The following paper was presented at The 9th Workshop on Disfluency in Spontaneous Speech (DiSS 2019) held at ELTE Eötvös Loránd University in Budapest, Hungary on 12–13 September, 2019.

Title: Acoustic-phonetic characteristics of Thai filled pauses in monologues

Author(s): Thanaporn Anansiripinyo and Chutamanee Onsuwan

Abstract: Filled pause (FP) is one type of disfluent phenomena that is commonly found in everyday speech. It has been widely studied in many languages, but little is known about this topic in Thai. This work explored three important acoustic-phonetic characteristics of Thai filled pauses in monologues. To elicit target monosyllabic tokens of FPs and those of regular word (RW) counterparts, 31 Thai adult females were asked to watch two short cooking videos and describe the contents. They were also asked to read out loud target word lists. Three acoustic measures: syllable duration, first (F1) and second formant (F2) frequencies were taken from 738 tokens. Across vowel contexts, only F2, not F1, in FPs, was significantly different from that in RWs. Differences in syllable duration between RWs versus FPs were near significant. The findings suggest that Thai speakers produced FPs in a presumably different way from RWs. In FPs, the syllable was relatively lengthened and the tongue position was moved towards the center of vowel space. Future directions include a detailed analysis of FPs in terms of amplitude, fundamental frequency, pause duration before/after fillers and other non-linguistic factors.

DOI: <https://doi.org/10.21862/diss-09-014-anan-onsu>

Citation (JIPA): Anansiripinyo, Thanaporn & Chutamanee Onsuwan. 2019. Acoustic-phonetic characteristics of Thai filled pauses in monologues. In: R. L. Rose & R. Eklund (eds.), *Proceedings of DiSS 2019, The 9th Workshop on Disfluency in Spontaneous Speech*, 12–13 September, 2019, Budapest, Hungary, 51–54.

The complete proceedings for DiSS 2019 are available as follows.

ISBN: 978-963-489-063-8

DOI: <https://doi.org/10.21862/diss-09>

DiSS 2019 was sponsored by The Faculty of Humanities, ELTE Eötvös Loránd University and the International Speech Communication Association (ISCA).

Acoustic-phonetic characteristics of Thai filled pauses in monologues

Thanaporn Anansiripinyo^{1,2}, Chutamanee Onsuwan^{1,2}

¹Department of English and Linguistics, Faculty of Liberal Arts, Thammasat University, Bangkok, Thailand

²Center of Excellence in Intelligent Informatics, Speech and Language Technology, and Service Innovation, Thammasat University, Bangkok, Thailand

Abstract

Filled pause (FP) is one type of disfluent phenomena that is commonly found in everyday speech. It has been widely studied in many languages, but little is known about this topic in Thai. This work explored three important acoustic-phonetic characteristics of Thai filled pauses in monologues. To elicit target monosyllabic tokens of FPs and those of regular word (RW) counterparts, 31 Thai adult females were asked to watch two short cooking videos and describe the contents. They were also asked to read out loud target word lists. Three acoustic measures: syllable duration, first (F1) and second formant (F2) frequencies were taken from 613 tokens. Across vowel contexts, only F2, not F1, in FPs, was significantly different from that in RWs. Differences in syllable duration between RWs versus FPs were near significant. The findings suggest that Thai speakers produced FPs in a presumably different way from RWs. In: FPs, the syllable was relatively lengthened and the tongue position was moved towards the center of vowel space. Future directions include a detailed analysis of FPs in terms of amplitude, fundamental frequency, pause duration before/after fillers and other non-linguistic factors.

Introduction

It is a well-known fact that words in a language may have several meanings and some may serve more than one function. To illustrate, English words such as ‘like’ and ‘well’, which are content words/ regular words (RWs), are also used as filled pauses in some spoken contexts (e.g. casual speech). As FPs, their meanings differ greatly from the original ones. Since FPs can be observed quite often in spoken data, they could present some challenges and obstacles for speech technology that relies on Automatic Speech Recognition (ASR) system, such as automatic translator (Gabrea & O’Shaughnessy, 2000; Ogata et al., 2009; Medeiros et al., 2013). We believe that understanding underlying acoustic differences between FPs and regular words is a useful step for differentiating between them.

Even though FPs have been studied widely in many languages (e.g. O’Shaughnessy, 1992;

Watanabe et al., 2004; Vasilescu, Adda-Decker & Nemoto, 2007; Gósy et al., 2014 etc.), research on FPs in Thai is insufficiently studied. Most of the studies focused on FPs used by Thai learners of English (Pletikosa & Rungrojsuwan, 2018; Williams & Korko, 2019). Only two studies (Chaimanee, 1996; Panichkul, 2003) have looked at the use of FPs in the Thai language.

Chaimanee (1996) compared frequency of occurrences of FPs in Thai conversations carried out between native (Thai) and non-native speakers (Chinese, Japanese, American, French, Frisian, and German). FPs that were found in this study include common FPs (e.g. [(?)u:m], and [(?)ɔ̃:] and Thai lexical FPs (e.g. [kɔ̃:] ‘also’, and [tɛháj] ‘yes’). Some speakers combined Thai lexical FPs with common FPs, for example, [kʰàʔ.ʔà:] (Thai ‘final particle’ [kʰàʔ] and common FP [ʔà:]).

Almost a decade later, Panichkul (2003) investigated acoustic characteristics of FPs in Thai monologues focusing on their duration, amplitude, and fundamental frequency. Data were collected from 30 Thai speakers talking about a given topic; 1,162 FPs were analyzed. She found that average duration of FPs was significantly longer than that of surrounding words, while the amplitude and fundamental frequency of FPs were significantly lower. It is noteworthy that similar finding for the duration differences were found in languages such as Japanese, English, French, and Spanish (Quimbo, Kawahara & Doshita, 1998; Vasilescu, Adda-Decker & Nemoto, 2007).

However, Panichkul (2003) did not include any Thai lexical FPs in her analysis and did not analyze other important features such as formant frequencies. Vocalic hesitations have demonstrated interesting features and have been studied in many languages (e.g. Vasilescu, Adda-Decker & Nemoto, 2007). In French, it is likely to be /œ/, while in American English, it is between /ʌ/ and /æ/, and /e/ in Spanish. Comparisons of F1 and F2 between FPs and RWs also suggested that there is a small tendency for higher F1 and lower F2 in FPs in Portuguese. (Proença et al., 2013)

This work presents findings from three main characteristics of common and Thai lexical FPs which are duration and first and second formant frequencies. More importantly, our data were collected and

analyzed in such a way that we could make systematic comparisons between FPs and RWs. The study is part of a larger project investigating acoustic characteristics of FPs (duration, f_0 , F1, and F2) and other non-linguistic factors, such as speaking style, topic, and age group.

Method

Participants

Thirty-one Thai adult females (24–68 years old; mean = 42.42 ± 15.64 SD) were recruited to perform two experimental tasks: describing cooking methods and reading target word lists. Participants of the same gender were selected for convenience in the acoustic data comparison.

Speech materials

In this study, participants were asked to perform two tasks: describing cooking methods and reading target word lists. In the first task, participants were instructed to watch two short cooking videos (3–5 minutes each) with Thai audio descriptions. The two videos differ in terms of complexity level ('easy' and 'more complicated') of cooking steps and variety of ingredients. The first 'easy' video is for "Coriander smoothie" and the second 'more complicated' video is a recipe for "Chor Muang" (type of Thai sweet; violet flower-shaped steam dumpling). They all started with the 'easy' video. After finish watching each video (one time each), they were asked to repeat the cooking methods, names of ingredients, and other important information as much as they could provide.

During the second task, they were asked to read target and non-target words in a sentence frame. Target words consisted of 24 monosyllables, 14 of which are FPs (7 common FPs [(?)u:m, (?)ù:m, (?)û:m, (?)ɾ:, (?)ɿ:, (?)à:, (?)ò:] and 7 Thai lexical FPs [bæ:p, kô:, kʰàʔ, kʰráp, kʰū:, læ:w, mǔan] as shown in Table 1. Ten non-target words share similar vowels with the FPs. These items randomly appeared three times (in three different word lists) so that each participant read 72 tokens in total.

Speaking rate was not directly controlled in both tasks (analysis of speaking rate is conducted separately and not included here); participants were encouraged to talk comfortably at their normal speaking rate.

Token selection

It should be noted that FPs are produced spontaneously and quite uncontrollably in speech. For the purpose of this study, acoustic analysis and comparison were conducted only on matching monosyllables from each speaker that were produced both as FPs (in describing task) and in RWs (in

Table 1. Target monosyllables (common and Thai lexical FPs) in word lists.

[(?)u:m]	[bæ:p] 'form, pattern'
[(?)ù:m]	[kô:] 'also'
[(?)û:m]	[kʰàʔ] 'final particle'
[(?)ɾ:]	[kʰráp] 'final particle'
[(?)ɿ:]	[kʰū:] 'is'
[(?)à:]	[læ:w] 'and'
[(?)ò:]	[mǔan] 'identical'

reading task). As a result, only eight monosyllables satisfied the condition; 256 tokens of RWs and 357 tokens of FPs as shown in Table 2. The last column shows a number of speakers (out of 31 participants) who produced these items during both tasks.

Table 2. Number of tokens for RWs and FPs used in acoustic analysis.

Monosyllable	No. of tokens		No. of speaker
	Regular word	Filled pause	
(?)ɾ:	17	8	4
(?)ɿ:	44	120	18
(?)à:	41	91	15
(?)ò:	8	9	6
bæ:p	24	33	8
kô:	60	63	20
kʰàʔ	33	23	12
kʰū:	21	10	7
Total	256	357	

Mispronounced tokens in the reading task were excluded resulting in unequal numbers of repetitions. In terms of occurrences of FPs, the highest occurrence was for common FPs [(?)ɿ:], followed by [(?)à:], and Thai lexical FP [kô:] 'also'. (Frequency of occurrences of all FPs from the first task is analyzed but not included here).

Acoustic analysis: For syllable duration measurement, we measured (with Praat; Boersma & Weenink, 2017) the entire syllable unit by marking the starting point at the start of initial sound and the ending point at the last sound of vowel or final (if exists). For formant frequencies, F1 and F2 values were extracted from the mid-point of vocalic portions; the point where formant frequency is likely to be the most stable.

Statistical analysis: Syllable duration, F1 and F2 values of each token were analyzed with statistical test and average and Standard Deviation (SD) for each factor were calculated. N-Way Analysis of Variance (N-way ANOVA) was used to test the differences of three factors (syllable duration, F1, and F2) between the tokens occurred in RWs and FPs.

Acoustic-phonetic characteristics of Thai filled pauses

Results of each factor will be presented as follows.

Syllable duration

Figure 1 shows that in 7 out of 8 items, syllable duration of FPs were longer than that of RWs, except in [bà:p]. In Table 3, average duration across all FPs was longer (339.90 ± 159.12 SD) than that of RWs (296.99 ± 85.73 SD). In addition, duration range of FPs was wider (53.50–1197.45 ms) than RWs (166.55 – 645.55 ms). However, the duration differences were near significant [$F(1,166) = 3.32$, $p = 0.0704$] (the dfs reflect number of target monosyllables that satisfied the condition, See Table 2). Post-hoc analysis indicated the same trend.

Formant frequencies (F1 and F2)

In Table 3, slightly higher average F1 and higher average F2 were found across all monosyllables (across different vowel contexts) in RWs than in FPs. N-Way Analysis showed that across vowel contexts, significant difference of F2 in FPs versus RWs was found [$F(1,166) = 5.57$, $p = 0.0194$], but not of F1 [$F(1,166) = 0.09$, $p = 0.7705$].

Interesting pattern emerges as we compare each individual word/vowel in Figure 2. In FPs (triangle shapes) (as opposed to RWs), their F2 values were located closer to the center of the vowel space. In fact, for the front vowel in [bà:p], F2 in FP was lower while for the back vowels in [(?)ò:] and [kô:], F2 in FP was higher (see Discussion section below).

Post-hoc analysis revealed that F2 of [bà:p] in RW was significantly higher than in FP ($p < 0.01$). For F1, post-hoc analysis showed that significant differences were found in common FPs, [ʔə:] and [ʔà:]. F1 of [ʔə:] in FP was significantly higher than that in RW

Figure 1. Average syllable duration of 8 monosyllables produced as RWs (left bar) vs. FPs (right bar). The whiskers indicate SDs

Table 3. Means and SDs of syllable duration, F1, and F2 values from 256 tokens of RWs and 482 tokens of FPs.

		Syllable duration (ms)	F1 (Hz)	F2 (Hz)
RWs	Mean	296.99	715.32	1606.79
	SD	85.73	230.00	355.06
FPs	Mean	339.90	712.82	1514.88
	SD	159.12	169.90	243.90

($p < 0.01$) and F1 of [ʔà:] in FP was significantly lower than RW ($p < 0.05$).

Discussions and future directions

We are aware that our data set was limited as naturally produced FPs are not easy to control and that the RW counterparts were produced in a connected sentence frame rather than spontaneous speech. Nevertheless, the findings revealed an interesting trend that Thai speakers produced RWs and FPs in a relatively different way. Syllables in FPs were lengthened to

Figure 2. Average F1 and F2 values of 8 monosyllables produced in RWs and FPs. Formant frequencies were taken from the mid-point of vocalic portions

some extent and the tongue position was moved closer towards the center of vowel space. F2 dimension was the one that showed the most significant effect.

Syllable duration differences were near significant, but the results largely agreed with previous studies (Quimbo, Kawahara & Doshita, 1998; Panichkul, 2003; Vasilescu, Adda-Decker & Nemoto, 2007) that duration of FPs tends to be longer than RWs. In the case of [bæ:p] in our data, the only exception for the pattern, we speculated that in FPs (but not in RWs), the tokens were likely pronounced with a different vowel, a contrastive short [bæp] rather than the long [bæ:p]. To a native ear, changing to [bæp] did not change the word meaning, but made it sound less formal.

As for formant frequencies, with most studies focusing on a few vowels in FPs, their results seemed to show that FPs had higher F1 and lower F2 (Vasilescu, Adda-Decker & Nemoto, 2007; Proença et al., 2013). In our data, we examined a number of different vowels and clearer patterns appeared to emerge (see Figure 2). In varying degrees, each vowel was produced differently in RWs versus FPs, but in FPs, the tongue position apparently (not surprisingly) moved towards the center; e.g. the vowel in [bæ:p] was pronounced more like [ə] or [a] in FPs.

Future directions include a detailed analysis of FPs in terms of amplitude, fundamental frequency, pause duration before/after fillers and other non-linguistic factors (such as age and gender). Preliminary results of FP frequency of occurrences seem to show that level of complexity of the task/topic has a noticeable effect. As the task becomes more complicated and challenging, speakers are more likely to show various disfluency strategies such as silent pauses, lengthening and FPs.

Acknowledgment

The project was partially supported by Center of Excellence in Intelligent Informatics, Speech and Language Technology, and Service Innovation, Thammasat University.

References

- Boersma, P. & D. Weenink. 2017. Praat: Doing phonetics by computer (ver. 6.0.56) <http://www.praat.org/> (Accessed 22 July 2017).
- Chaimanee, N. 1996. Communicative pauses in Thai. In: *Pan-Asiatic linguistics: Proceedings of the 4th International Symposium on Language and Linguistics*, Mahidol University, 8–19 January 1996, Bangkok, Thailand, 174–182.
- Gabrea, M. & D. O'Shaughnessy. 2000. Detection of filled pauses in spontaneous conversational speech. In: B. Yuan, T. Huang, & X. Tang (eds.), *Proceedings of the International Congress of Phonetic Sciences*, 16–20 October 2000, Beijing, China, 678–681.
- Goto, M., K. Itou & S. Hayamizu. 1999. A real-time filled pause detection system for spontaneous speech recognition. In: G. Olaszy, G. Németh, & K. Erdőhegyi (eds.), *Proceedings of the 6th European Conference on Speech Communication and Technology*, 5–9 September 1999, Budapest, Hungary, 227–230.
- Gósy, M., J. Bóna, A. Beke & V. Horváth. 2014. Phonetic characteristics of filled pauses: The effects of speakers' age. In: S. Fuchs, M. Grice, A. Hermes, L. Lancia, & D. Mücke (eds.), *Proceedings of the 10th International Seminar on Speech Production (ISSP)*, 5–8 May 2014, Cologne, Germany, 150–153.
- Medeiros, H., H. Moniz, F. Batista, I. Trancoso & H. Meinedo. 2013. Experiments on automatic detection of filled pauses using prosodic features. In: F. Bimbot, C. Cerisara, C. Fougeron, G. Gravier, L. Lamel, F. Pellegrino, & P. Perrier (eds.), *Proceedings of the International Speech Communication Association*, 25–29 August 2013, Lyon, France, 2629–2633.
- Ogata, J., M. Goto & K. Itou. 2009. The use of acoustically detected filled and silent pauses in spontaneous speech recognition. In: *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 19–24 April 2009, Taipei, Taiwan, 4305–4308. <https://doi.org/10.1109/ICASSP.2009.4960581>
- O'Shaughnessy, D. 1992. Recognition of hesitations in spontaneous speech. In: *Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing*, 23–26 March 1992, San Francisco, USA, 521–524. <https://doi.org/10.1109/ICASSP.1992.225857>
- Panichkul, S. 2003. *An acoustic study of Thai pause fillers in relation to their syntactic positions in monologues*. Master's dissertation, Chulalongkorn University, Bangkok, Thailand.
- Pletikosa, J. & S. Rungrojsuwan. 2018. The use of filled pauses in monologue by Thai learners of English with different English language proficiency levels. *Vacana Journal of Language and Linguistics*, 6(1): 20–35.
- Proença, J., D. Celorico, A. Veiga, S. Candeias & F. Perdigão. 2013. Acoustical characterization of vocalic fillers in European Portuguese. In: R. Eklund (ed.), *Proceedings of DiSS 2013, the 6th Workshop on Disfluency in Spontaneous Speech and TMH-QPSR Volume 54(1)*, KTH Royal Institute of Technology, 21–23 August 2013, Stockholm, Sweden, 54(1): 63–66.
- Quimbo, F. C., T. Kawahara & S. Doshita. 1998. Prosodic analysis of fillers and self-repair in Japanese speech. In: R. H. Mannell & J. Robert-Ribes (eds.), *Proceedings of the International Speech Communication Association*, 30 November – 4 December 1998, Sydney, Australia, 3313–3316.
- Vasilescu, I., M. Adda-Decker & R. Nemoto. 2007. Acoustic and prosodic characteristics of vocalic hesitations across languages. *Scientific Report 2007*.
- Watanabe, M., Y. Den, K. Hirose & N. Minematsu. 2004. Clause types and filled pauses in Japanese spontaneous monologues. In: S. H. Kim, & D. H. Youn (eds.), *Proceedings of the International Speech Communication Association*, 4–8 October 2004, Jeju, Korea, 2981–2984.
- Williams, S. A. & M. Korko. 2019. Pause behavior within reformulations and the proficiency level of second language learners of English. *Applied Psycholinguistics*, 40: 723–742. <https://doi.org/10.1017/S0142716418000802>