturong Tantibundhit[2,3]
[1]Department of English and Linguistics, Faculty of Liberal Arts, Thammasat University, Thailand
[2]Center of Excellence in Intelligent Informatics, Speech and Language Technology and Service Innovation (CILS), Thammasat University, Thailand
[3]Department of Electrical and Computer Engineering, Faculty of Engineering, Thammasat University, Thailand
babblenoise@gmail.com, consuwan@tu.ac.th, sync_nero@hotmail.com, tchartur@engr.tu.ac.th

Abstract

This work investigates Lombard Effect (clean versus babble) on the realization of Thai lexical tones across two settings: with conversation partner (map task) and without partner (sentence reading). For both, thirty location names whose final syllable varying in lexical tones were constructed and used. Ten pairs of Thai adults participated in the study. The findings showed that in noise condition, regardless of whether the tones were produced with or without conversation partner, F0 values for all tones were significantly higher. Importantly, Lombard Effect on the tones was significantly increased for all but low tone when a partner was involved.

Index Terms: Lombard Effect, Thai, babble noise, lexical tone, map task

1. Background

In 1911, Lombard conducted a study and made an important discovery that during conversation with noise (of various types), speakers adjust their use of language and speech patterns [1]. From then on, the concept of Lombard Effect/Lombard speech was originated, which was the phenomenon where a speaker increases loudness and modifies her speech in noise condition to allow a listener to understand the intended message [2]. Lombard Effect has received extensive attention from speech scientists and relevant studies have been carried out in many languages (e.g., English [3], Spanish [4], Cantonese [5]).

By and large, it was clear that apart from an increase in intensity, speakers make other significant speech adjustments in noise condition (as opposed to clean condition). Specifically, it was found that speakers made adjustment in their production of vowels in terms of vowel duration and intensity [6]. Fundamental frequency and formant frequency were also higher when communication was changed from clean to noise condition [6]. Importantly, these effects were reported across languages (e.g., Spanish [4], French [7]).

However, several studies concluded that Lombard Effect did not take place as a consequence of noises only. They asserted that thoughts, mental states, and reactions of conversation partners also had influences on speech modifications [8]. Lane and Tranel [9] reported that in identical noise conditions, speakers with larger number of conversation partners made greater degrees of speech adjustment_than those in ones-sided communication. In addition, Fitzpatrick, Kim and Davis [10] concluded that during communication in noise condition with conversation partners, vowel duration and F0 were significantly higher than without partner. Those previous findings highlighted an important aspect of communication in Lombard speech where there is at least a 2-way communication and the sender aims to make the receiver understand what she wants to say. Thus, the sender is willing and ready to adapt her speaking strategies for the effectiveness of communication [1]. Moreover, Zeng and Liu [11] stated that individual's speaking patterns/forms usually change abruptly when noises occur during communication or when there are errors in speaking and listening from conversation partners.

As mentioned, many acoustic characteristics were reported to associate with Lombard speech. Recently, Lombard Effects on lexical tones have been explored in tonal languages, such as Cantonese and Thai. Zhao and Jurafsky [5] compared speech adjustments of 6 lexical tones in clean versus white noise conditions (with no conversation partner) from 8 Cantonese speakers aged between 20-52 years old. The findings showed that in white noise condition, F0 in all tones increased, especially in mid and mid-rising tones.

Kasisopa et al. [12] examined speech production in noise, focusing on lexical tones in Thai. Speech data was obtained from six females aged between 27-34 years old in clean and white noise conditions. It was found that F0 and tone contours for the five tones in isolated words were higher than those in sentence frame. Importantly, it was found that F0 for four tones (mid, low, high, and rising tones) in white noise condition was higher than in clean condition.

Lombard Effect not only reflects speech adjustment but also reveals aspects of human speech perception process. A number of studies have investigated perception patterns of Thai lexical tones in noise conditions (e.g., white noise in Onsuwan et al. [13] and pink noise in Mixdorff et al. [14]). To our knowledge, only the study by Kasisopa et al. [12] has addressed the question of Lombard Effect in the production of Thai lexical tones. Therefore, we would like to explore this further by 1) using various types of noise: clean, white, and babble (results in babble noise are only presented here) 2) including two communicative settings (with and without conversation partner). To give readers a background of Thai lexical tones, F0 contours of the five tones (mid, low, falling, high, and rising) from this present study in clean condition where there was no conversation partner are given in Figure 1.

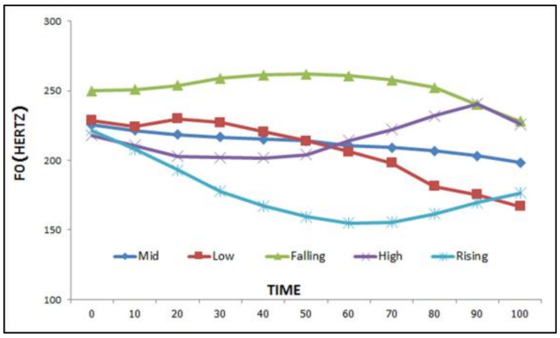

Figure 1: *F0 of five Thai lexical tones (mid, low, falling, high, and rising) from this study in clean condition with no conversation partner.*

2. Method

2.1. Participants

Participants consisted of 10 conversation pairs (one member is a 'speaker' and the other 'listener') aged between 20-40 years old; all pairs had at least one female member (male-female or female-female). Members in each pair know each other well and only one female member was assigned the speaker role.

2.2. Speech materials

Ten target syllables were /pa:/ /pà:/ /pâ:/ /pá:/ /pǎ:/ /na:/ /nà:/ /nâ:/ /ná:/ /nǎ:/. Thirty location names were constructed, each with a target syllable in a final (stressed) syllable. For example, /khrua.khun.**ná:**/ 'Auntie's kitchen', /hâ:ŋ.tà.wan.**ná:**/ 'Tawanna shopping mall'. A same list of target location names was used in map task and sentence reading.

2.3. Data recording

Data were collected from two separate settings: with conversation partner (map task) and without partner (sentence reading). Each recording session, which includes map task and sentence reading, always started with map task and lasted about 20 to 30 minutes.

2.3.1. Map task (with conversation partner)

Following a map task designed by Viethen, Dale and Cox,F [15], three sets of maps were developed; each set consisted of two corresponding maps, one for the speaker and the other the listener as shown in Figures 2 and 3. Three different sets of maps (3 scenes: city scene, beach scene and country scene) were created so that three different noise conditions (clean, white noise, and babble noise) could be randomly be distributed across 10 participant pairs. Before the task begun, each participant was assigned a role, either as a speaker or a listener. They were informed that their task was for the speaker to guide the listener from a starting point (only indicated on the speaker's map) to a finishing point (only indicated on the speaker's map) via a specified route (only indicated on the speaker's map) for each map set. Each map set has 18 assigned locations (10 of which were target location names and 8 were distractors). They were also told that one of them would be hearing some noise from headphones. One short practice trial with a simplified set of maps was given to each pair.

The members were in two separate rooms and could only communicate through microphones and headphones. Only one member (the speaker) heard the noise from headphones. Three types of noise condition were introduced: clean, white noise, and (Thai) multi-talker babble noise (for this paper, only results from babble noise were reported) at 60-75 dBSPL. Recordings of speech from both members were made, but only those from the speakers were analyzed. It should be noted that for each map set, when multiple repetitions of target location names were elicited, but only the best two for each were selected.

2.3.2. Sentence reading (without conversation partner)

The speakers were asked to read at a normal speed 72 sentences (with 30 target location names and 24 distractors) for three times in each noise condition (randomly assigned). The sentence frame was /tɕʰǎn.paj … ʔ a:tʰ ít.ní:/ 'I go to…this week.'

Figure 2: *Sample of listener's map (city scene).*

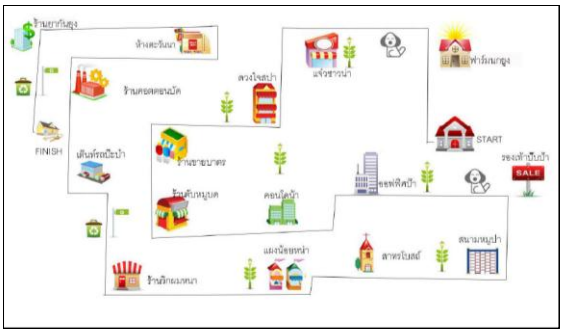

Figure 3: *Sample of speaker's map (city scene).*

2.4. Data analysis

Target syllables were manually segmented and two parameters were measured, i.e., segmental duration (not reported here) and F0 using Praat [16]. Specifically, the F0 values of tone contour of each syllable were extracted in ten equidistant (time-normalized points) from 0% to 100% as shown in Fig. 4. In total, speech tokens analyzed here are composed of 400 tokens from map task (10 syllables × 2 repetitions × 2 noise conditions × 10 speakers) and 600 tokens from sentence reading (10 syllables × 3 repetitions × 2 noise conditions × 10 speakers).

Then, the data were separately analyzed in five repeated three-way ANOVA, i.e., one for each of the five Thai tones. Three independent variables in each of these analyses are Condition (Clean/Lombard), Communication (Map/Sentence), and Time points (10% points), respectively.

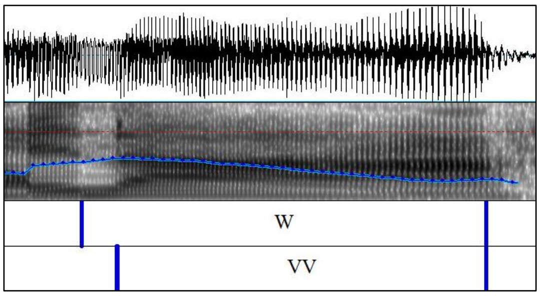

Figure 4: *Waveform and spectrographic display (with F0 line) of a syllable /nà:/ produced in clean condition in sentence reading.*

3. Results

Table 1 shows significant F-values for all factors and their interactions for each of the five tones on fundamental frequency (F0) by comparing the clean and noise condition and with (map) and without (sentence) conversation partners. Findings are explained below.

Clean vs. Lombard: The results revealed that in the babble noise condition, F0 for all of the five tones were significantly higher than those in the clean condition, both in map and in sentence settings (see Table 1). This can clearly be seen when comparing between the two solid curves (map Lombard vs map clean) and between the two dashed curves (sentence Lombard vs sentence clean) in Figures 5-9.

Map vs. Sentence: For map task (both in clean and noise), F0 for all but falling tones were significantly higher than those in sentence setting. The Clean/Lombard × Map/Sentence interactions were significantly different for four tones, i.e., mid, falling, high, and rising, showing that Lombard Effect was stronger in map than in sentence settings. (see Table 1). As for F0 shapes, when comparing the two solid curves (map) with the two dashed curves (sentence), the dashed ones appear to be flatter and lower, but overall shapes remain relatively the same for mid, falling, and high tones (Figures 5, 7 and 8). A notable pattern can be seen in rising tone where the second half of the F0 contours showed a gradual fall rather than a rise in sentence setting.

Trend Analyses: The trend analyses showed that there were significant trends over time for all but high tone. Interestingly, trends interacted with Map/Sentence for mid, falling, and rising tones.

Table 1: *Significant F-values for all factors and their interactions for each of the 5 tones.*

	Mid	Low	Falling	High	Rising
C vs L	**6.94**	**9.02**	***34.58***	***52.20***	***47.89***
M vs S	***61.46***	***18.02***	N/S	***30.11***	***57.43***
TP	*7.74*	***23.07***	***11.54***	N/S	***23.73***
C/L x M/S	***14.03***	N/S	*6.39*	**10.55**	***11.89***
C/L x TP	N/S	N/S	N/S	N/S	N/S
M/S x TP	**2.42**	N/S	***3.62***	N/S	***20.09***
C/L x M/S x TP	N/S	N/S	N/S	N/S	N/S

Note: C stands for speech in clean condition; L stands for Lombard speech; M stands for map task; S stands for sentence reading; TP stands for time points; and N/S stands for not significant. *Italic figure* = *p*<.05; **Bold figure** = *p*<.01; and ***Bold Italic figure*** = *p*<.001.

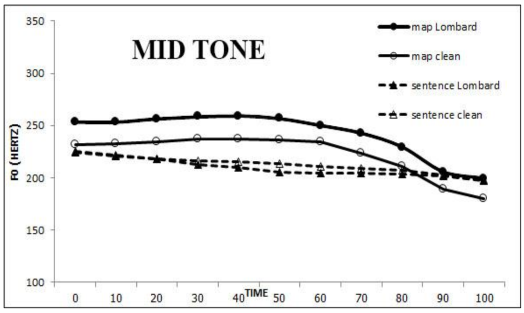

Figure 5: *F0 curves of mid tone across 10 normalized time points in clean vs. babble noise (Lombard) and with conversational partner (map) vs. with no partner (sentence).*

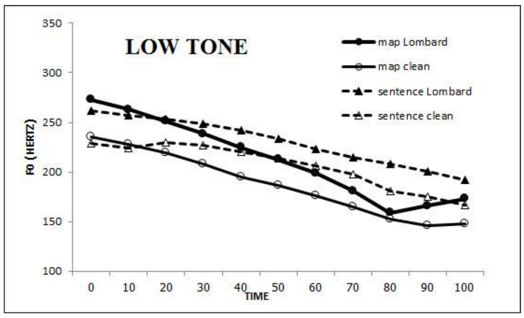

Figure 6: *F0 curves of low tone across 10 normalized time points in clean vs. babble noise (Lombard)and with conversational partner (map) vs. with no partner*

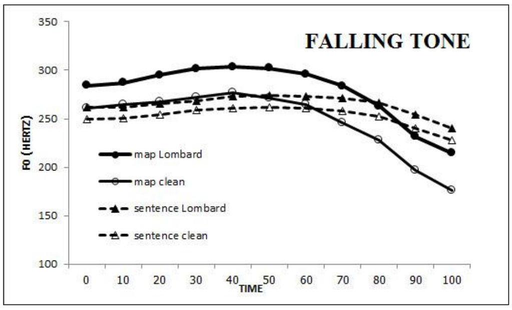

Figure 7: *F0 curves of falling tone across 10 normalized time points in clean vs. babble noise (Lombard) and with conversational partner (map) vs. with no partner (sentence).*

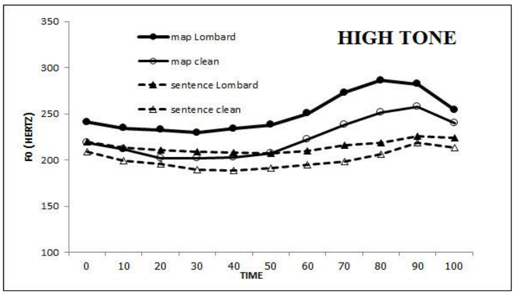

Figure 8: *F0 curves of high tone across 10 normalized time points in clean vs. babble noise (Lombard) and with conversational partner (map) vs. with no partner (sentence).*

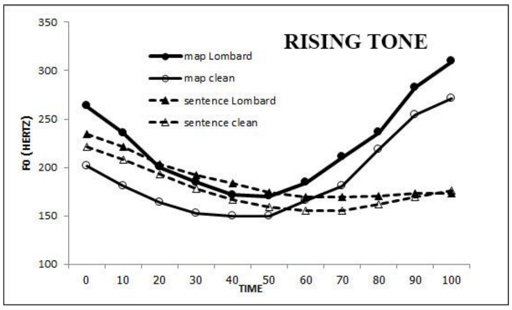

Figure 9: *F0 curves of rising tone across 10 normalized time points in clean vs. babble noise (Lombard) and with conversational partner (map) vs. with no partner (sentence).*

4. Discussion and Future Work

In general, our findings are in line with previous studies on Lombard speech [1-7]. Specifically, in babble noise condition, regardless of whether the tones were produced with or without conversation partner, F0 values for all tones were higher than in clean condition. Interestingly, Lombard Effect on Thai lexical tones was significantly increased (higher F0 for mid, falling, high, and rising tones) when conversation partner was present. On this latter point, it was consistent with what has been reported [9-11]. Thus, apart from noise condition, having a conversation partner can be another important factor for speakers to significantly adjust their speech. This is possibly because more of speaker's effort would require to obtain reactions from conversation partners when it comes to real and active communication situations.

From the current set of speech data, detailed analyses are being conducted on lexical tones in white noise condition and Lombard Effect on segmental durations with emphasis on adjustment of vowel duration among contrastively short and long vowels. Analysis of other acoustic correlates such as intensity and spectral tilt could also be valuable.

Another important contribution of this study is the design of map task (in Thai) and the procedure involved in the task. We believe that our map task could be beneficial for relevant studies in which speech eliciting in a natural conversation setting is required.

5. References

[1] Lau, P. "The Lombard effect as a communicative phenomenon," UC Berkley Phonology Lab Annual Report, 1-9, 2008.

[2] Anglade, Y. and Junqua, J-C., "Acoustic-phonetic study of-Lombard speech in the case of isolated-words," STL Research Reports, 129-135, 1990.

[3] Hazan, V. and Baker, R., "Acoustics-Phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions," *J. Acoust. Soc. Am*., 2139-2152, 2011.

[4] Castellanos, A., Benedi, M. and Casacuberta, F., "An analysis of general acoustic-phonetic Features for Spanish speech produced with the Lombard effect," Speech Commun, 20:23–35, 1996.

[5] Zhao, Y. and Jurafsky, D., "The effect of lexical frequency and Lombard reflex on tone hyperarticulation," Journal of Phonetics, 37:231-247, 2009.

[6] Summers, W., Pisoni, D., Bernacki, Pedlow, R., and Stokes, M.,"Effects of noise on speech production: acoustic and perceptual analyses," *J. Acoust. Soc. Am.,* 84: 917-928, 1988.

[7] Patel, R. and Schell, K. W., "The influence of linguistic content on the Lombard effect," *J.Speech Lang. Hear.*, 51:209-220, 2008.

[8] Junqua, J-C., Fincke, S., and Field, K., "Influence of the Speaking Style and the Noise Spectral Tilt on the Lombard Reflex and automatic speech recognition," *In International Conference on Spoken Language Processin*g, 467-470, 1998.

[9] Lane, H., and Tranel, B., "The Lombard sign and the role of hearing in speech," W J. Speech Hear., 14:677-709, 1971.

[10] Fitzpatrick, M., Kim, K. and Devis, C., "The effect of seeing the interlocutor on auditory and visual speech production in noise," *In Proc. of International Conference on Auditory-Visual Speech Processing*, 31-35, 2011.

[11] Zeng, F.G. and Liu, S., "Speech Perception in individuals with auditory neuropathy," Journal of Speech, Language, and Hearing Research, 367-380, 2006.

[12] Kasisopa, B., Attina, V. and Burnham, D., "The Lombard effect with Thai lexical tones: an acoustic analysis of articulatory modifications in noise," Proceedings of Interspeech 2014, 15th Annual Conference of the International Speech Communication Association, Singapore, September 14-18, 2014.

[13] Onsuwan, C., Tantibundit, C., Saimai, T., Saimai, N., Chootrakool, P., and Thatphitthakkul, S., "Analysis of Thai tonal identification in noise," Proceeding of the 14th Australasian International Conference on Speech Science and Technology (SST). Sydney, Australia: Macquarie University, 2012.

[14] Mixdorff, H., Charnvivit, P., and Burnham, D. K., "Auditory–visual perception of syllabic tones in Thai," In E. Vatikiotis-Bateson, D. Burnham, & S. Fels (Eds.), Proceedings of the Auditory–Visual Speech Processing International Conference. Adelaide, Canada: Causal Productions. 3-8, 2005.

[15] Viethen, D. and Cox, F., "Designing a new map task (manuscript)," Sydney: Western Sydney University, 2010.

[16] Boersma, P. and Weenink, D., "Praat: Doing phonetics by computer [Computer Program] Version 5.3.57," retrieved 1 November 2013 from http://www.praat.org/

[17] Bapineedu, G., "Analysis of Lombard Effect Speech and its application in speaker verification for imposter detection," in Proc. of Interspeech 2013.