

SPEECH RHYTHM OF THE DANISH-CHINESE INTERLANGUAGE RELIES ON RHYME STRUCTURE

Marjoleine Sloos^{1,3}, Jie Liang², Xuechao Ne³, Rasmus Puggaard Hansen^{3,4}, Mengzhu Yan³, Chun Zhang⁴

¹Aarhus University (Interacting Minds Centre), ²Tongji University (School of Foreign Languages),
³Aarhus University (Linguistics), ⁴Aarhus University (Chinese Studies)

ABSTRACT

Although speech rhythm is a highly important factor in the evaluation of nativeness, it is usually not explicitly taught in second language acquisition. Moreover, acquisition studies of Mandarin prosody typically focus on the acquisition of tone. We investigated durational patterns of Mandarin Chinese as produced by Danish learners. Native Mandarin Chinese rhythm is characterized by, first, sentence-final syllable lengthening, then tone-intrinsic duration, and finally, rhyme structure. Danish speech rhythm is fundamentally different, since Danish is a stress language with only primary stress on the level of the phonological phrase, and extensive reduction of unstressed syllables. We show that the interlanguage of Danish learners of Chinese is comparable to native Mandarin Chinese duration patterns with regard to rhyme structure. However, differentiation of duration based on tone and final lengthening lags behind in Danish acquisition of Chinese.

Index Terms— Second language acquisition, final lengthening, prosodic duration, Mandarin Chinese, Danish

1. SPEECH RHYTHM IN SECOND LANGUAGE ACQUISITION

Each language has its own speech rhythm, the recurring pattern of syllables with similar duration. This speech rhythm is usually not explicitly taught in second language acquisition [22]. However, rhythm seems to be the most important factor for native speakers in the judgement of nativeness [22]. Speech rhythm should thus have priority in the acquisition of a second language. The fact that speech rhythm is not taught in the classroom does not imply that it is not acquired, however. In a study on the acquisition of the rhythm of German among Chinese, Italian, and Polish learners, vowel reduction was to a significant degree acquired, indicating that the acquisition of rhythm is a low-

level, unconscious, process [9]. Which are the factors that determine speech rhythm? What has to be acquired exactly?

We investigate these questions for Danish learners of Mandarin Chinese. Since the factors of speech rhythm are to a large extent language-specific, let us first consider the rhythmic aspects of Chinese and Danish.

2. CHINESE AND DANISH SPEECH RHYTHM

Chinese and Danish differ drastically from each other on the suprasegmental level. The most pervasive difference is that Chinese is a tone language whereas Danish is a stress language. Chinese also has simpler syllable structure than Danish. But the two languages have a common feature: extensive reduction. In this section, we describe the durational aspects of tone, stress, a syllable's position in the sentence, as well as reduction in native Chinese and Danish.

2.1. Tone

Tones have an intrinsic duration. The common view is that, based on citation tones, T4 and T1 are shorter than T2, which is shorter than T3. Based on connected speech, however, the following (mean) durations are reported [21]:

T1 212ms, T2 201ms, T3 178ms, T4 195ms.

T3 is relatively short in medial position, but in sentence-final position, T3 is the longest tone [21]. Apparently, in connected speech, T3 is not fully pronounced. The duration of the neutral tone T0 is about 50%-60% of the lexical tones (see [15] and references cited there), which makes duration a primary cue for recognition of the neutral tone ([24]:49). Duration of tone interacts with the position in the sentence and rhyme structure, which we discuss in section 2.3 and 2.4 below. First, we turn to the role of duration as a function of stress.

2.2. Stress

In terms of duration, stressed syllables are longer than unstressed syllables [14]. Spectral reduction may occur in unstressed syllables such that vowels are realized as schwa, and schwa syllables may undergo deletion [12]. Although stress plays a subordinate role in Chinese, durational reduction occurs in unstressed syllables. Typically, these syllables have a neutral tone, like *zi* [tsɿ] in *beizi* [peitsɿ] ‘cup’ (see also 2.5). This may even lead to syllable contraction ([3,19] among others). Syllable contraction is the extensive reduction resulting in deletion of the rhyme of the unstressed syllable and resyllabification of the onset of the unstressed syllable as the coda of the preceding stressed syllable. For example, *beizi* can be reduced to a single syllable [peits] ‘cup’.

Syllable contraction (at least in Taiwanese Mandarin) appears mainly to be a result of speech rate [2] and lexical frequency [19]. The highly frequent function words and suffixes (often with a central vowel like in *ge* [gə], *men* [mən], *shi* [ʃɿ]) are most susceptible to reduction.

Unlike Chinese, Danish is a typical stress language, in which stress assignment has been claimed to be free to a very large extent with only one stressed syllable per word [6]. Unstressed syllables are subject to extensive reduction. Even stressed syllables can be reduced within a syntactic unit [6]. Reduction is a topic that attracted special attention in the literature on Danish (see e.g. [16] for an overview). Although reduction usually occurs more often in informal registers, also in Danish [10], this does not prevent extensive reduction in careful speech [11]. Dramatic vowel and consonant reduction—four or even five syllables can be contracted to a single one—have been described in research on the Danish Phonetically Annotated Spontaneous Speech (DanPASS) corpus [8,17].

2.3. Final lengthening

There is vast evidence that, cross-linguistically, sentence-final syllables are often subject to lengthening as initially reported by Klatt [13]. In Chinese, the final syllable of an intonational phrase appears to be 28 to 61% longer than sentence-medial syllables [4,15]. As such, it turns out to be the strongest predictor for duration, followed by tone [4]. The following values for identical syllables in medial and final position are provided by Feng [5]:

Table 1. Tone duration in sentence-medial and sentence-final position [5].

Position	Tone 1	Tone 2	Tone 3	Tone 4
medial	248	259	249	248
final	274	320	335	268

By contrast, in Danish, final lengthening does not or only optionally occur (according to the regiolect) [7].

2.4. Syllable structure

Rhyme structure has been shown to correlate with duration in Chinese. For instance, Tseng & Lee [18] distinguish nine clusters of rhymes based on duration. In their study, the shortest rhymes consisted of the central vowels [ɿ ʌ ə]. But the other categories were highly inconsistent regarding (i) the number of segments in the rhyme, (ii) whether the vowel was a monophthong (V) or a diphthong (VV), (iii) whether the rhyme had a coda nasal (N) or not, and (iv) whether the onset was followed by a glide (G) or not. E.g. one group of rhymes with comparable duration consisted of /ai ou uei əng ong ing/. In a similar analysis, Chu & Feng [4] arrived at a more homogeneous *(but still not fully consistent) clustering (see Table 2).

Table 2. The order of rhyme structures according to their duration [4].

Rhyme Structure Type	Rhymes
1	a o e i ʌ u y
2	ei ui ie iu ye uo er
3	ai ao ua ou ia uai iao
4	in un yn en ing
5	an ang eng ong uan yan ian uang ueng iong iang

Summarizing, the most important predictor for duration in Mandarin Chinese is final lengthening, followed by tone and rhyme structure. In Danish, final lengthening does not occur often. The effect of rhyme structure on duration is not exactly known. Syllable contraction occurs in both languages although the surface forms differ for the two languages. Given these differences, we hypothesize that Danish-accented Chinese has (almost) no final lengthening and that reduction may occur. Differentiation in tone duration may not be native-like.

3. MATERIAL

3.1. Speakers

We recorded 12 undergraduate students enrolled in the Chinese Studies program at Aarhus University in Denmark. All students were Danish native speakers. They had spent one semester in Beijing as part of their studies. At the time of the recordings, all participants were about to graduate.

3.2. Stimuli

The stimuli consisted of 89 simple sentences of 6 syllables. We deliberately included 66 neutral tones and also 83*T3 that were subject to sandhi (i.e. T3 surfaces as T2 before an adjacent T3).

We distinguish the following rhyme types (including glides G): central vowels, V, VV, VN, GV, GVV, GVN, Vr. The exact numbers of the distribution are given in Table 3. These distributions are not equally balanced since we could use only basic vocabulary and simple sentences such that the students could more or less fluently read the sentences, e.g.:

- (1) 我有十个苹果。
Wǒ yǒu shíge píngguǒ.
 'I have ten apples'

Table 3. The number of rhyme types represented in the material.

Type	Rhyme	Number
V[central]	ə ɿ ɪ	79
V	a e i u o	145
GV	ia ie iu ua ue	18
VV	ai ao ei ou uo ui	155
GVV	iao	6
VN	an an gen eng in ing ong un ün	120
Vr	er	1
GVN	ian iang	23

3.3. Procedure

The participants were asked to read the 89 sentences from slides of a Microsoft PowerPoint presentation. Each sentence was presented on a single slide in Chinese characters and pinyin (tones were indicated with standard diacritics e.g. mā, má, mǎ, mà, ma). The experiment was self-paced and took 15-20 minutes on average. The speech was recorded with a Tascam DR05 V2 recorder in the linguistics lab at Aarhus University.

4. RESULTS

To investigate the variance of syllable duration in Danish-accented Mandarin Chinese, we measured the duration of the rhyme of each syllable with Praat speech processing software and script [1,23]. We speaker-normalized the data by subtraction of the rhyme duration from the speaker mean. Subsequently, we conducted a linear regression test on these deviations of the mean and the predictors (i) position in the sentence, (ii) tone, and (iii) rhyme structure.

4.1. Final lengthening

As compared to syllables in medial position (Mean 180ms), sentence-final syllables were 12.3% longer (Mean 202ms, $t = 14.7$, $p < 0.001$). Initial syllables, by contrast, were 16.1% shorter (Mean 152ms, $t = -11.9$, $p < 0.001$). Final lengthening interacted with tone duration such that in final position, T1 was 32.5% longer, and T2 T3 T4 and T0 were respectively 17.3%, 32.0%, 21% and 0% longer.

4.2. Tone

We observed variance in tone duration, which interacted with the position in the sentence (as mentioned in 4.1). The means are provided in Figure 1. In medial position, the order from the shortest tone to the longest was as follows: T0 < T4 < T2 < T1 < T3. In sentence-initial and sentence-final position, this order was somewhat blurred. Most remarkably, T3 in initial position turned out as short as T4. We also observe neutralization between T0 and T4 in sentence-final position.

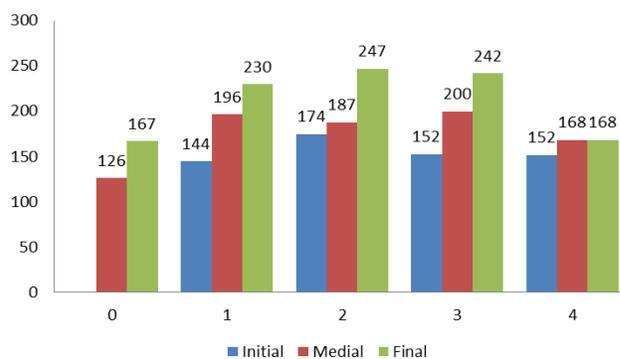


Figure 1. Mean rhyme durations divided by tone and by position in the sentence (in ms).

The duration of the neutral tone T0 was 67.1% of the mean duration of the lexical tones in sentence-medial position and

74.6% of the mean duration of the lexical tones in sentence-final position. The durational differences on the basis of tone were significant, as well as all interactions (Table 4). Table 4 also shows that the effect size (estimates) of the position in the sentence was larger than that of the lexical tones.

Table 4. The Estimates, t-value, and p-value of the rhyme duration as predicted by tone and position in the sentence. Significance on the .95 level is indicated by an asterisk. The reference levels are medial position and tone 2.

Factor	Est.	t-value	p-value
(Intercept)	-44.52	-15.06	< 0.001*
Sentence-final	38.67	6.56	< 0.001*
Sentence-initial	-20.33	-3.93	< 0.001*
Tone 0	-32.93	-9.05	< 0.001*
Tone 1	12.81	4.31	< 0.001*
Tone 3	12.90	4.43	< 0.001*
Tone 3 sandhi	-7.80	-1.98	0.047*
Tone 4	-7.48	-2.57	0.010*
Initial:Tone 1	-17.65	-2.74	0.006*
Final:Tone 3	16.09	2.10	0.035*
Initial:Tone4	15.69	2.05	0.041*
Final:Tone 4	-42.12	-5.43	< 0.001*

4.3. Rhyme type

Finally, rhyme structure was a strong factor for speech rhythm among the Danish L2 learners of Chinese. We provide the means of the duration of the rhyme types in medial position in Table 5.

Table 5. Mean rhyme durations divided by rhyme type (in ms, ordered from shortest to longest).

Rhyme Type	Rhyme	Duration
V[central]	ə ɿ ɪ	131.5
V	a e i u o y	157.7
GV	ia ie iu ua ue	168.9
VV	ai ao ei ou uo ui	189.0
GVV	iao	202.9
VN	an an gen eng in ing ong un yn	207.3
Vr	er	248.5
GVN	ian iang	257.9

In the same linear regression model as discussed above, rhyme type was significant for all levels. Table 6 lists the

effects of the different rhyme types in order of increasing effect size as compared to the reference level of the shortest rhyme (central vowel). In comparison with the factors position in the sentence and tone, rhyme structure has the largest effect size (as expressed in the estimates).

Table 6. The estimates, t-value, and p-value of the rhyme duration as predicted by rhyme structure. The reference level is central vowels.

Factor	Est.	t-value	p-value
(Intercept)	-44.5	-15.06	< 0.001*
V	20.7	8.26	< 0.001*
GV	32.0	6.93	< 0.001*
VV	55.5	22.20	< 0.001*
GVV	67.0	9.40	< 0.001*
VN	73.2	28.9	< 0.001*
Vr	100.7	6.0	< 0.001*
GVN	117.9	29.7	< 0.001*

5. DURATION IN THE DANISH-CHINESE INTERLANGUAGE

We investigated the rhythm of Danish-accented Chinese in relation to the position in the sentence, duration of tone, and rhyme structure. The results show that in the interlanguage, rhyme structure was the strongest predictor, ordered according to their duration as follows: V[central] < V < GV < VV < GVV < VN < Vr < GVN. In addition, sentence-final syllables were 12.3% longer and initial syllables were 16.1% shorter than syllables in medial position. Further, in medial position, the order from the shortest tone to the longest was T0 < T4 < T2 < T1 < T3. The duration of the neutral tone T0 was 67.1% of the mean duration of the lexical tones in medial position and 74.6% of the mean duration of the lexical tones in final position.

How do these results relate to the variance of syllable duration in Chinese? The most important predictor for syllable duration in native Mandarin Chinese is final lengthening, whereby the final syllable has a duration that is 28-61% longer than medial syllables [4,18]. In our Danish-accented Chinese data, final lengthening does occur, but depending on tone, only 0-32.5%. Although this suggests that final lengthening is acquired, it does not approach native-like production. Besides, initial syllables were significantly shorter than medial sentences, which is likely to be a task-related result: the sentences were read one-by-one and the first syllable was probably the best prepared.

If we turn to the effect of syllable structure in our data to that of native production, we find that duration follows roughly the same pattern as reported by [4] for native Mandarin. For convenience, we list the order of lengthening in native Chinese compared to the order of lengthening in our data in Table 7.

Table 7. The order of rhymes according to their structure in Danish-accented and native Mandarin Chinese.

Rhyme Type	Danish-accented	Native
V[central]	ə ʌ ɪ	ə ʌ ɪ a o i u y e
V	a o i u y e	
GV	ia ie iu ua ue	ie iu ye uo + ei ui (VV), + er (Vr)
VV	ai ao ei ou uo ui	ai ao ua ou + ia (GV) + uai iao (GVV)
GVV	iao	
VN	an an gen eng in ing ong un yn	in un yn en ing
Vr	er	
GVN	ian iang	uan yan ian uang ueng iong iang + an ang eng ong (VN)

An exception in the regularity of the rhyme durations of Danish and native speakers is the *-er* rhyme. Whereas it is relatively short in native production, it is lengthened in Danish-accented speech. We suspect the typically Germanic tendency to lengthen vowels before a rhotic [20] is responsible for this process.

Finally, we can relate variance in duration to tone identity. For instance, T0 is shorter than the lexical tones in Danish-accented Chinese. But whereas in native speech the duration of T0 is 50-60% of the lexical tones, T0 produced by Danish speakers is 66-77% of the lexical tone. This indicates that the Danish learners do acquire reduction of T0, but acquisition is not (yet) complete.

In order to compare the realization of tone duration of the Danish L2 learners to native realization, Figure 2 provides the duration of the tones in our data, along with the values of native speakers [21]. We added the value for T0 as 55% of the mean of the lexical tones for native production (see section 2.1). Figure 2 shows that T3 is the longest among the Danish L2 speakers but the shortest for native speakers. We assume that the L2 learners adhere to the

durational differences in citation forms that they learned rather than imitating native speakers in connected speech, who appear to reduce T3 in connected speech (see Table 1).

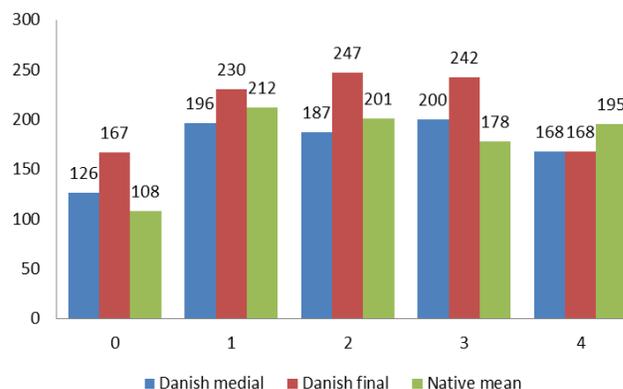


Figure 2. Tone-intrinsic durations in Danish-accented and native Mandarin Chinese in ms. Further, see text.

The most surprising finding is perhaps that the Danish speakers' rhyme durations are often shorter than the native speaker's ones in [21]. Also if the native speaker in [21] would have been using careful speech style, it is unexpected that second language learners have shorter durations. A possible explanation is that Danish is so much reduced, that relatively much phonological (and semantic!) information is realized very quickly in a given time frame (conform [11] in a comparison of speech rate and reduction between Danish and its very closely related sister languages Norwegian and Swedish). It would be interesting to compare different Danish-accented languages with the same languages spoken by native speakers as well as other non-native speakers to see whether Danish speakers in general tend to produce more reduced speech in their second languages than other speakers.

Summarizing, we suggest that Danish-accented Chinese is characterized by an extremely quick phonological realization. Syllable duration is mostly influenced by rhyme structure, approaching nativeness. However, we observed that in the interlanguage, final lengthening is not acquired—even though it is the most important factor in native speech. Tone duration, the second important factor for native speakers, had a non-native character.

7. ACKNOWLEDGEMENTS

We thank Jeroen van de Weijer for comments on a previous version of the paper and Jette Jørgensen and Barbara Fuchs for assistance during the recordings.

10. REFERENCES

- [1] Boersma, P., & Weenink, D., *Praat: Doing phonetics by computer* [computer program], 2010.
- [2] Cheng, C., & Xu, Y., "Extreme reductions: Contraction of disyllables into monosyllables in Taiwan Mandarin", *INTERSPEECH*, pp. 456-459, 2009.
- [3] Cheng, Y., "Syllable contraction in Mandarin A-not-A construction", 12th National Conference on Linguistics, 2001.
- [4] Chu, M., & Feng, Y., "Study on factors influencing durations of syllables in Mandarin", *Proc. Eurospeech, Scandinavia*, 2001.
- [5] Feng, L., "北京話語流中聲、韻、調的時長" (Duration of initials, finals and tones in Beijing dialect). In T. Lin, & L. J. Wang (Eds), *北京語音試驗錄 (Notes of Phonetic Experiments in Beijing Dialect)*, Beijing University Publishing, Beijing, pp. 131-195, 1985.
- [6] Grønnum, N., "Danish", *Journal of the International Phonetic Association*, 28(1-2), pp. 99-105, 1998a.
- [7] Grønnum, N., "Intonation in Danish", In D. Hirst & A. Di Cristo (Eds), *Intonation Patterns: A Survey of Twenty Languages*, CUP, Cambridge, pp. 131-151, 1998b.
- [8] Grønnum, N., "DanPASS-a Danish phonetically annotated spontaneous speech corpus", *Proceedings of the 5th LREC*, pp. 1578-1583, 2006.
- [9] Gut, U., "Non-native speech rhythm in German", *Proceedings of the ICPhS Conference*, pp. 2437-2440, 2003.
- [10] Heegård, J., & Thøgersen, J., "Style-dependent rules for reduction: Mixed-model multiple linear regression analysis of reduction phenomena in Danish news readings", *Nordic Journal of Linguistics*, 37(3), pp. 325-365, 2014.
- [11] Hilton, N. H., Schüppert, A., & Gooskens, C., "Syllable reduction and articulation rates in Danish, Norwegian and Swedish", *Nordic Journal of Linguistics*, 34(02), pp. 215-237, 2011.
- [12] Johnson, K., "Massive reduction in conversational American English", *Proceedings of the 1st Session of the 10th International Symposium Spontaneous Speech: Data and Analysis*, pp. 29-54, 2004.
- [13] Klatt, D. H., "Vowel lengthening is syntactically determined in a connected discourse", *Journal of Phonetics*, 3(3), pp. 129-140, 1975.
- [14] Klatt, D. H., "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence", *The Journal of the Acoustical Society of America*, 59(5), pp. 1208-1221, 1976.
- [15] Li, A., Gao, J., Jia, Y., & Wang, Y., "Pitch and duration as cues in perception of neutral tone under different contexts in standard Chinese", *Asia-Pacific Signal and Information Processing Assoc.*, 2014.
- [16] Pharao, N., *Consonant reduction in Copenhagen Danish*. Ph.D. Dissertation, University of Copenhagen, 2010.
- [17] Schachtenhaufen, R., "Looking for lost syllables in Danish spontaneous speech", *Linguistic Theory and Raw Sound*, pp. 61-88, 2010.
- [18] Tseng, C., & Lee, Y., "Speech rate and prosody units: Evidence of interaction from Mandarin Chinese", *Speech Prosody*, 2004.
- [19] Tseng, S., "Contracted syllables in Mandarin: Evidence from spontaneous conversations", *Language and Linguistic*, 6(1), pp. 153-180, 2005.
- [20] Vennemann, T., "Phonetic detail in assimilation: Problems in Germanic phonology", *Language*, 48(4), 863-892, 1972.
- [21] Wang, J., "Syllable duration in Mandarin", *Proc. of the Fifth International Conference on Speech Science and Technology*, pp. 322-327, 1994.
- [22] Wenk, B. J., "Speech rhythms in second language acquisition", *Language and Speech*, 28(2), pp. 157-175, 1985.
- [23] Xu, Y. "ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis", *Proc. Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, Aix-en-Provence, France, 2013.
- [24] Yang, B., *Perception and Production of Mandarin Tones by Native Speakers and L2 Learners*, Springer, Berlin/Heidelberg, 2015.