The productive acquisition of dental obstruents by Danish learners of Chinese

Rasmus Puggaard

Published as: Puggaard, Rasmus. 2020. The productive acquisition of dental obstruents by Danish learners of Chinese. In Dongyan Chen & Daniel Bell (eds.), *Explorations of Chinese theoretical and applied linguistics*, 168–195. Cambridge Scholars.

Page numbers from published version indicated in double square brackets in the running text.

Abstract

Standard Chinese has phonemic dental stops and affricates /t t^h ts ts^h/, while the aspiration of the Danish alveolar stop is manifested as affrication /t^s/. This paper tests whether Chinese dental obstruents are particularly problematic to acquire for Danish L1 speakers, as suggested by Wang (2014). It also tests whether Danish /t^s/, assumed to be problematic for the acquisition of the corresponding Chinese phoneme, is transferrable to Danish L1 speakers' pronunciation of English. 25 students of Chinese at different levels were asked to utter short sentences in Danish, English, and Chinese. The results unequivocally show that the participants of the experiment transfer affrication to their production of English, and that most Danes learning Chinese have problems with the production of the dental obstruents. The results also show that Danish L2 learners of Chinese pronounced /ts/ as a fricative, which is clearly distinct from Chinese /t^h ts^h/ and that the duration is significantly longer than for L1 speakers of Chinese. The primary distinguishing feature between /t^h ts^h/ appears to vary in learners' perception of duration but not in affrication. Moreover, for most first and second year students, and among half of the participating third year students, affrication is prominent in the production of Chinese phonemes /t^h t^{sh}/. To further complicate the picture, the students were also found to gradually develop fortis before they were able to pronounce Chinese /t^h/. This paper argues that if their pronunciation is purely aspirated this does not cause problems, but if affrication is retained in their pronunciation it only serves to make their production of /t^h ts^h/ more similar. *[[169]]*

Keywords: Second language acquisition, Chinese as L2, L2 phonology, acoustic phonetics

1. Introduction

The minor phonetic differences between otherwise similar phonemes in two different languages often lead to difficulty in second language (L2) acquisition. In a paper presented at Fudan University in October 2014 (Wang 2014), some specific production and perception difficulties were predicted for Danish speakers learning Standard Chinese (SC) as an L2. These predictions were based on existing theoretical models for the acquisition of L2 phonology, and specifically the acquisition of SC dental obstruents was emphasised. There is plenty of literature on L2 acquisition of SC phonology, particularly on the acquisition of tones (e.g. Yang 2015). But the study of SC acquisition by native speakers of Danish is a new field, albeit one that has been active as of late. Recent work on Danes' acquisition of SC includes a study by Sloos et al. (2015) on the acquisition of speech rhythm. Sloos et al. find that native speakers of Danish do not observe SC intrinsic tone duration, and do not have the characteristic final syllable lengthening of SC. Sloos et al. (2016) tested the pitch range and tone realization of SC, and found that the pitch variability of native Danish speakers was significantly lower than that of native SC speakers, and that tone realisation was often fairly poor, presumably due to the small pitch range of Danish. Yan and Sloos (2019) tested Danish learners' perception of the SC dental stops and affricates and found that native (L1) speakers of SC and Danish L2 learners rely on different cues when distinguishing them. Teachers of Chinese at Aarhus University also reported that productive acquisition of these sounds could be problematic. By comparing the phonologies of Danish and SC, this paper sets up a hypothesis for what problems Danish speakers may have when producing SC consonants in general, and tests whether the dental stops and affricates cause particular problems by analysing recordings of students at different levels producing them. As predicted by Wang (2014), it was found that a specific quirk of Danish phonology, namely the prominent affrication of alveolar stops, results in difficulties with the production of the SC dental obstruents. The problem is tenacious and is still found among some 3rd year students.

In Section 2 below, three models accounting for the perception of L2 sounds will be presented, before Section 3 uses these models to make predictions of difficulties Danish learners may meet when acquiring the simple onset consonants of SC. Section 4 will present the speakers, stimuli and procedure *[[170]]* used in the recordings for this paper, before Section 5 gives the method used for analysing the recordings. In Sections 6 and 7 the recordings are analysed, with Section 6 being a small experiment testing Danes' pronunciation of aspirated alveolar stops in English. The results are discussed in relation to the predictions in Section 8, while Section 9 sums up and concludes.

2. Theoretical background

Developing productive competence in the phonological system of an L2 as an adult is never easy, and Flege (1980: 119) notes that the development of a new phonological system appears slow compared to more abstract linguistic competences such as development of lexicon and syntax. When learning an L2 as an adult, the L1 phonological system interferes with that of the L2 in complex ways, both in perception and production of the L2. There are three well-known models accounting for how L2 sounds are filtered through the phonology of the L1: Kuhl's Native Language Magnet model (NLM; e.g.

Kuhl et al. 1992); Best's Perceptual Assimilation Model (PAM; e.g. Best et al. 2001); and Flege's Speech Learning Model (SLM, e.g. Flege 1995).

The NLM proposes that listeners develop acoustic prototypes for L1 phonemes very early in life. These acoustic prototypes serve as "magnets" in the surrounding perceptual space, making it harder to perceive variation surrounding L1 prototypes. Different non-prototypic members of a category will be perceived as more similar to the prototype than to each other, regardless of the actual acoustic difference between them (Kuhl et al. 1992: 607). Likewise, according to the NLM, L2 phones that are similar to L1 prototypes will be dragged into the perceptual space of the prototypes, and listeners will not develop prototypes for novel L2 categories (Best et al. 2001: 776). As evidence opposing the NLM, Lively and Pisoni (1997) report that the results supporting the NLM are much more robust for small children than for adults, and also find that phoneme prototypes are not stable across different phonetic contexts.

PAM was originally developed to account for the fact that speakers of American English were found to discriminate Zulu clicks surprisingly well; according to PAM, this is because the clicks were not perceived as speech sounds by L1 speakers of English, in which case the L1 phonology neither aids nor hinders discrimination (Best et al. 1988). L2 sounds that are not similar to any L1 sounds are deemed non-assimilable speech sounds, and discrimination among them is expected to be as good as the actual difference between the acoustic cues allows. Given that there is no interference from *[[171]]* L1 phonology in these cases, discrimination is expected to range from good to excellent. Discrimination between two L2 sounds is expected to be accurate when they are separated by phonological boundaries that also exist in L1, or if one L2 phoneme is similar to an L1 phoneme while the other cannot be easily identified with a specific L1 phoneme. When two L2 sounds are equidistant to one L1 phoneme, however, discrimination is expected to be poor. This is referred to as Single Category assimilation. Two L2 sounds may also be like one L1 phoneme, albeit with one being more similar than the other; this is termed a Category Goodness difference. Here, discrimination is also expected to be poor, but not as poor as in Single Category discrimination (Best et al. 2001: 777).

Of the three models, the SLM is the only one with primary focus on production of L2 sounds as opposed to perception, although the two aspects are assumed to be closely linked. The model mainly focuses on the production of fairly advanced L2 learners. The SLM assumes that L2 sounds are classified based on perceptive similarity to the nearest native category. The further the distance is from an L2 sound to the nearest L1 category, the better the chance is for a new stable sound category to develop. If an L2 sound is reasonably close to the nearest L1 sound, there is a good chance of it being perceived as part of the L1 category, which may result in production like a good exemplar of the L1 category, even though that may be a poor exemplar of the L2 sound (Flege 1995: 239). If any phonetic differences are discerned by the learner, a new phonetic category can be developed; however, the learner's category may not be similar to the corresponding category of a native speaker, since it may be based on different phonetic cues. An example of this is the production of English stops by native speakers of Canadian French: the

English aspirated stops /p t k/ have a much longer release duration than the corresponding French unvoiced stops. Caramazza et al. (1973, cited from Flege 1981: 450) report that L1 speakers of French produce English stops with a release duration somewhere in between that of their own French stops and that of an L1 speaker of English, indicating that while they have different categories for English and French /p/, their English /p/ category is still different from that of a native speaker. Besides, the SLM maintains that some L2 production errors are not perceptually motivated, but due to motoric output constraints from the L1 (Flege 1995: 238). PAM and SLM will both be used as predictors and explanatory models in the coming sections.

Apart from the interference of one phonological system with another, there is also evidence that there is an influence of L1 orthography-phonology correspondences when learning an L2. For example, Bassetti (2006a, cited *[[172]]* in Bassetti 2008: 197) found that Italian learners of SC were likely to pronounce SC unvoiced unaspirated stops /p t k/ as voiced /b d g/ due to Pinyin's use of <b d g> for these. Meng (1998, cited in Bassetti 2006b: 100) made a similar case for English learners of SC. This explanation should only be used when an observed problem cannot be explained using a phonological model; this problem can just as easily be explained using PAM or SLM.

3. Phonology of Danish and SC

There are many challenges for the Danish learner wishing to learn SC as an L2: a complex tone system has to be developed from scratch; in spite of the huge vowel phoneme inventory of Danish, there are still new ones to be learned; the learner must acquire retroflex consonants; and the learner must master a prohibitively complex new writing system, among other challenges. These will recur for speakers of most European languages, but there is one aspect of the Danish phonology that poses a unique challenge: the Danish alveolar stop is highly affricated in simple onsets.

Given that SC syllables only allow simple onsets (and a very restricted coda of either /n ŋ ɻ/), and given Flege's (1995: 239) hypothesis that L1 and L2 sounds are related to each other at a position-sensitive allophonic level and not a phonemic level, only the simple onset allophones of Danish will be taken into account here. In Section 3.1 and 3.2 below, consonants allowed in simple onsets of Danish and SC will be presented. In Section 3.3 the mapping possibilities of the phonemes will be analysed, highlighting the problem of the dental obstruents.

3.1 Danish

3.1.1 Danish consonants in onset position

The initial consonant allophones of Danish can be seen in Table 1, following Grønnum (1998: 39): *[[173]]*

	Labial	Alveolar	Palatal	Velar	Uvular	Glottal
Unaspirated plosive	b̥	d̥		g̊		
Aspirated plosive	p^{h}	t^{s}		k^{h}		
Fricative	f	s				h
Nasal	m	n				
Approximant	ʋ		j		ʁ	
Lateral		l				

Table 1. Initial consonant allophones of Danish.

Grønnum generally uses the less specific transcriptions /b d g p t k/ for the stop phonemes because their realizations differ in coda position; the above are the onset-specific realizations. The exact position and manner of articulation of the Danish rhotic is hard to describe, as it is not static at any point and has elements of both a glide and a trill (Grønnum 2007: 115).

Unlike e.g. English, Danish stops do not have a voicing distinction, but are exclusively voiceless – and nor is there a distinction between fortis and lenis in Danish stops, which are exclusively lenis (Grønnum 1998: 107, 263). This gives the set /b̥ d̥ g̊ p^{h} t^{s} k^{h}/, though narrowly the simple onset allophones of the aspirated set can be given as [b̥^{h}, d̥^{sh} ~ d̥^{s}, g̊^{h}]. The "voiced voiceless" symbols are used because lenis pronunciation of the unaspirated stops makes them acoustically similar to voiced stops in other languages (Grønnum 1998: 16).

3.1.2 Danish alveolar obstruents

As shown above, Danish has three phonemic alveolar obstruents /d̥ t^{s} s/. /t^{s}/ is of primary focus in this paper: as opposed to the other aspirated plosives, the majority of its release phonetically consists not of aspiration, but of affrication. It is not analysed as a phonological affricate, as that would result in an awkward gap in the phoneme inventory, and because that is hardly assumed to be in accordance with the cognitive organization of sounds. Instead, it is analysed as an aspirated consonant in which the majority of the aspiration occurs with the tip of the tongue in close proximity to the alveolar ridge, acoustically resulting in affrication. The affrication is sometimes followed by a short period of regular aspiration (glottal frication) after the tongue leaves the alveolar ridge, depending on the quality of the following vowel. As mentioned above, the shift from affrication to aspiration is gradual. Nothing in the recordings made for this paper (see Section 4) or the existing literature suggests that this affrication is currently a dialectal, *[[174]]* regiolectal, or gender-specific feature, although it was missing in some traditional Danish dialects (Puggaard 2018). Likewise, nothing in the recordings or the existing literature suggests that affrication of /t^{s}/ in simple onset is dependent on the phonetic environment.

In the recordings made for this paper, the mean voice onset time (henceforth VOT) of Danish /t^s/ in simple onset (henceforth DA-*t*) was 94ms. For the average token, this consisted of 72ms of affrication followed by 21ms of aspiration. Approximately half of the tokens (50%, n=138) consisted of no or only very short aspiration. In contrast, only 1% (n=3) of tokens had no affrication, and these tokens were all uttered by the same speaker. These numbers are similar to the results of Mortensen and Tøndering (2013: 52), who found an average VOT of between 83.8ms and 86.8ms depending on subsequent vowel height. The relatively shorter VOT found in that paper is assumed to be due to Mortensen and Tøndering's data consisting of spontaneous speech, as opposed to the data basis of this paper.

Mean VOT	**Mean frication**	**Mean aspiration**	**%w/o frication**	**% w/o aspiration**	**Range**
93ms	72ms	21ms	1	50	3-17cs

*Table 2. DA-*t *VOT.*

3.2 Standard Chinese

3.2.1 Standard Chinese consonants in onset position

Following Duanmu (2007: 24) the initial consonant phonemes of SC can be seen in Table 3. When Pinyin transcriptions differ from the phonetic symbols, they will be given in *italics* after the phoneme. Palatals are given as phonemes here, even though Duanmu does not consider them to be phonemic.

	Labial	**Dental**	**Retroflex**	**(Palatal)**	**Velar**
Unaspirated plosive	p, *b*	t, *d*			k, *g*
Aspirated plosive	p^h, *p*	t^h, *t*			k^h, *k*
Unaspirated affricate		ts, *z*	tʂ, *zh*	tɕ, *j*	
Aspirated affricate		ts^h, *c*	tʂ^h, *ch*	tɕ^h, *q*	
Fricative	f	s	ʂ ʐ, *sh r*	ɕ, *x*	x, *h*
Nasal	m	n			
Lateral		l			

Table 3. SC consonant phonemes in onset position.

[[175]]

Duanmu proposes a minimal phonemic inventory of SC, as he considers the glides [w j ɥ]to be allophones of the high vowels. Chao (1968: 21) describes the velar fricative as a uvular /χ/. Chao also describes /ʐ/ as a voiced continuant but considers it to be a rhotic /r/ phonologically. This makes sense from the point of view of phonological economy, since /ʂ ʐ/ otherwise make up the only phonemic set in SC distinguished by voicing. /ʐ/ is used here as it most closely resembles the phonetic realization.

Like Danish, SC stops are not distinguished by voicing, but by aspiration. Unlike Danish, SC has a lenis-fortis distinction. The phonological status of the palatals [tç tçʰ ç] is unclear; while Cheng (1973: 40) gives them phonological status, they have also been identified as allophones of the dental affricates (Hartman 1944: 38) and the velar stops (Chao 1968: 21) respectively, while Duanmu (2007: 31) analyses them as consonant-glide combinations, specifically combinations of the dental sibilants and palatal glide. Kratochvil (1968: 25) furthermore describes the dental stops as alveolars /t tʰ/, but his claim appears to be unfounded, while a dental place of articulation is backed up by x-ray and palatographic data (Zhou and Wu 1963, cited in Duanmu 2007: 25).

3.2.2 SC dental obstruents

The focus of this paper is the dental obstruent series /t̪ t̪ʰ ts̪ ts̪ʰ s/, and particularly /t̪ʰ ts̪ ts̪ʰ/. Henceforth the Pinyin letters for the aspirated dental stop and dental affricates SC-*t*, SC-*c*, SC-*z* will be used instead of phonetic symbols.

In the recordings of native SC speakers made for this paper, the mean VOT of SC-*t* lasted 86ms, consisting either of purely glottal frication, or velar frication before back vowels. This is slightly shorter than for Danish, but longer than for English (Lisker and Abramson 1964: 394), even though all three languages are sufficiently similar in this respect to be grouped together in Cho and Ladefoged's VOT typology (1999: 223). The VOT mean is similar to that found by Chao and Chen (2008: 223), who measured a mean VOT of 81ms for SC-*t*, but markedly shorter than that found by Rochet and Fei (1991: 105), who measured a mean VOT of 98.7ms for SC-*t*. This difference is presumably due to Rochet and Fei measuring elicited single syllable utterances. The mean VOT of SC-*z* in the recordings for this paper lasted 73ms, consisting solely of coronal frication. These numbers are similar to those of Liu and Jongman (2013: 4) who report a release between 60ms and 70ms. The mean VOT of SC-*c* in the recordings for this paper was 152ms, consisting of roughly two thirds affrication and one third aspiration *[[176]]* (99ms and 53ms on average, respectively). Similar to SC-*t*, the aspiration often consisted of velar frication before back vowels. 11% of the tokens had no audible aspiration; these were the tokens occurring in syllables with a syllabic voiced continuant /z̩/ in the rhyme, which is to be expected, as the place of articulation of the continuant is homorganic to that of the affricate (e.g. Hartman 1944: 31). The mean VOT measured for SC-*c* is similar to that found by Liu and Jongman (2013: 4). The VOT values in the recordings made for this paper are generally similar to previously findings.

Phoneme	Mean VOT	Mean frication	Mean asp.	% w/o frication	% w/o aspiration	Range
SC-*t*	86ms	0ms	86ms	100	0	3-15cs
SC-*z*	73ms	73ms	0ms	0	100	2-15cs
SC-*c*	152ms	99ms	53ms	0	11	8-29cs

Table 4. SC phonemes VOT.

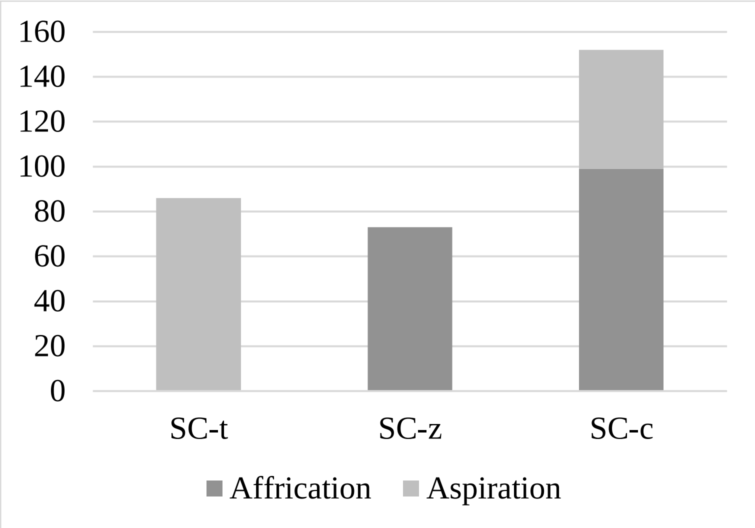

Figure 1. Illustration of the duration and contents of the releases of SC dental aspirated stops and affricates.

3.3 Mapping expectations

There are plenty of minor differences in the Danish and SC simple onset consonants. Many of these are likely to lead to different degrees of accented pronunciation of SC by native speakers of Danish.

SC /p k m f/ are expected to map to the corresponding sounds in Danish with no or only minor problems, as the sounds are highly similar. /pʰ kʰ/ are also *[[177]]* expected to map to the corresponding sounds in Danish, the only difference being that the Danish stops have lenis pronunciation. Given that the respective VOTs are similar (Mortensen and Tøndering 2013; Chao and Chen 2008), fortis pronunciation is not expected to be adopted during SC acquisition. /t s n l/ are alveolar in Danish and dental in SC, which should also be unproblematic, as the acoustic differences are minimal. Following SLM, it is likely that an alveolar place of articulation will be retained. Regarding the stop /t/, Cho and Ladefoged (1999: 220) find that dental and alveolar places of articulation are not in themselves expected to result in differences in VOT, but Danish /d̥/ has been found to be up to 10ms longer than Chinese /t/ (Mortensen and Tøndering 2013: 51; Chao and Chen 2008: 223).

The nearest Danish equivalent to both the retroflex and palatal sets /tʂ tʂʰ ʂ tç tçʰ ç/ is the alveolopalatal set [dʑ tɕ ɕ], phonologically considered to be /tˢj d̥j sj/ by Grønnum (1998: 264-266). In PAM terms the difference here is assumed to be one of Category Goodness, with the SC palatals being the better match. Lai (2009: 1268) found that perceptive discrimination of retroflex and palatal affricates was generally unproblematic for native speakers of Malay and Burmese, both of which only have postalveolar affricates.

The SC velar/uvular fricative /x/ does not fit neatly into any Danish phoneme category. Pinyin interference may trigger /h/ for some Danish learners, particularly since /h/ is phonetically the most similar Danish segment, although a very poor fit. /x/ is expected to be sufficiently dissimilar to any Danish sound to form a novel category. Alternatively, many learners may already have formed a category /x/ due to experience learning German as an L2.

SC-*t*, SC-*z*, and SC-*c* are all fairly close to DA-*t* in different respects, making acquisition either a case of Single Category assimilation or a Category Goodness difference in PAM terms. SC-*t* is the closest to DA-*t* in duration. The complete lack of affrication makes it initially seem very different from DA-*t*, but since the affrication is presumably of low cognitive significance to the native speaker of Danish, it may not be of much importance to the transfer. Danish native speakers are also assumed to normally maintain the feature of strong affrication when pronouncing aspirated alveolar plosives /t^h/ in other L2s (see Section 6). Orthographic interference from Pinyin may also trigger affricated realization. *[[178]]*

Inspecting only VOT, SC-*z* appears very similar to DA-*t*. The mean VOTs are reasonably similar and considering that 50% of DA-*t* tokens have no audible pure aspiration, those tokens should be very similar to the native SC-*z* pronunciation. However, there are reasons why they are less similar than these numbers indicate: SC-*z* never has aspiration, and the affrication of SC-*z* and that of DA-*t* are dissimilar, due to the gradual shift of DA-*t*'s affrication to more aspiration-like noise. The affrication noise of DA-*t* peaks immediately after the plosive release, while the affrication noise of the SC-*z* release is more constant. The affrication noise of SC-*z* also generally peaks at a higher frequency than that for DA-*t*. An unrelated reason for Danish speakers to be more likely to recognise SC-*z* as a separate category is that German has a similar alveolar affricate /ts/ in syllable-initial position (Antonsen 2007: 26). Finally, the relatively weak plosive release and short duration of SC-*z* means that it may be perceived as closer to Danish /s/ than DA-*t*. In this case both SC-*z* and SC /s/ would compete for the Danish /s/ category, with SC-*z* being the weaker fit. The results of the current paper point towards the latter analysis.

SC-*c* is much longer than DA-*t*, but the phonetic realization is otherwise much more similar than the other contenders, as it is the only one of the consonants in which both affrication and aspiration are important cues. Aspiration proper is more prominent in SC-*c*, and unlike DA-*t*, the shift from affrication to aspiration in SC-*c* is not gradual. But the distribution of affrication and aspiration makes SC-*c* phonetically very similar to Danish DA-*t*.

SC-*t* and SC-*c* are more or less equidistant to DA-*t*, being similar to it in different respects. An outcome of this could be that the early SC interlanguage of Danish L1 speakers will have SC-*t* and SC-*c* grouped in a single L2 category, /t^s/, which must then be split at a later stage. Otherwise, both SC-*t* and SC-*c* are expected to retain unwanted features from DA-*t*, i.e. affrication is assumed to be retained in SC-*t* and a gradual shift from affrication to aspiration is assumed to be retained in SC-*c*, which would make the two L2 categories highly similar. SC-*z* is not assumed to be an equally good candidate, but could turn out to be grouped with the other two in the early interlanguage, especially considering Yan and Sloos' (2019) findings that the phonetic cues of affrication and aspiration are weighted differently for Danish learners than for L1 SC speakers when distinguishing between SC-*t*, SC-*c*, and SC-*z*. *[[179]]*

4. Materials

4.1 Speakers

25 undergraduate students were recorded for this paper. All students were native speakers of Danish and enrolled in the Chinese Studies program at Aarhus University. The students were between 20 and 42 years old, though all except two participants were between 20 and 25. More than half of the participants were from Central Jutland, the remainder coming from all across Denmark. 11 were male and 14 were female, which reasonably reflects the gender distribution of the study program. All of the participants spoke English, all except two had studied German for at least three years in elementary school, and approximately half had studied French in high school. 10 of the participants had studied Chinese in high school, 9 more reported having some knowledge of Chinese before starting university, and 6 reported having no previous knowledge of Chinese whatsoever prior to university. The subjects were unpaid and did not receive credit for participating. None of the participants reported hearing, vision, or reading deficiencies. They were unaware of the particular hypothesis being tested. The subjects were at three different stages of their studies:

1st year students (henceforth Y1). Seven of the participants had begun their studies approximately five weeks before the recordings took place. At this point they were reasonably proficient at using *Pinyin but* had only a small inventory of Chinese characters and a small SC vocabulary.

2nd year students (henceforth Y2). 12 of the participants had begun their studies a little more than one year previously. At this point they had a reasonably large inventory of Chinese characters and vocabulary. They had been introduced to all characters and words used in the recordings.

3rd year students (henceforth Y3). Six of the participants had begun their studies a little more than two years previously. As part of their studies they had recently finished a semester at Peking University, where teaching had been done solely in SC, so they had been exposed to a large amount of both spoken and written Chinese.

In addition to the Danish participants, 7 native speakers of SC were recorded, 4 males and 3 females. They were all studying in Aarhus as exchange students. They were from different areas of China, but all spoke SC as their native variety of Chinese. No factor other than study level resulted in major differences in the results for the Danish students. *[[180]]*

4.2 Stimuli

The stimuli consisted of 64 simple sentences of 6 syllables each. The sentences contained the same number of syllables to control for speech rate. 12 of the sentences were in Danish to get samples of the participants' L1 speech; 12 were in English; and 40 were in SC. The Chinese participants were only asked to read the SC sentences. The Danish and English sentences were both designed to include the aspirated alveolar stop in front of a broad variety of vowel combinations. For both sets, 9 sentences included the alveolar stop and

3 were filler sentences. The SC sentences were designed to include the dental obstruents in focus before a broad variety of vowel and tone combinations. 30 of the sentences included one or more of these consonants while 10 were filler sentences.

4.3 Procedure

The participants were asked to read the sentences presented above from slides of a Microsoft PowerPoint presentation. All sentences were presented on single slides, the Chinese ones in both characters and Pinyin. The recordings were self-paced and lasted 3-6 minutes on average. The speech was recorded with either a Zoom APQ3HD or a Sony PCM-D50. The recordings mostly took place in the Interacting Minds Centre lab at Aarhus University, but also in three other empty, quiet meeting rooms at Aarhus University and VIA University College respectively.

5. Analysis methodology

To investigate how L1 Danish speakers' pronunciation of dental obstruents differed from native pronunciation in SC, the Praat speech processing software (Boersma and Weenink 2016) was used to measure and analyse the release portion of the stops and affricates. The term VOT will be used to broadly signify the period from the plosive release until the beginning of voicing, thus covering the entire release of plosives and affricates. For each of the obstruents in focus, the VOT was measured and, if applicable, split into an affrication segment and an aspiration segment, which were measured independently. If either affrication or aspiration lasted less than 1cs they were considered inaudible, and thus given as 0. If consonants were mispronounced in a manner that suggested that they were speech errors, they were excluded from the analysis. This mainly covers consonants pronounced at an entirely wrong place of articulation, e.g. dentals that were pronounced as retroflexes or velars. *[[181]]*

VOT, affrication segments, and aspiration segments were measured in cs. For SC speakers, delimitation of affrication and aspiration was generally clear and unproblematic, and could reasonably have been given in ms, but for Danish speakers, more precision proved problematic, due to the gradual transition from affrication to aspiration in Danish pronunciation. This gradual transition was prominent among Danish participants in both their L1 and their L2s. Delimitation of affrication and aspiration was based on spectrographic and auditory inspection of the individual tokens; if possible, high-pitched irregular noise was taken to mean affrication while pre-voicing formant traces was taken to mean aspiration. However, in the Danish speaker tokens, these frequently overlapped, in which case delimitation was based mainly on auditory judgment. An example of this can be seen in Figure 1, where unvoiced vowel formants can clearly be seen simultaneously with diminishing high-pitched irregular noise. When voicing did not start simultaneously for all formants, voice onset was taken to be at the beginning of voicing at F1. Fischer-Jørgensen and Hutters (1981) make a good case for considering everything until voicing at

F2 and F3 aspiration, but this approach is better suited for measuring the duration of the whole release portion of a stop, and not for analysing different aspects of the release.

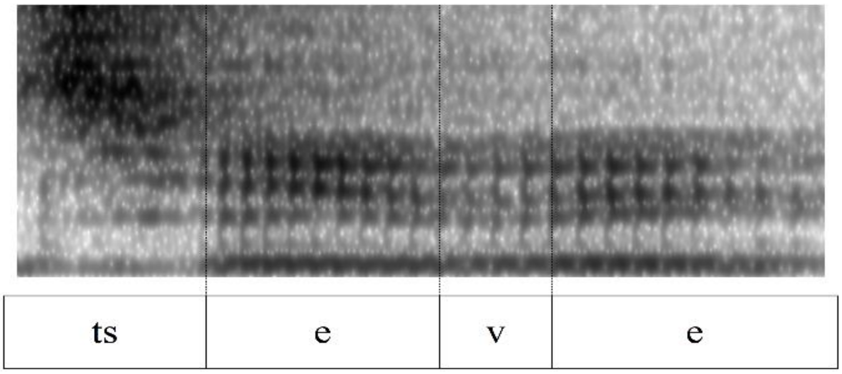

Figure 2. Unvoiced vowel formants simultaneous with diminishing fricative noise.

6. Affrication in L2 English

As noted above, it has often been mentioned, but not tested, that Danes often pronounce /t^h/, fortis /t/ and similar phonemes in different L2s as [t^s], even in languages which they are very proficient in. If true, this is assumed to increase the likelihood for CH-*t* to be pronounced as [t^s]. To test this hypothesis, a subpart of the current experiment tested the pronunciation of the English aspirated alveolar stop by L1 speakers of Danish. The participants all spoke English with a high degree of proficiency. 25 students of Chinese can naturally not be said to be representative of the whole *[[182]]* population of English-speaking Danes, so work is still needed on this topic, but the results are sufficiently clear to have some merit. No native English speakers were recorded for comparison, so comparison of VOT will be made to the values reported in Docherty (1992: 116) and Klatt (1975: 689). They find mean VOTs for /t^h/ at 63ms and 65ms for British English and American English respectively. Neither of them refers to affrication in these stops, and such a feature is generally not mentioned in the literature.

It turned out that there was a small but significant difference in the VOT of DA-*t* and English /t^h/ as pronounced by L1 Danish speakers (henceforth EN-*t*), with $t(523)=2.6$, $p<0.01$. The mean EN-*t* had a VOT of 88ms, which is just 6ms shorter than DA-*t*, and still around 20ms longer than native pronunciation of EN-*t*. Likewise, affrication was slightly less prominent in EN-*t* than in DA-*t*; EN-*t* had a mean affrication duration of 66ms, thus taking up an average of 74% of the release (compare 77% for DA-*t*). In total, affrication was avoided in 6% (n=16) of EN-*t* tokens, which is a lot more frequent than DA-*t*, but still means that there is affrication in most tokens. Half of the tokens with no affrication (n=8) were spoken by only two of the study participants. 48% (n=120) of EN-*t* tokens had no period of aspiration proper, which is roughly similar to DA-*t*. In support of SLM, it appears that at least some of the speakers are aware of an acoustic difference between EN-*t* and DA-*t*, and have developed a separate category for EN-*t* which is based on their perceived difference between the two.

Phoneme	Mean VOT	Mean frication	Mean asp.	% w/o frication	% w/o aspiration	Range
EN-*t*	88ms	66ms	23ms	6	48	4-16cs
DA-*z*	93ms	72ms	21ms	1	50	3-17cs

*Table 5. EN-*t *vs DA-*t *VOT*

[[183]]

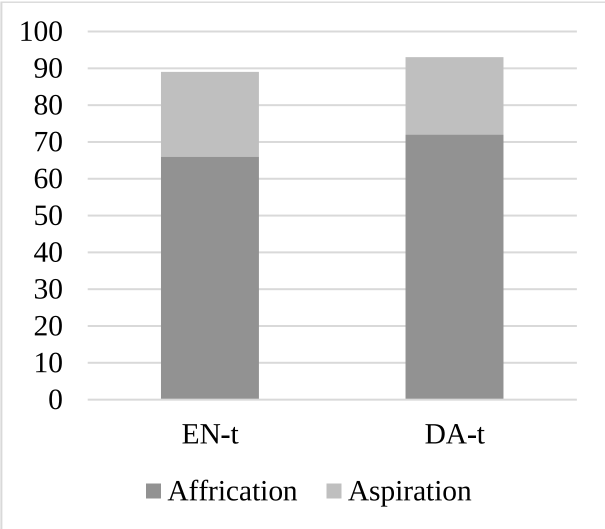

*Figure 3. Illustration of the duration and contents of the EN-*t *and DA-*t *release.*

In conclusion, EN-*t* is remarkably close to DA-*t* in pronunciation, and it can safely be said that affrication has been retained as a prominent feature in EN-*t*. As with DA-*t*, the feature is prominent for all recorded speakers – including those who consistently avoided it in their pronunciation of Chinese SC-*t* (see below). This means that even those students who had successfully established /tʰ/ as a stable category in SC retained [tˢ] in their pronunciation of English.

7. Results

*7.1 SC-*t

Unlike SC-*z* and SC-*c*, SC-*t* can appear before /i/ and can be palatalized, and both /t̪ʰi/ and initial /t̪ʰʲ/ were involved in the experiment. They were, however, left out of the final analysis for two reasons: 1) since the dental affricates cannot appear before high front vowels, SC-*t* will most likely not be perceived as an affricate whether pronounced with affrication or not, and 2) aspiration preceding a high front vowel is acoustically very similar to affrication, making delimitation of either very difficult and ultimately too random. The latter proved true for both Danish and native SC speakers.

SC-*t* was the only analysed consonant in which the participants' total VOT became progressively less native-like as they advanced in their studies. The mean total VOT for Y1 students was 92ms and 95ms for Y2 students, and while both were longer than the mean VOT for native Chinese speakers, neither differed significantly from it. The Y3 students, however, had a mean *[[184]]* total VOT of 112ms, which was significantly longer than that of native speakers, with $t(205)=5.34$, $p<0.001$. It should be mentioned that the Y1 and Y2 students' VOTs were not significantly different than the VOT of DA-*t*, while those of the Y3 students' were, with $t(368)=5.49$, $p<0.001$. It should also be mentioned that the long VOT for Y3 students can be partially explained by one of the participants of this group having a particularly long mean VOT of 152ms, raising the Y3 mean by a fair amount. However, if the outlier participant is excluded, the Y3 mean VOT is at 104ms, which remains significantly higher than both the Y1, Y2, and native SC groups.

The average Y1 SC-*t* consists of 41ms of affrication followed by 50ms of aspiration. None of the actual tokens look like this; this average is a product of highly variant pronunciation. More than half (65%, n=73) of the tokens had affrication, and among those the mean affrication duration was 64ms. Out of the Y1 participants, 4 out of 7 affricated all or almost all of their tokens, 1 affricated approx. half of their tokens, and 2 almost never affricated. A third (30%, n=34) of the tokens had less than 10ms of or no aspiration. The majority of Y1 SC-*t* releases consisted of a fairly long period of affrication sometimes followed by a fairly short period of aspiration.

The average Y2 SC-*t* consists of 24ms of affrication and 71ms of aspiration. This is also a product of variant pronunciation, but clearly shows that by this stage of study, affrication has become much less prominent in the production of SC-*t*. Less than half (43%, n=83) of Y2 tokens had affrication, while only 9% (n=18) lacked aspiration. None of the Y2 students affricated all of their tokens, but only 2 out of 12 consistently avoided it. The majority of the Y2 SC-*t* releases consisted either only of aspiration or of a fairly short period of affrication followed by a longer period of aspiration.

The average Y3 SC-*t* consists of 31ms of affrication and 81ms of aspiration. Similar to Y2, 44% (n=42) of tokens had affrication, and 11% (n=10) lacked aspiration. These numbers are less native-like than the Y2 ones. There are two explanations to account for this development: 1) a longer VOT makes affrication a less prominent part of the release and 2) a longer VOT is a product of the students developing fortis pronunciation of the phoneme. A more fortis pronunciation of a Danish affricated SC-*t* would acoustically result in a longer release retaining both affrication and aspiration. A possible adverse effect of this strategy is that the prolonged VOT could simply cause these tokens to resemble SC-*c* more than SC-*t*. For an extreme example of this, see Figure 4, which is a SC-*t* token with a very long release (290ms) that has both affrication and aspiration. Presumably the prolonged VOT is an attempt to stress the aspiration of the release, but in doing so, affrication *[[185]]* also becomes more prominent, and the token comes to altogether resemble native SC-*c* more than SC-*t*.

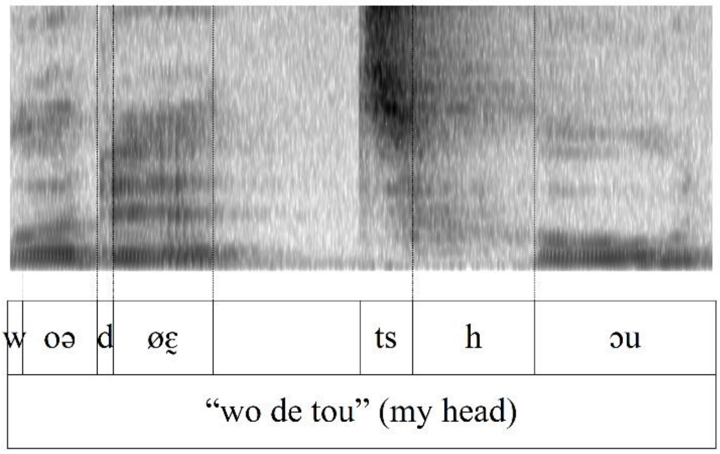

*Figure 4. Example of a very lengthy SC-*t *token with both affrication and aspiration being prominent.*

Half of the Y3 students consistently or near-consistently avoided affrication, which is a clear improvement from Y2. The majority of the Y3 SC-*t* releases had no affrication, while some consisted of both affrication and aspiration, but with fairly long VOTs, which both stresses the aspiration of these tokens, and results in a less native-like release duration.

Group	Mean VOT	Mean frication	Mean asp.	% w/o frication	% w/o aspiration	Range
Y1	92ms	41ms	50ms	35	30	4-21cs
Y2	95ms	24ms	71ms	57	9	3-19cs
Y3	112ms	31ms	81ms	56	11	5-29cs
L1	86ms	0ms	86ms	100	0	3-15cs

*Table 6. SC-*t *VOT*

[[186]]

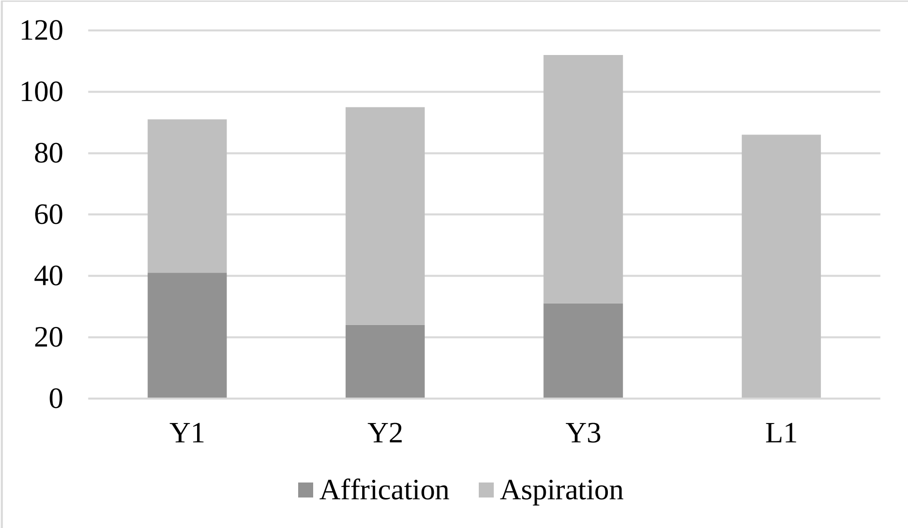

*Figure 5. Illustration of the duration and contents of the SC-*t *release at different stages of study and for L1 speakers.*

*7.2 SC-*z

The pronunciation of SC-*z* became progressively more native-like as the students advanced in their studies. The mean total VOT for Y1 students was 131ms. The Y2 students' VOT was significantly shorter at 119ms with $t(239)=2.64$, $p<0.01$. Likewise, the Y3 students' VOT was significantly shorter than the Y2 students at 106ms with $t(231)=2.06$, $p<0.05$. However, the Y3 students' VOT was still significantly longer than that of the native SC speakers, with $t(179)=6.10$, $p<0.001$. Except for a few mispronounced tokens, the participants never produced it with aspiration, and it is clearly pronounced distinctly from SC-*t* and SC-*c*. The only difference across the stages of study is the duration of the affrication. Note that while the duration gradually improves, the Y3 mean duration is still much longer than the native mean.

Danish production of SC-*z* often has no plosive release at all, making it phonetically very similar or identical to /s/. This was a very frequent issue, happening in 25% (n=84) of all Danish tokens. Some speakers never made this error, but more than half of the participants made it at least once. There was no significant improvement as their studies progressed: between 20% and 30% of all tokens for a given group lacked plosive release. For an example, see Figure 6 showing a SC-*z* token with no plosive release as said by a Y3 speaker. At 11cs, the token is near the average duration for Y3 students. *[[187]]*

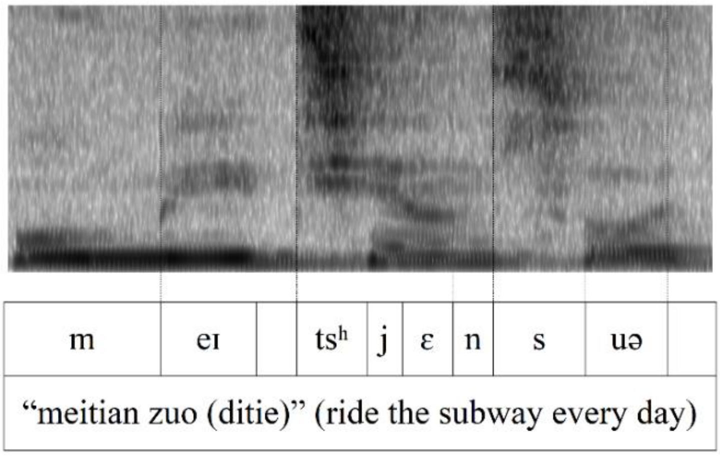

Figure 6. Example of a SC-z token with no plosive release.

It should be noted that a few of the native SC tokens also lacked plosive release, but these cases can all be accounted for by rapid speech, which cannot be said for the Danish speakers.

Group	**Mean VOT**	**Mean fr.**	**Mean asp.**	**% w/o fr.**	**% w/o asp.**	**% w/o pl.rel.**	**Range**
Y1	131ms	131ms	1ms	0	99	23	5-23cs
Y2	119ms	118ms	1ms	1	98	27	5-29cs
Y3	106ms	105ms	1ms	0	99	22	4-23cs
L1	73ms	73ms	0ms	0	100	4	2-15cs

Table 7. SC-z VOT.

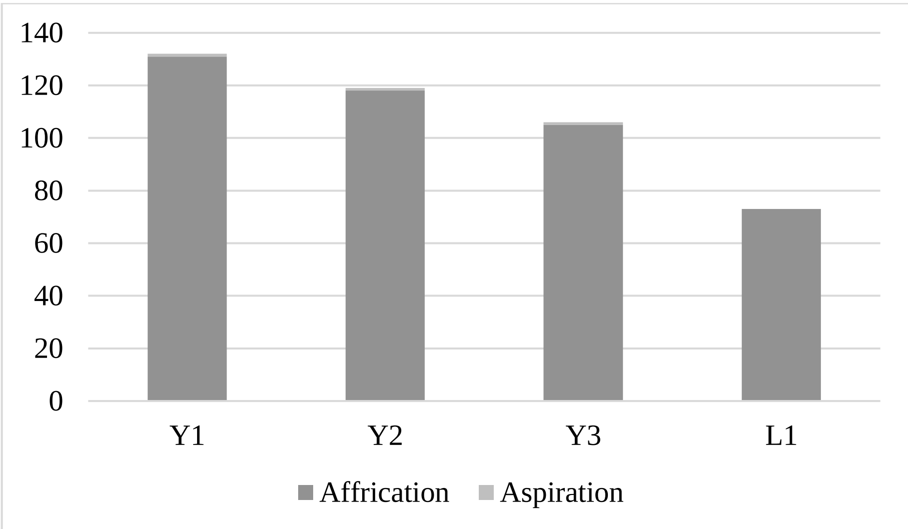

Figure 7. Illustration of the duration and contents of the SC-z release at different stages of study and for L1 speakers.

*7.3 SC-*c

The pronunciation of SC-*c* became progressively more native-like as the students advanced in their studies. The VOT is not significantly different between Y1 and Y2 at 110ms and 109ms respectively. The Y3 mean VOT is significantly longer at 125ms, with $t(171)=3.13$, $p<0.01$. While still much shorter than the native pronunciation, it is a clear improvement.

The average Y1 SC-*c* consists of 72ms of affrication followed by 38ms of aspiration. The mean distribution is like the L1 SC production, with 65% of the release consisting of affrication, but once again the average is a result of very variant pronunciation. This early in their studies, the participants clearly had troubles with the pronunciation of SC-*c*; 14% (n=10) of the tokens were mispronounced to the point of exclusion, frequently as a velar stop /k/ which can only be explained by orthographic interference from Danish or English. 45% (n=27) were pronounced with no or minimal aspiration, though a clear fortis pronunciation typically distinguished these from SC-*z*. To further complicate the picture, 13% (n=8) were pronounced with no affrication, making these tokens more like native SC-*t*, while still typically longer. This is presumably a form of hypercorrection, demonstrating the students' problems with keeping SC-*t* and SC-*c* separate. An example of this can be seen in Figure 8, where a Y1 student pronounces a SC-*c* token with no signs of affrication, but a fairly long period of aspiration (13cs).

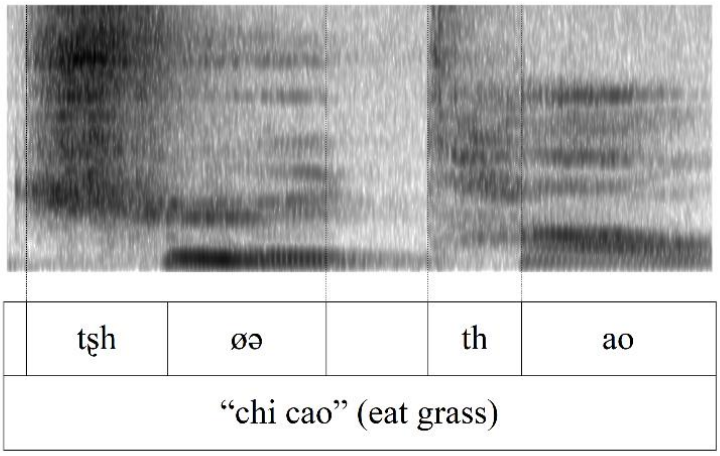

*Figure 8. Example of lengthy SC-*c *token with no affrication.*

[[189]]

The average Y2 SC-*c* consists of 64ms of affrication followed by 45ms of aspiration. This means that the mean distribution is less native-like than at Y1, with only 59% of the average VOT consisting of affrication. This is actually a result of the pronunciation becoming more stable, and aspiration becoming more prominent: only 27% (n=31) of tokens were pronounced without aspiration. Along with much fewer tokens discarded due to mispronunciation (6%, n=7), it is clear that SC-*c* production has stabilized and become less problematic. Lack of affrication was still surprisingly prominent, occurring in 17% (n=19) of Y2 SC-*c* tokens.

The average Y3 SC-*c* consists of 86ms of affrication followed by 39ms of aspiration. Along with duration, the distribution has also grown more native-like, with affrication lasting an average of 69% of the total duration. Lack of aspiration is still a significant issue, seen in a third (33%, n=20) of tokens. Lack of affrication has, however, ceased to be a significant issue (3%, n=2).

Group	Mean VOT	Mean frication	Mean asp.	% w/o frication	% w/o aspiration	Range
Y1	110ms	72ms	38ms	13	45	3-25cs
Y2	109ms	64ms	45ms	17	27	5-20cs
Y3	125ms	86ms	39ms	3	33	6-23cs
L1	152ms	99ms	53ms	0	11	8-29cs

*Table 8. SC-*c *VOT.*

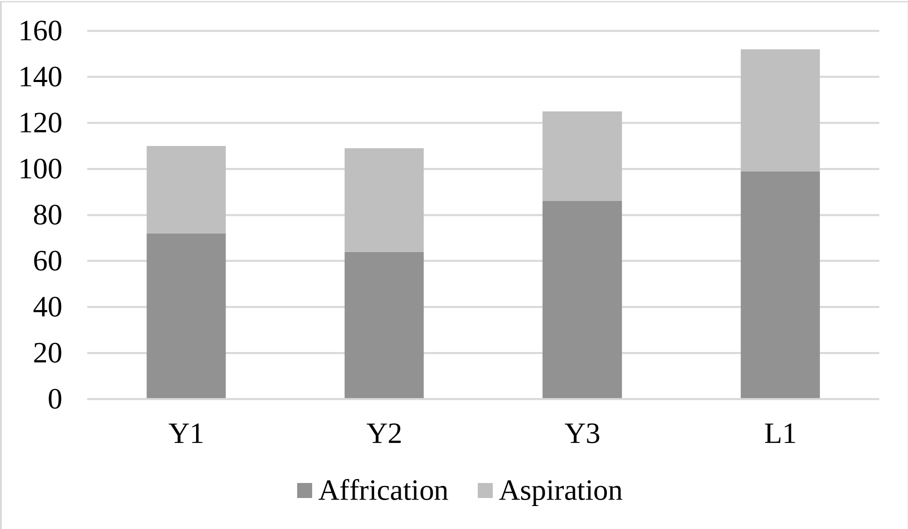

*Figure 9. Illustration of the duration and contents of the SC-*c *release at different stages of study.*

[[190]]

8. Discussion

The results of the experiment presented in Section 6 showed that at least some participants had developed a novel category EN-*t* separate from DA-*t*, and that for all of these participants the difference between the two categories was one of duration and not one of affrication. It appears that affrication is a very stable feature of the DA-*t* category, which should pose a problem to SC acquisition; in some ways it does, but in some respects it does not. Looking at the data in Section 7, it appears that different categories for SC-*t*, SC-*z* and SC-*c* are established early on in SC acquisition.

The category for SC-*z* is established as distinct early on, being much longer than SC-*t* and SC-*c* and never having aspiration. This may be because SC-*z* is more likely to be perceived as Danish /s/ rather than DA-*t*. In any case, plosive release in SC-*z* is not as important a cue for Danish speakers as it is for L1 speakers. Danish speakers may use other acoustic cues to distinguish the SC-*z* from /s/; the recordings indicate that an initially very narrow constriction of SC-*z* may be an important cue. The possibility that SC-*z* might be categorically closer to /s/ than to DA-*t* was not taken into account when the experiment was designed, but a detailed comparison of SC-*z* and SC /s/ as pronounced by Danish speakers would be very enlightening, and could also help provide a solid explanation for the extremely long release of SC-*z* among Danish speakers. It would also be very interesting to see whether SC /s/ sometimes gets plosive release when pronounced by Danish speakers, which is likely to be the case.

SC-*t* and SC-*c* are also more or less established as separate categories for most or all speakers even by the 5th week of studying Chinese. This is corroborated by the fact that SC-*c* is 18ms longer than SC-*t* on average for the Y1 group. But both categories show very differing pronunciation patterns within-group, and there is a significant overlap of SC-*t* and SC-*c* pronunciation. The SC-*t* category appears to still be identical to DA-*t* for some Y1 speakers. For many speakers at Y1 level, the distinguishing feature between SC-*t* and SC-*c* appears to be one of intensity, and not affrication. Affrication is typically a stable feature of both categories, so they are distinguished instead by a more fortis pronunciation in SC-*c*, which results in a longer release duration. Of course, in native SC, both SC-*t* and SC-*c* are fortis, but that is presumably hard for Danish listeners to perceive if SC-*t* is assimilated with DA-*t*. The categorization strategy may work well for perceptual distinction, but for productive distinction it is problematic, considering the significant overlap it is likely to create between the two categories in casual speech. *[[191]]*

At Y2 and Y3 level, SC-*c* becomes progressively more native-like as its release duration increases, but it does not necessarily become progressively more distinct from SC-*t*. Generally, SC-*t* gradually develops a more fortis pronunciation, causing an increase in release duration, even though it was already long compared to native pronunciation. This is true for almost all participants in the experiment; at Y3, the individual participants have average release durations ranging from 97ms to 152ms for SC-*t*, meaning that even the speaker with the shortest average VOT has a longer VOT than the average for all Y2 participants. For some speakers, affrication ceases to be a feature of SC-*t*, and for some it does not. As far as productive distinction goes, this makes all the difference: if SC-*t* is pronounced with a prolonged release duration but with no affrication, a native speaker of SC is expected to readily perceive it as SC-*t*. But if it is pronounced with both a prolonged release duration and affrication, what is left to distinguish it from SC-*c*? The release of SC-*c* is only 13ms longer on average, so a very significant overlap between the two is to be expected. As mentioned above, only half of the Y3 participants consistently avoided affrication, which means that this issue is very tenacious.

While different categories for SC-*t* and SC-*c* were established early on, they were relatively unstable for at least Y1 and Y2. A significant number of SC-*c* tokens were realized simply as aspirated stops /t^h/. There are two possible explanations for this: 1) It is due to hypercorrection among students who are particularly aware that they have to aspirate and not affricate their SC-*t*s; the fact that this transfers to SC-*c* shows that the distinction between the two is still not entirely straightforward; or 2) affrication is not a distinguishing feature of either SC-*t* or SC-*c* for the students who made the error; the non-affricate tokens typically have a long release, indicating fortis pronunciation, which is an important feature for SC-*c*. This would support the idea that affrication of coronal stops is not very cognitively salient to native speakers of Danish. It may simply be that some students distinguish SC-*t* from SC-*c* by duration or fortis pronunciation, with affrication being an optional feature of both. The reason is probably a combination of both factors. As Flege (1981: 446) points out, it is not at all clear whether "native-like pronunciation at the level of phonetic implementation is even necessary for accent-free speech". But in this case, the difference between phonetic implementation of the Danish /t^s/ category and the SC /t^h ts^h/ categories may cause significant communicative problems. Further research is needed on the extent of these communicative problems, and the extent to which these sounds are perceived as the right or wrong categories by native speakers of SC remains unknown. *[[192]]*

A solution to the problem may lie in the early teaching of SC pronunciation. As mentioned above, affrication of DA-*t* may not be a very cognitively salient feature for the native speaker, and Danes are likely to be unaware of the feature. The feature will certainly be easier to eradicate from their SC pronunciation if they are aware of it. When teaching SC phonetics to Danish learners, teachers are advised to make students aware that affrication is a prominent feature of aspirated Danish stops and not of aspirated SC stops.

9. Conclusion

Danish phonology is poorly equipped to help categorize the SC phonemes /t^h ts ts^h/ in a meaningful manner, and this paper set out to test whether Danes could productively discriminate between the sounds. The findings suggest that Danes generally make a productive distinction between the sounds even at an early level of their studies, though the cues used for the productive discrimination are problematic. SC /ts/ turns out to be productively more similar to Danish /s/ than /t^s/, and while it is certainly a category in itself, Danes fairly often produce the affricate with no initial plosive release, making it simply a fricative which sometimes has a very narrow constriction at the onset. Danish phonology makes native speakers of Danish likely to perceive and produce aspirated stops as lenis; an effect of this is that the main discriminating feature between SC /t^h ts^h/ for some Danes turns out not to be affrication, but duration – especially in the early Chinese interlanguage. Over time, fortis pronunciation of /t^h/ is developed, which only results in a longer and less native like release duration, especially since some speakers fail to develop the distinction between affrication and aspiration. In the third year of their studies, some participants consistently produced /t^h/ with no affrication, while some still struggled to pronounce it distinctly from /ts^h/. The results are in line with the predictions made based on PAM, and SLM provides a good explanatory basis for the problems in production.

Acknowledgments

I would like to thank all the participants in the study, and especially Zhang Chun for her help in organising it, and Jakob Steensig for lending me recording equipment. I would also like to thank William McGregor, Marjoleine Sloos, and Mads Lundegaard for their useful comments on a previous version of this paper, as well as the participants of the first ISOCTAL. *[[193]]*

References

Antonsen, Elmer H. 2007. *Elements of German. Phonology and morphology*. University of Alabama Press.

Bassetti, Benedetta. 2006a. Orthographic input and phonological representations in learners of Chinese as a foreign language. *Written Language and Literacy* 9(1), 95–114. doi:10.1075/wll.9.1.07bas.

Bassetti, Benedetta. 2006b. Pinyin orthographic input and CFL teachers' pronunciation. Paper presented at *The British Chinese Language Teachers' Seminar*, Cambridge University.

Bassetti, Benedetta. 2008. Orthographic input and second language phonology. In Thorsten Piske & Martha Young-Scholten (eds.), *Input matters in SLA* (Second Language Acquisition), 191–206. Multilingual Matters. doi:10.21832/9781847691118-013.

Best, Catherine T., Gerald W. McRoberts & Elizabeth Goodell. 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America* 109(2), 775–794. doi:10.1121/1.1332378.

Best, Catherine T., Gerald W. McRoberts & Nomathemba M. Sithole. 1988. Examination of perceptual reorganization for nonnative speech contrasts. Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance* 14(3), 345–360. doi:10.1037/0096-1523.14.3.345.

Boersma, Paul & David Weenink. 2016. *Praat. Doing phonetics by computer*. Version 6.0.21. URL:fon.hum.uva.nl/praat.

Caramazza, Alfonso, Grace H. Yeni-Komshian, E.B. Zurif & E. Carbone. 1973. The acquisition of a new phonological contrast. The case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America* 54(2), 421–428. doi:10.1121/1.1913594.

Chao, Yuen Ren. 1968. *A grammar of spoken Chinese*. University of California Press.

Cheng, Chin-Chuan. 1973. *A synchronic phonology of Mandarin Chinese* (Monographs on Linguistic Analysis 4). Mouton de Gruyter. doi:10.1515/9783110866407.

Cho, Taehong & Peter Ladefoged. 1999. Variation and universals in VOT. Evidence from 18 languages. *Journal of Phonetics* 27, 207–229. doi:10.1006/jpho.1999.0094.

Docherty, Gerard J. 1992. *The timing of voicing in British English obstruents* (Netherlands Phonetics Archives 9). Foris. doi:10.1515/9783110872637.

Duanmu, San. 2007. *The phonology of Standard Chinese* (The Phonology of the World's Languages). 2nd ed. Oxford University Press.

Flege, James Emil. 1980. Phonetic approximation in second language acquisition. *Language Learning* 30(1), 117–134. doi:10.1111/j.1467-1770.1980.tb00154.x.

Flege, James Emil. 1981. The phonological basis of foreign accent. A hypothesis. *Teachers of English to Speakers of Other Languages Quarterly* 15(4), 443–455. doi:10.2307/3586485.

Flege, James Emil. 1995. Second language speech learning. Theory, findings, and problems. In Winifred Strange (ed.), *Speech perception and linguistic experience. Issues in cross-language speech research*, 233–277. York Press.

Grønnum, Nina. 1998. *Fonetik og fonologi. Almen og dansk*. Akademisk Forlag.

Grønnum, Nina. 2007. *Rødgrød med fløde. En lille bog om dansk fonetik*. Akademisk Forlag.
Hartman, Lawton M. 1944. The segmental phonemes of the Peiping dialect. *Language* 20(1), 28–42. doi:10.2307/410379.
Klatt, Dennis H. 1975. Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research* 18, 686–706. doi:10.1044/jshr.1804.686.
Kratochvíl, Paul. 1968. *The Chinese language today. Features of an emerging standard*. Hutchinson.
Kuhl, Patricia A., Karen A. Williams, Francisco Lacerda, Kenneth N. Stevens & Björn Lindblom. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255(5044), 606–608. doi:10.1126/science.1736364.
Lai, Yi-hsui. 2009. Asymmetry in Mandarin affricate perception by learners of Mandarin Chinese. *Language and Cognitive Processes* 24(7/8), 1265–1285. doi:10.1080/01690960802113850.
Lisker, Leigh & Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops. Acoustical measurements. *Word* 20, 384–422. doi:10.1080/00437956.1964.11659830.
Liu, Jiang & Allard Jongman. 2013. American Chinese learners' acquisition of L2 Chinese affricates /ts/ and /tsh/. *Proceedings of Meetings on Acoustics* 18. doi:10.1121/1.4798223.
Lively, Scott E. & David B. Pisoni. 1997. On prototypes and phonetic categories. A critical assessment of the Perceptual Magnet Effect in speech perception. *Journal of Experimental Psychology* 23(6), 1665–1679. doi:10.1037/0096-1523.23.6.1665.
Meng, Zimin. 1998. 对外汉语语音教学中使用 ”汉语拼音方案” 的几个问题. In Jinming Chao & Zimin Meng (eds.), 语音研究与对外汉语教学. Beijing Language and Culture University Press.
Mortensen, Johannes & John Tøndering. 2013. The effect of vowel height on voice onset time in stop consonants in CV sequences in spontaneous Danish. *Proceedings of Fonetik 2013. The XXVIth annual phonetics meeting*, 49–52. Linköping University.
Puggaard, Rasmus. 2018. Realizations of /t/ in Jutlandic dialects of Danish. *Linguistica Lettica* 26, 368–393.
Rochet, Bernard L. & Yanmei Fei. 1991. Effect of consonant and vowel context on Mandarin Chinese VOT. Production and perception. *Canadian Acoustics* 19(4), 105–106.
Sloos, Marjoleine, Jie Liang, Xuechao Ne, Rasmus Puggaard Hansen, Mengzhu Yan & Chun Zhang. 2015. Speech rhythm of the Danish-Chinese interlanguage relies on rhyme structure. *Proceedings of the Oriential COCOSDA/CASLRE conference*, 305–310. Jiaotong University.
Sloos, Marjoleine, Jie Liang, Mengzhu Yan & Chun Zhang. 2016. Acquisition of pitch in Chinese by Danish leaners. In Marjoleine Sloos & Jeroen van de Weijer (eds.), *Proceedings of the 2nd workshop on Chinese accents and accented Chinese*, 34–43. Shanghai: Fudan University.
Yan, Mengzhu. 2016. *Perceptuel cue weighting of Chinese /t, t^h, ts, ts^h/ by Danish learners of Chinese and native Chinese listeners*. MA thesis, Aarhus University.
Yang, Bei. 2015. *Perception and production of Mandarin tones by native speakers and L2 learners*. Springer. doi:10.1007/978-3-662-44645-4.
Zhou, Dianfu & Zongji Wu. 1963. 普通话发音圖譜. Shangwu Yinshuguan.