Acoustic Patterns in Hong Kong Cantonese Hesitation Markers: Vowel Quality and Omnisyllabic Tone

Abstract: Hesitation markers (HMs) lie somewhere on the dividing line between linguistic and sub-linguistic, evincing at the same time crosslinguistic commonalities as well as language-specific features (Candea et al. 2005; Dingemanse & Woensdregt 2020). This study seeks to expand on understanding of these lexically peripheral items by analyzing their acoustic properties in Hong Kong Cantonese (HKCT), including vowel quality and F0. Recent work by Dingemanse and Woensdregt (2020) discusses phonetic similarities in HMs across languages, which are constrained by adherence to the phonologies of their respective languages. However, it is unclear how HMs are incorporated into HKCT tonal phonology, which can be characterized as 'omnisyllabic' (Matisoff 1995), every syllable being associated with a lexical tone.

The present study gathers acoustic data (F0, F1-F2, duration) from the PolyU Corpus of Spoken Chinese (http://wongtaksum.no-ip.info:81/corpus.htm) across 10 speakers and 196 HMs, comparing them against 525 lexical items. Results for F1-F2 clustered around mid-front /ɛ/, which is in line both with crosslinguistic trends and HKCT phonology. Results for F0 clustered around the lower end of the pitch range, approaching Tones 3 and 6, which is expected given effort minimization trends. However, establishing a connection with a particular lexical tone was complicated by the similarity of Tones 3 and 6, which could not be statistically distinguished. Cross-speaker analysis showed considerable variation in F0, indicating that context, intonation, or idiolectal variation may play a substantial role. This has implications for our understanding of how peripheral items like hesitation markers are treated in omnisyllabic tone languages like Cantonese.

Key words: Hong Kong Cantonese, hesitation markers, acoustic phonetics, phonology

香港粵語猶豫標記的聲學模式：元音音色及全音節

提要：猶豫標記置身語言學與副語言學的邊界，佔據一個既有跨語言共性又具特定語言特徵的區域 (Candea et al. 2005; Dingemanse & Woensdregt 2020)。本研究旨在透過分析猶豫標記在香港粵語中，包含元音音色及 F0 的聲學特性，擴充對這些語義周邊項目的理解。Dingemanse and Woensdregt (2020) 最近的研究論及猶豫標記有跨語言聲學相似性，亦受制於各語言對自身的音段音韻學的持守。不過，香港粵語作為「全音節語言」 (Matisoff 1995)，每個音節均附上聲調，猶豫標記應如何納入香港粵語的聲調音韻學，這仍不清楚。

本研究從理工大學中文口語語料庫 (PolyU Corpus of Spoken Chinese, http://wongtaksum.no-ip.info:81/corpus.htm) 集結 F0, F1-F2 及時長的聲學資料，橫跨十位語者，共 196 個猶豫標記，與 525 個詞匯作比較分析。F1-F2 的結果圍繞半開前元音 /ɛ/ 叢集，與跨語言趨勢及香港粵語音韻學均能匹配。F0 的結果則叢集於近乎最低的音高，接近調三及調六，合乎最少付出傾向的預想。可是，要與特定字詞聲調建立聯繫仍很複雜，因為調三及調六相似，不能以統計學區分。誇語者分析顯示 F0 有相當變化，提示語境、語調或個人習語變化可能有重大影響。這對理解全音節語言如粵語如何應對猶豫標記等等周邊項目有所提示。

關鍵字：香港粵語、猶豫標記、聲學語音學、音韻學

1. INTRODUCTION

Hesitation Markers ('fillers', 'planners', 'delay markers', 'disfluencies', etc.) and their phonetics/phonology are an understudied aspect of Sinitic linguistics, particularly with varieties other than Mandarin. A good deal of research has focused on the place of these items relative to the general vocabulary in other languages (Candea et al. 2005; Vascilescu et al. 2005; Dingemanse & Woensdregt 2020). Meanwhile, two authoritative studies on Sinitic hesitation markers (HMs), Zhao & Jurafsky (2005) and Yuan et al. (2016), dealing with Mandarin exclusively, do not provide an exhaustive definition of the form HMs take. Liesenfeld (2019), focusing on Malaysian Cantonese, provides valuable insights into the acoustic qualities of minor turn-initial particles, but does not provide a definition for hesitation markers as a class of their own, and in fact finds that they have inconsistent vocalic qualities, which is unexpected to the extent that HMs have conventionalized forms (Clark & Fox Tree 2002; Candea et al. 2005). In addition, while language internal functions and sociolinguistic variables commonly associated with these items in other languages have been addressed extensively (Levelt 1983; Clark and Fox Tree 2002; DeLeeuw 2007; Tottie 2016), ongoing research on these items in Dingemanse (2017) and Dingemanse & Woensdregt (2020) has raised interesting questions about how hesitation markers should be analyzed through a cross-linguistic approach, looking at how their form and function interact to produce observed phonological trends.

In addressing these gaps, the present research looks at how hesitation markers (HMs) are integrated into the phonology of Hong Kong Cantonese (HKCT), primarily in terms of F0 and vowel quality, and in turn how HKCT hesitation markers are integrated into the larger crosslinguistic trends observed for HMs, as documented in Dingemanse (2017) and Dingemanse & Woensdregt (2020). Additionally, the present research hopes to explore and complement previous acoustic research on minor particles in understudied varieties of Sinitic, such as the work of Liesenfeld (2019) on Malaysian Cantonese. To accomplish this, the present study extracts acoustic and categorical data from the PolyU Corpus of Spoken Chinese (hosted at Hong Kong Polytechnic University), a collection of over four hours of naturalistic interviews and conversations in HKCT, and compares it against acoustic data from lexical vocabulary. The ultimate goal is a more rigorous definition of the phonetic/phonological form of HMs in HKCT, as well as a better understanding of the place Cantonese occupies within cross-linguistic HM typology. In addition, further understanding of the way HMs interact with considerations of least-effort and phonological specificity is sought as well.

An additional concern of the present research is the 'omnisyllabic tone' (Matisoff 1995) associated with Cantonese: all syllables are held to assimilate to one of the lexical tones in the language. As words, it is assumed that this should apply to HMs as well; however, it is not clear how or whether this applies to such sub-lexical vocabulary. Evidence from the treatment of sentence final particles (Sybesma & Li 2007; Matthews & Yip 2013: 339) seems to point to a default tone which may serve as a stand-in for HMs as well. This 'omnisyllabic tone' hypothesis for Cantonese will also be explored in the discussion of how HKCT incorporates HMs into its tonal and intonational patterns.

The acoustic data from the PolyU Corpus of Spoken Chinese support the argument that Hong Kong Cantonese HMs are in line with general crosslinguistic trends; vowel quality is consistently mid-front, simultaneously consistent with the limitations of CT phonology and the trend for mid-central vowels in HMs, and F0 is consistently low relative to other vocabulary. However, HM tonal contour cannot be conclusively defined; the data do not support a strong connection between lexical level tones and hesitation marker F0, either in terms of pitch tracks or average, even though a weak connection with the low-level Tone 6 of Cantonese can be established. This presents an apparent exception to the requirement that all HKCT syllables assimilate in production to some lexical tone, although several alternative explanations are entertained.

2. BACKGROUND

Hesitation markers, following characterizations in Clark & Fox Tree (2002) and Tottie (2016), are here defined as discourse markers which signal a delay or pause in speech, and can either follow, precede, or fill the pause. More generally, they function as placeholders as speakers attempt to formulate linguistic output, and therefore have an integral function in easing communication as repairs during lapses in conversation (Fox Tree 2001). A distinction can be made between lexical and non-lexical hesitation markers[1]; the former refers to HMs which have analogues in the general vocabulary (and which therefore have dual or multiple functions), such as *like* and *so* in English, while the latter refers to HMs without such analogues, such as English *uh* and *um*. The distinction should not be understood as related to lexical status, as both sets are conventionalized as discourse markers. The present study focuses solely on non-lexical HMs, since these, lacking analogues with 'arbitrary' forms, are believed to display evidence of crosslinguistic trends in form convergence.

Non-lexical HMs are most often found in two contexts; either at the beginning of an utterance, before anything else has been uttered, or between two utterances in succession. The following examples illustrate these two contexts and the lexical/non-lexical contrast (HMs in brackets); (1a-b) are from Cantonese (Erbaugh 2001; Matthews and Yip 2013: 356), and (1c) is from Mandarin (Erbaugh 2001):

1) a. [誒] … 講個 … [誒] … 有個工人掛 (C01, line 1.1)

[*eh*] *... gong2 go3 ...* [*eh*] … *jau5 go3 gung1-jan4 gwaa3*

HM … say CL … HM … be CL worker hang

'Eh… It's about… eh… there's a laborer hanging…'

b. [咁啊] … 等我諗吓先

[*gam2 aa4*] *... dang2 ngo5 lam2-haa5 sin3*

HM … wait 1.sg think-DEL first

'Well… let me think about it first.'

c. [嗯] … 有一個農人在路邊的一顆樹上...[嗯] (M04, lines 2.1-2)

[*én*] *... yǒu yī gè nóngrén zài lù biān de yī kē shù shàng ...* [*én*]

HM … be one CL farmer be.at road side SP one CL tree above … HM

'Uh… there was a farmer in a tree by the road… uh'

(CL = classifier; DEL = delimitative aspect; SP = subordinative particle)

[1] See Dingemanse, Torreira, & Enfield (2013) for a similar distinction for Other-Initiated Repair strategies (OIRs)

Examples (1a) and (1c) represent common non-lexical HMs in these two languages; in both the first HM occurs at the beginning of the utterance, and the second occurs between two phrases or phrase finally. Example (1b) represents a common lexical HM in Cantonese, *gam2* 'so, well', which in addition to its function as an HM can also be used as an adverb or conjunction. This illustrates the differences between these two classes of HMs: while *gam2* in (1b) is entirely arbitrary in form, drawn from the lexical vocabulary of Cantonese, the HMs in (1a) and (1c), despite having no etymological relationship and coming from two languages separated by many hundreds of miles, resemble each other to a great extent. Henceforth the focus of this study will be relegated solely to non-lexical hesitation markers such as those in (1a) and (1c), focusing primarily on Cantonese, with passing reference to Mandarin for comparison. In addition, the remainder of the piece will focus almost entirely on the *form* of HMs, while discussion of their *function* as discourse moderating items will be solely for illustrative purposes. It is also important to note that while the inherent function of hesitation markers (that is, signaling a pause or delay in speech) appears to contribute to their form similarities across languages (see Dingemanse 2017), focus on immediate communicative context is outside the scope of the present research.

2.1 HM Research

In prevailing theories of grammar leading up to the present, hesitation markers and other language 'disfluencies' were regarded as errors in producing speech (Chomsky 1965), and were therefore considered extrinsic to language. Increased attention was eventually placed on the communicative functions of these items, within the more general study of repair strategies, focusing on their regular usage, form, and positioning (Schegloff, Jefferson, & Sacks 1977; Levelt 1983; Fox, Hayashi, & Jasperson 1996). Research on the lexical status of HMs reached a milestone in Clark & Fox-Tree's (2002) study, which concluded that English *uh* and *um* were in fact words of English, with a conventionalized form, specified functions, and regular distributional properties. Following this, the general assumption here is that non-lexical HMs should be treated as words in the languages in which they appear, treated similarly to discourse markers or interjections in terms of their function and lexical status.

Increased investigation led researchers to note cross-linguistic similarities not just in terms of HM function, but also in terms of form (Kobayashi et al. 1993; Shriberg 2001; Candea et al. 2005; Braun & Rosin 2015), or both (Wieling et al. 2016; Dingemanse & Woensdregt 2020). The general finding across these studies is that non-lexical HMs have considerable phonetic similarities across languages; this finding is of particular relevance to the current research and will be expanded on in the following section.

2.2 Hesitation Markers in Cross-linguistic Perspective

Most languages appear to have some kind of non-lexical hesitation marker which fulfills the communicative functions specified above. In addition, studies on the acoustic qualities of these items have found that there are significant ways in which HMs differ from arbitrary vocabulary, supporting the lexical/non-lexical split. For instance, Shriberg (2001) finds that English average F0 is dramatically lower in HMs relative to the average F0 of the utterance, while Braun & Rosin (2015) find that this is true relative to a German speaker's average pitch. It has also been shown that the decrease in F0 is largely predictable and preserves intonational contours (Shriberg & Lickley 1993; Shriberg 2001). Other studies have also found an exaggerated duration of HMs relative to other vocabulary (Shriberg 2001; Clark & Fox Tree 2002; Candea et al. 2005; Yuan et

al. 2016). For instance, Shriberg finds that English HM vowels can be longer by around 250ms or more. Lowering of pitch and dramatic increase in length seem to reliably distinguish HMs from arbitrary vocabulary; other studies have also shown this for Mandarin (Zhao & Jurafsky 2005; Yuan et al. 2016). A final point to note is Braun & Rosin's (2015) finding that choice of HM can be speaker-specific, indicating that there is at least some variation from speaker to speaker.

Even more interesting is the apparent segmental convergence in form that these items have across a wide sample of languages (Candea et al 2005; Dingemanse & Woensdregt 2020). Most languages seem to have some iteration of a monophthongal, mid-to-low vowel HM which can be extended to mark a pause, accompanied in many cases by a bilabial nasal in the coda. Candea et al. (2005) find a general trend of acoustic convergence in HM vowels across seven languages (French, Spanish, Mandarin Chinese, Am. English, German, Standard Arabic, and Italian). The findings showed that if the F1 and F2 of non-lexical HMs were pooled, all languages in the sample would approximate a vowel quality in the mid to central region of the vowel space. Furthermore, Dingemanse & Woensdregt (2020) find that in five languages (Arabic, Am. English, Japanese, Mandarin, and German) the segments found in 'continuers' (HMs) are more restricted relative to arbitrary vocabulary, the most common segments being [m, n, ɑ, ə]. There are however language specific variations in quality; for instance, while English and Mandarin tend toward a vowel quality in the mid-back region, /ʌ/ and /ɤ/ respectively, Spanish tends toward a mid-front vowel /e/. In addition, there is evidence that speakers can use HM acoustic features to distinguish languages (Vasilescu et al. 2005), indicating that language-specific HM acoustic features are salient. However, in both cases, the preferred vowel is consistently mid and does not vary much from this acoustic space, while the exact choice of vowel seems to be dependent on language-specific phonology. Therefore, while HM vowels display language-specific qualities dependent on phonological inventory, they also display evidence of form convergence across unrelated languages in terms of vowel quality (as well as F0, duration, and choice of accompanying nasal).

HMs display similarities across a variety of acoustic features, with certain predictable differences from language to language. Dingemanse (2017) and Dingemanse & Woensdregt (2020) go on to discuss explanations of how this apparent phonological form convergence has evolved. These studies theorize that the interactional environment of HMs shapes their form in order to fulfill certain criteria for efficient communication. For instance, Dingemanse & Woensdregt (2020) suggest that considerations of ease of production, minimal disruption to communication, and maximal distinctiveness from arbitrary vocabulary are key factors shaping optimal HMs. This can be most clearly observed in how languages across families consistently adopt monosyllabic, mid-to-low vowel HMs with low pitch and primarily bilabial nasals. This theory of HM optimization presupposes that there exists such a thing as a 'natural' hesitation marker, deviations from which we may consider odd. For instance, one might wonder why no language uses a high-front or high-back vowel to consistently precede pauses in speech, and if there were such a language, it might be considered strange or disruptive to communication due to its saliency.

It appears that the factors at work involve the interplay of effort preservation, the phonological system of a language, and the inherent functions of HMs. Since HMs serve to fill pauses in speech and can be quite lengthy (>1000ms in some cases), it is reasonable that they should consist primarily of nasals and vowels, sounds that are amenable to extension over a longer period of time with relatively few articulatory gestures. At rest, the vocal tract is in the position required to produce a bilabial nasal, and mid-vowels involve the least variation from schwa, which would require less effort to produce relative to high vowels. The average low pitch these items tend to evince is also in keeping with the preference for low salience and low effort. At the same time, however, languages do not unilaterally converge on schwa or even a bilabial nasal without exception; rather, the closest alternative is chosen from the set of sounds (segmental or suprasegmental)

available in the language in question, which fulfills the above criteria as closely as possible. There therefore seems to be some kind of blueprint for an optimal hesitation marker, one which fills the criteria mentioned in Dingemanse & Woensdregt (2020).

Three views on the phonological form of HMs relative to arbitrary vocabulary can be extrapolated, moving from least to most restrictions on form. If HMs are assumed to be somewhat random in their form, with no consistent quality, then it stands to reason that they should be more variable than general or purely arbitrary vocabulary, in line with claims about their non- or sub-linguistic status as 'disfluencies' (Chomsky 1965; Levelt 1983; Shriberg 2001; McDougall & Duckworth 2017); this is in line with findings by Liesenfeld (2019) for Malay Cantonese. In assuming that HMs are solely dependent on least-effort considerations for their phonetic output, one would assume that this would conventionalize them to a considerable extent, perhaps reducing them to schwa or some other 'neutral vowel quality' (Wieling et al. 2016), irrespective of the language in which they occur. Both of these assumptions are too strong, and fail to capture the full range of possibilities for HM form.

If HMs are conventionalized in form due to interactional environment (Dingemanse, Torreira, & Enfield 2013; Dingemanse 2017; Dingemanse & Woensdregt 2020), that is, the combined pressures of easing articulation, minimizing disruption to conversational flow, and maximizing distinctiveness from other vocabulary, but also subject to language specific phonological patterns (Candea et al. 2005; Vasilescu et al. 2005), it is possible they have a conventionalized form, and vary acoustically in ways comparable to arbitrary vocabulary, although their form remains somewhat predictable across languages (e.g., bilabial nasals, monophthongal central vowels, etc). Therefore, are HMs more, less, or equally as variable as other vocabulary? The descriptive qualities of HMs in Hong Kong Cantonese, and whether and how these criteria are met, will be addressed in the next section.

2.3 Hesitation Markers in Sinitic: Mandarin and Cantonese

Sinitic languages, like all languages, have lexical and non-lexical hesitation markers with cross-linguistically common functions. Most research on Sinitic HMs is restricted to Mandarin and tends to deal with usage and sociological factors. For instance, Zhao & Jurafsky (2005) discuss the placement of HMs within phrases, their duration, and average pitch, as well as the quantifiable differences between lexical and non-lexical HMs, while Yuan et al. (2016) discuss these as well as sociolinguistic trends in the usage of different HM forms by gender and education. Studies such as Strassel et al. (2005), identify four characters in the spoken corpora used as stand-ins for HMs: 嗯 *en* 唔 *um/em* 呃 *eh* 啊 *ah*. Of these, Yuan et al. (2016) identify 嗯 *en* and 呃 *eh* as the most common.

These studies provide clues for a phonological definition of what constitutes a Mandarin non-lexical HM, and seem to agree that the most common come with V and VN syllabic structure. While none of the sources discuss vowel quality in very much detail, it can be inferred from the discussion that the vowels of the two most common Mandarin HMs are most likely mid-back [ɤ] to mid-central [ə], these two vowels being allophones of a single phoneme /ɤ/ (the latter allophone found preceding nasal codas). In terms of tone, Zhao & Jurafsky (2005) mention that Mandarin HMs have particularly low pitch relative to the phrase, and are produced with the Mandarin 'neutral tone', where the F0 contour of the syllable is supplied by context (Li & Thompson 1981; Wiedenhof 2015). None of these facts are surprising given what is known about common trends in HM form; what is surprising is the choice of nasal, which is more often [n] than [m], although this could be ascribed to language specific properties of Mandarin, such as the fact that it has no

syllable final or syllabic [m]. That said, the bilabial nasal does occur as a possible HM, in either VN or N sequences.

The most common HMs in HKCanCor (Kwong and Wong 2015), the corpus of spoken Cantonese, are 誒/欸, and 嗯, commonly romanized as *e6*/*e4* or *em6*/*m6* in Jyutping. These two HMs make up more than 35% of all items classified as 'interjections' in the corpus (around 3000 interjections in total)[2]. In fact, the most common interjection overall is 誒 *e6*/*e4*, making up close to one third of all interjections in the corpus. Figure (1), created using Jackson Lee's PyCantonese library (Lee 2015), summarizes the findings.

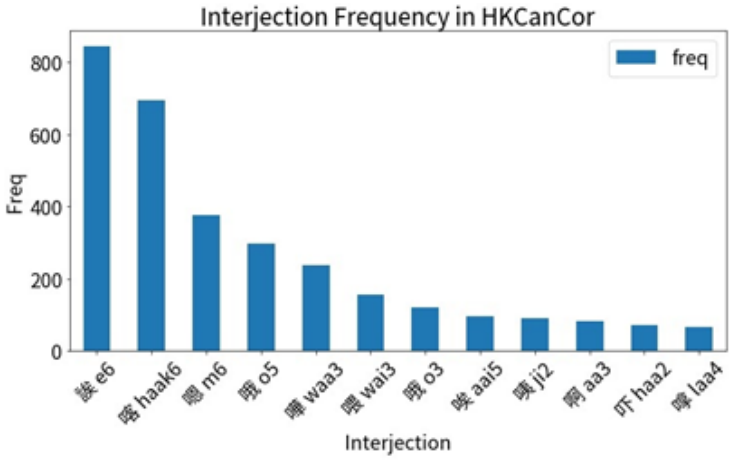

Figure 1. Interjections in HKCanCor by frequency; first and third are common HMs

These items in Cantonese are known to be quite varied in form, and it is quite possible that items represented by a single character may cover a plurality of HMs. The syllabic form of these two items can be V, VN, and N, with the most common vowel, based on the romanizations, approximating a mid-front [ɛ]. The tone, once again based on the romanizations, is by default assumed to be the low-level Tone 6 or low-falling Tone 4 (see §2.4 for details), which is consistent with the low pitch associated with HMs cross-linguistically. However, as following sections will show, there appears to be more variation in terms of pitch than would be the case for lexical tones in purely arbitrary vocabulary. For this reason, what tonal contour may be the default for these items, if any, will be further explored in later sections.

Research on similar items in other varieties of Cantonese turns up different conclusions. For instance, Liesenfeld (2019) finds that Malaysian Cantonese 'turn-initial management tokens' (of which HMs are a subclass) have no consistent vowel quality, ranging from [ɐ], [ɔ], [œ], [ø], [o], and even [u]. This is surprising for several reasons. HKCT has 11 phonemic vowels (Bauer & Matthews 2017); these include six mid-central and central vowels [e ɛ ø œ ɐ a]. Of these, only [ɛ], [a], and [œ] occur in open CV syllables: *ze6* [tsɛ:˩] 'to thank', *caa4* [tsʰa:˩] 'tea', *hoe1* [hœ:˥] 'boot'. It is therefore surprising that [ɐ] is found to be a common vocalic quality for HMs, as it does not occur in open syllables, whereas [ɛ] and [a] are found to be far less common.

[2]The character 嗯 *em6/m6* can also be used in affirmations (Liesenfeld 2019) and is therefore not exclusively an HM in the corpus; HKCanCor does not distinguish between these two uses. However, 誒 *e6* alone accounts for almost one third of all interjections in the corpus, therefore even if the former item is excluded, HMs are clearly quite common interjections.

Liesenfeld additionally holds that while pitch tends toward low level and falling F0, which agrees with assumptions about HKCT hesitation markers thus far, the study does not find it to be a 'distinctive feature'. These two conclusions about vocalic quality and F0 are surprising given that languages can be identified by their HMs (Vasilescu et al. 2005); if the phonetic quality of HMs were not at least somewhat consistent, it seems unlikely this would be the case. At the same time, Liesenfeld's study excludes the most common HM in CT, namely '誒', and therefore the results are bound to be inconsistent with those above.

Although Cantonese and Mandarin are two closely related Sinitic languages, their HMs are in fact quite different. While there are some general similarities (low pitch, mid vowels), these are more readily ascribed to common trends for HMs across languages, rather than to their close genetic affinity. Exploring how these language-specific differences interact with cross-linguistic trends in order to produce optimal HMs in Hong Kong Cantonese, is therefore the goal of the present study. The following study attempts to acoustically define HMs by analyzing samples of naturalistic spoken HKCT, as well as evaluate the competing claims of Liesenfeld (2019) and Dingemanse & Woesndregt (2020) for HKCT.

2.4 Omnisyllabic Tone in HK Cantonese

As a tonal language, Cantonese HM tonal contour and its relationship to phonological tones is also of interest to the present study. Matisoff (1995) discusses the concept of 'omnisyllabic' tone, which refers to tonal systems in which all syllables in a language assimilate to some lexical tone. Cantonese seems to adopt this pattern, as all lexical syllables must bear one of the lexical tones (Matthews & Yip 2013). HKCT has a tonal inventory of 6 lexical tones (Bauer & Matthews 2017). These include the following: high-level Tone 1, high-rising Tone 2, mid-level Tone 3, low-falling Tone 4, low-rising Tone 5, low-level Tone 6; the present research focuses on the three level tones of Cantonese: Tones 1, 3, and 6. Consider the following minimal pairs:

si1 [si˥] 'silk' *si4* [si˩] 'time'

si2 [si˧˥] 'history' *si5* [si˨] 'market'

si3 [si˦] 'meaning' *si6* [si˩] 'thing, matter'

The Omnisyllabic Tone Hypothesis assumes that all syllables of Cantonese are incorporated into this system, regardless of their lexical status. Not all Sinitic languages function like this; notably, Standard Mandarin has a contextually-defined 'neutral' tone that gets its F0 contour from surrounding tones. Cantonese has no equivalent to the neutral tone; peripheral or sub-lexical material usually appear in one of the mid to low tones. For instance, sentence-final particles (SFPs) with 'neutral' meanings seem to gravitate towards a neutral or default mid-level Tone 3 (Sybesma & Li 2007; Matthews & Yip 2013: 339); compare the following SFP tonal triplets, from Tsang (2020):

laa3 'change of state (neutral)'	*aa3* 'question (neutral)'	*lo3* 'suggestion (neutral)
laa1 'request (playful)'	*aa1* 'suggestion (playful)'	*lo1* 'suggestion (playful)'
laa4 'change of state (query)'	*aa4* 'question (surprise)'	*lo4* 'suggestion (surprise)'

What is important to note about these triplets is that, regardless of the primary function of the SFP, different tonal contours convey different shades of pragmatic meaning. Consistently, the high-level Tone 1 and the low-falling Tone 4 are associated with playfulness and surprising information, respectively; the mid-level Tone 3, on the other hand, is not associated with any additional layers of meaning. Similarities with

intonation patterns in other languages has led to the claim that these items might be inherently toneless, and instead get their tone through intonation (Cheung 1972), which is then assimilated to one of the lexical tones. Others (Law 1990) have claimed that the differences in meaning arise through 'tonal particles': segmentless floating tones. In either approach, Tone 3 is a 'default' in CT sentence-final particles.

Additional evidence for the omnisyllabicity of HKCT tone comes from foreign-loan phonology, where a default high-low or high-mid tonal pattern is adopted by speakers in order to assimilate English stress to HKCT lexical tone; for instance, unstressed English syllables in bisyllabic words take the low-level Tone 6 if final (Kiu 1977)[3], e.g., *latte* [ˈlaɭ.tʰɛɻ], or the mid-level Tone 3 if initial, e.g., *guitar* [kitɬ.tʰaɭ] (Silverman 1992). Silverman (1992) also notes that when an illegal onset cluster is divided into two syllables, the low-level tone is found, e.g., *stamp* [siɻ.ˈtamɭ]. It is possible that HMs are in some way similar to borrowed unstressed syllables, requiring a default in a low or mid-level tone.

If HMs are incorporated into the CT tonal system as lexical vocabulary, it is predicted that they will default to one of two lexical tones: the mid-level Tone 3 (like SFPs) or the low-level Tone 6 (like foreign-loan vocabulary). If on the other hand HMs are not incorporated into the CT tonal system, it is predicted that they will obtain their F0 contours from intonation or context. An issue that naturally arises is how to distinguish these two manipulations of F0 in HKCT (intonation and lexical tone), which influence each other and overlap to some extent (Lee 2004; Ma, Ciocca, & Whitehill 2011). However, the majority of the effects of intonation can only be discerned in HKCT in utterance final position (Lee 2004; Matthews & Yip 2013), whereas hesitations can occur in a variety of positions in the utterance (see §4.1). In final position, linear models were run in order to check for effects on F0 (see §4.3).

The final question this study seeks to address is the degree to which HM tonal contours assimilate to the lexical tones of HKCT, in terms of their acoustic quality. If HMs are words of Cantonese, they should assimilate to some lexical tone; which of the tones HMs assimilate to will depend on their lexical status within HKCT. Three hypotheses are entertained: either HMs will be treated as SFP-like elements which default to Tone 3, as foreign loan vocabulary which default to Tone 6, or as non-lexical and dependent for their F0 contour on intonation. Evidence supporting the 'omnisyllabic tone' hypothesis, where all syllables assimilate to the tonal system regardless of lexical status, or for a special status for HMs within HKCT phonology, will be discussed in §4.2.

3. METHODOLOGY

Data was gathered form the PolyU Corpus of Spoken Chinese (http://wongtaksum.no-ip.info:81/corpus.htm), which provides naturalistic, conversational interviews in Hong Kong Cantonese conducted between 2010 and 2015. The interviews are transcribed in written Chinese with the original recordings provided. In total, 10 adult speakers (6 female, one of them being the interviewer) and 4h5'35" hours of recordings are included in the corpus. Data was gathered from one recording per speaker, which varied in length from 2'09" to 11'14". The corpus does not provide demographic data on the participants; Chor (2014), who makes use of the corpus, mentions that the respondents are university students and the interviewers are research assistants, aged 18-23.

Table 1. Breakdown of items in F0 analysis

	HM	**Lex1**	**Lex3**	**Lex6**	**All**

[3] Or a rising tone, as Silverman (1992) notes. It is also worth noting that the same source lists monosyllabic borrowings from English as taking a default high-level Tone 1, e.g., *card* [kʰatɭ].

Female	156	117	136	118	527
Male	40	37	37	38	152
Total	196	154	173	156	679

In total, 695 items were included: 196 hesitation markers and 499 lexical items. The most common syllabic format for HMs is V (n = 170), with VN the second most common (n = 20). Items analyzed for F0 and items analyzed for F1-2 overlapped for the most part; items were not included in the former if F0 was too low to register (n = 16), and were not included in the latter if they were composed of syllabic nasals or diphthongs. Within the lexical items analyzed for F0, three tonal contours were collected: high-level Tone 1 (n = 154), mid-level Tone 3 (n = 173), and low-level Tone 6 (n = 156; see Table 1). Vowels extracted from the lexical set included [a] (n = 65), [e] (n = 121), [i] (n = 76), and [o] (n = 70). Most of the lexical items are common grammatical particles or sentence final particles in Cantonese, chosen due to their tonal contour, syllable structure (V, CV, or CVV), and frequency in HKCT speech (see figure 2). The six most common are the attributive/possessive particle 嘅 *ge3*, the general classifier 個 *go3*, the topic-marking particle 呢 *le1*, the copular verb 係 *hai6*, the coordinating conjunction 就 *zau6* 'then', and the plural classifier, comparative marker, and adverb 啲 *di1* 'some, a few', which together account for over half the sample (n = 260).

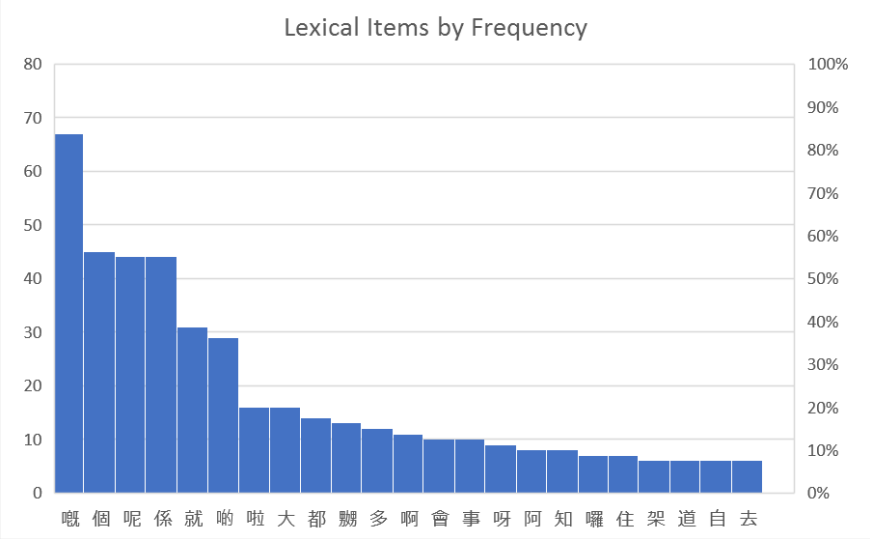

Figure 2. Items in the lexical sample ranked by frequency (23 most common)

Lexical items were restricted in form to V, CV, and CVV syllables for ease of comparability with HMs, and because these syllable types are predicted to carry clearer target tones compared to closed syllables. 2 HMs were excluded from F0 analysis due to their F0 being too low to measure reliably (< 50Hz). Each item was analyzed for duration, context, tonal contour (Level, Falling, Rising, etc), and fundamental frequency, both as an average and at regular points across the syllable. Items with monophthongal vowels were analyzed for F1 and F2 as well (diphthongs excluded). All acoustic analysis was conducted in Praat (Boersma and Weenink 2021), and all statistical analysis was conducted using R (R Core Team 2021).

Average F0 for HMs and lexical items (in Hertz) was taken across the periodic portion of the syllable, excluding sudden spikes or falls caused by errors (e.g., the quality or circumstances of the recording) or by syllable initial consonants (e.g., fricatives, affricates). For lexical items, which were restricted to CV and CVV syllables, F0 was measured only on vowels and diphthongs, since including the consonant might

affect the reading. Previous research (Khouw & Ciocca 2007) has shown that the level tones under analysis (Tones 1-3-6) can be distinguished through average F0 alone. When calculating the average, the effects of tonal context were preserved; for instance, if an HM was preceded by a T2 which demonstrably raised its F0 at the beginning of the syllable, this was not corrected and was included in the average F0 for that HM. This was done in an effort to determine potential contextual effects on average F0. In addition, position in the utterance (Isolated, Initial, Final, and Medial) was also coded for HMs in order to determine whether there were any effects on average F0.

In addition to an average, F0 at regular points across the syllable was also measured, in order to compare pitch tracks across HMs and lexical tones. Using a modified Praat script written by Dr. Jonathan Havenhill, each syllable was divided into 20 equal timepoints, at each of which a measurement of F0 was taken. Certain items were excluded from this analysis due to irregularities in their contours, such as abrupt rises or falls caused by creaky voice; in total 605 items were included for pitch tracking. In the analysis, values below 15% and above 85% of the length of the syllable were excluded, in order to avoid contextual effects.

F1 and F2 was taken as an average over the steady-state portion of the vowel. In the lexical set, formants were extracted for cardinal monophthongs, including [a], [ɛ], [ɔ], [i], in order to define the acoustic space in which HKCT vowels occur; the high-back vowel [u] was not included, however, since this sound was only found twice as a monophthong in the recorded data[4]. The other monophthongs of Cantonese, namely [œ] and [y], were also not analyzed since they are predicted to be primarily differentiated through F3; HMs were not found to be rounded in the sample. In total, 525 vowels were included in the analysis (HM = 193). The formant values were pooled and normalized using the NORM Vowel Normalization Suite (Thomas and Kendall 2007). Finally, duration was measured starting from the beginning to the end of activity on the waveform, in milliseconds (ms).

If HKCT hesitation markers are viewed in a naturalistic communicative environment, two outcomes are predicted: 1) HM vowels will assimilate in production to mid-front [ɛ], and 2) the pitch of HMs will assimilate to low-level Tone 6. This is in accordance with cross-linguistic trends for HMs to occur with mid-vowels and low or variable F0 relative to other vocabulary, predicted by Candea et al. (2005) and Dingemanse & Woensdregt (2020). The choice of [ɛ] is due to the vocalic inventory of Cantonese not including any central vowels in open syllables, while Tone 6 is predicted as it is the level tone with the lowest F0 in HKCT[5]. Alternatively, HMs might have a more central vowel quality which is not found in the segmental inventory, perhaps a schwa or other 'neutral' vowel quality (Wieling et al. 2016), and may default to another tone (such as Tone 3; see above) or to no tone in particular, with the F0 contour provided by context or intonation. This would offer support to Liesenfeld's (2019) findings on the lack of consistency of HM vowels and lack of distinctiveness for HM pitch in Malaysian Cantonese. As the following section will show, portions of both of these predictions seem to hold; this will be elaborated on further in the discussion in §5.

[4] In fact, [u] is uncommon in Cantonese CV syllables generally; in CV sequences, it only occurs with the labiodental [f] and CantoDict (Sheik 2021) lists only 45 characters pronounced <*fu*> (using Jyutping orthography). The two instances identified in the corpus were non-standard pronunciations of 都 *dou1* as *du1*.

[5] Technically the lowest pitch belongs to Tone 4, the 'low-falling' tone (Matthews & Yip 2013). However, this tone is left out of the analysis here for several reasons. Firstly, it is shorter and more abrupt than the level tones (Tones 1, 3, and 6) and therefore cannot extend to cover the length of an HM, which can be up to 1s. Secondly, this tone is often associated with creaky voice (Mok, Zuo, & Wong 2013), which causes problems with readings of F0 in Praat. Finally, there is little evidence that this tone can be a default in peripheral items, as it is seen to add an additional dimension to the meaning (see §2.4).

4. RESULTS

This section covers results of the acoustic analysis of the hesitation markers and lexical vocabulary in the corpus. A general overview of the results is covered below in §4.1, followed by results for F1-F2 in §4.2, and F0 in §4.3.

4.1 General Overview

Hesitation markers in the sample, despite displaying some degree of variation from speaker to speaker (see Appendix), possess certain consistent features. For the most part HMs tended to be fairly lengthy (avg. 380ms), to have low F0 relative to other vocabulary (with one notable exception), and to have mid-front towards central vowel quality (see §4.2 and §4.3 for detail). There were predictable differences between female and male values for F0 and F1-2, while duration is consistent (see Table 2).

Table 2. Average values for hesitation markers (SD in parenthesis)

	Female	**Male**
F0 (Hz)	200.1 (33.5)	92.2 (22.2)
F1 (Hz)	709.6 (98.3)	607.7 (52.2)
F2 (Hz)	2043.9 (260.9)	1828.2 (221.6)
Duration (ms)	382.3 (194.6)	377.1 (187.3)

HM-F0 is significantly lower for female ([$t = -6.9837$, $df = 240$, $p < 0.0001$]) and male ([$t = -4.17$, $df = 34.793$, $p = 0.000192$]) speakers when compared to the overall F0 of the phrases in which they appear (228.949 for females, 129.611 for males). This is analogous to what is seen in Shriberg (2001) and Braun & Rosin (2015), and shows that HMs are consistently lower in F0 than other vocabulary.

Average duration was around 380ms with a range between 100-1000ms; compare this with lexical syllables, which averaged 262ms with a range of 51-1150ms (see Table I in Appendix). The difference in means between the two was highly significant ([$t = 8.0229$, $df = 300.08$, $p < 0.0001$]), supporting claims that HMs are substantially longer than lexical vocabulary (Candea et al. 2005; Yuan et al. 2016). The density distribution by item can be observed in Figure 3 (created with R).

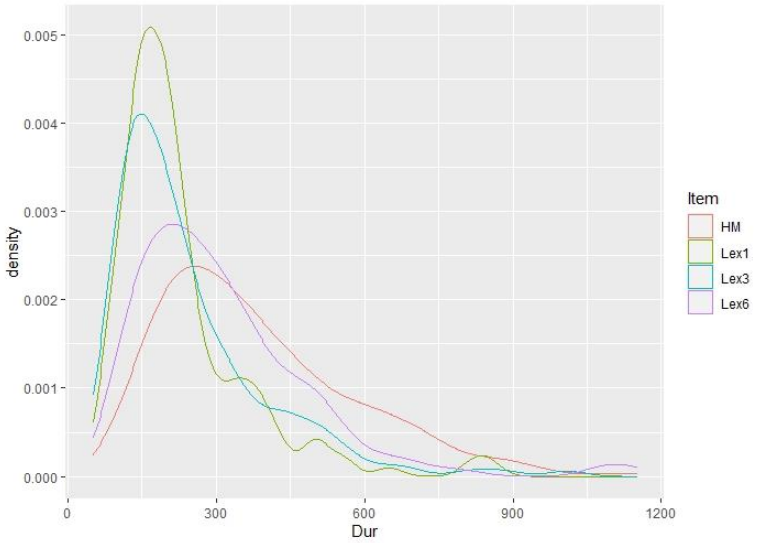

Figure 3. Density for Duration (ms) by Item (all speakers)

Differences in duration between the lexical items in Figure (3) can be attributed to word choice. For instance, the most common Lex6 item was *zau6* 'then' (see Figure 2 in §3), which is also a common lexical HM. Dual use as a lexical word and an HM increases length for Lex6 dramatically, approximating non-lexical HM duration. Meanwhile, the most common Lex1/Lex3 items were *di1* and *ge3*, which are common grammatical particles and therefore often shortened. Regardless, non-lexical HMs are still the longest items in the sample.

HM F1-F2 overlaps considerably with lexical [ɛ], and can be quantifiably differentiated from [a i ɔ] in the sample (see §4.2). F0 for HMs can be reliably distinguished from lexical Tone 1, whose value in Hz is considerably higher (see Table 3). HM pitch is comparable (both in terms of average and contour) to Tone 3 and Tone 6, but it is not possible to determine which of the two they approximate. Interestingly, Tones 3-6 cannot be reliably distinguished; see §4.3 for details.

Table 3. Average F0 (Hz) for lexical tones (SD in parenthesis)

	Female	**Male**
Tone 1 (High Level)	274 (64.7)	148.3 (34.7)
Tone 3 (Mid Level)	207.3 (36.7)	109.2 (19.8)
Tone 6 (Low Level)	200.8 (26.7)	102.8 (18)

The difference in Hz between T3 and T6 is rather small, and may be due to ongoing sound change in HKCT, where these two tones appear to be merging (Mok, Zuo & Wong 2013). In terms of position in the utterance, HMs were coded as one of the following: phrase initial (n = 69), phrase final (n = 46), phrase medial (32), and isolated (n = 46) (this last category was defined as silence both preceding and following the HM). The means for F0 by position can be found in Table V in the appendix; differences between categories did not appear substantial, but an LME was run to determine if the differences were significant (§4.3). Further discussion of the acoustic data will take place in following sections.

4.2 First and Second Formants

Pooling together the F1 and F2 values for each vowel in the sample (n = 526) across all speakers, the un-normalized vowel space looks like what we see in Figure 4 below[6]. Already we can see somewhat of an overlap between the vowels in blue and red, which represent the front mid vowel [ɛ] and the vowel found in HMs (marked with an 'x', in red), respectively.

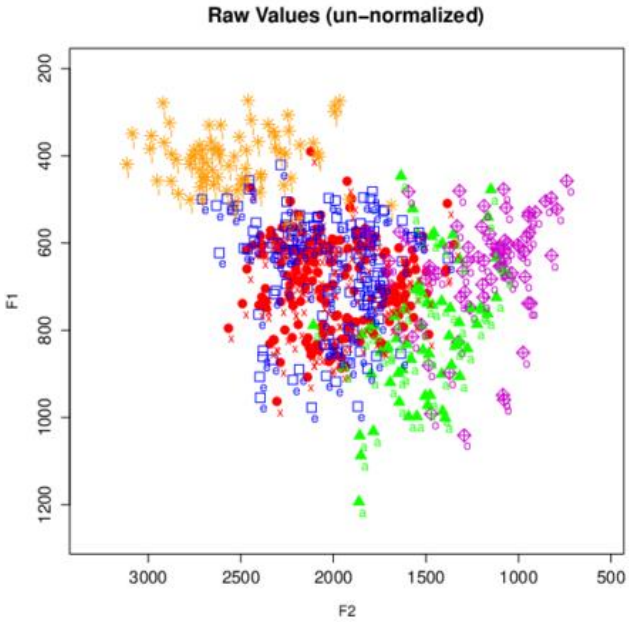

Figure 4. Raw F1-2 values for all vowels in the sample

Relative to the other vowels, HMs seem to cluster around the mid-front to mid-central region, which is unsurprising given the trends outlined in §2.2, but is surprising given HKCT's lack of a mid-central vowel. There also seems to be a rather high degree of variation among all vowels (consider the variation for [a], for instance). Raw formant values are particularly sensitive to cross-speaker physiological variation. Therefore, a vowel extrinsic Lobanov normalization (1971) is conducted, converting the measurements in Hz to standard values.

Normalization results for F1 and F2 across all speakers were consistent with expectations for prediction (1): the vowel space occupied is overwhelmingly mid-front, with considerable overlap with HKCT [ɛ] from lexical vocabulary. Unlike what we see in Figure 4, there is no mid-central clustering, which appears to have been due to differences across speakers. This outcome is summarized in the chart below, which shows Lobanov normalized F1 and F2 values for each vowel token (HMs in red, with an 'x').

[6] All figures in §4.2 were created with NORM (Thomas & Kendall 2007)

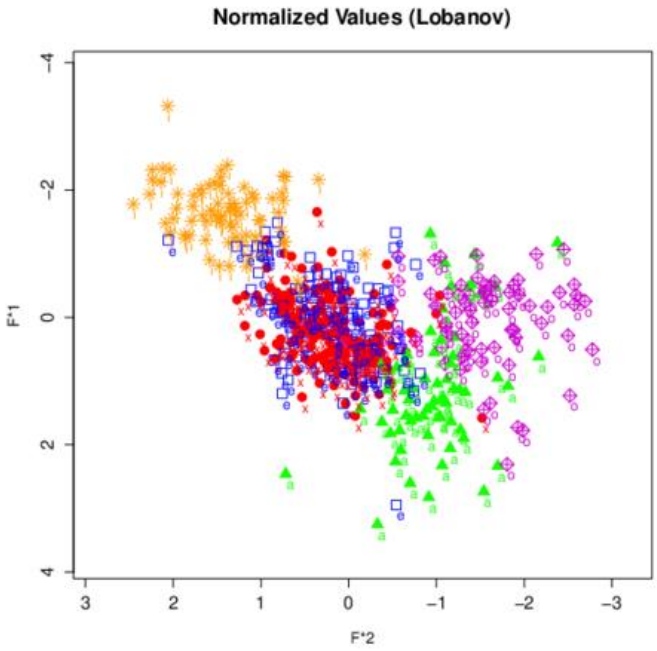

Figure 5. Lobanov normalized F1-2 values for each vowel token

Figure (5) shows that there is almost complete overlap of CT [ɛ] (in blue) with HM vowels, minus a few outliers. It should be noted, however, that [ɛ] is also variable, with some tending towards [i] or even [a]. Importantly, however, this normalization process (Lobanov 1971) treats all HMs as belonging to the same vowel category, which may not be the case. For instance, some may be lower, tending towards [a], and others may be more centralized; the normalization would then end up ironing out some crucial differences in the vowel qualities. In addition, the normalization procedure works best if all vowels (cardinal vowels in particular) from the target language are included; as mentioned in §3, [u] was excluded for lack of tokens, while [y] and [œ] were not included since F3 does not figure in the analysis. Still, however, we can see that the normalization gives a clearer picture of the overlap between the two vowels in red and blue, which is in line with our predictions of a mid-front quality. Figure (6) gives a better idea of the degree of overlap, by plotting the un-normalized vowel means.

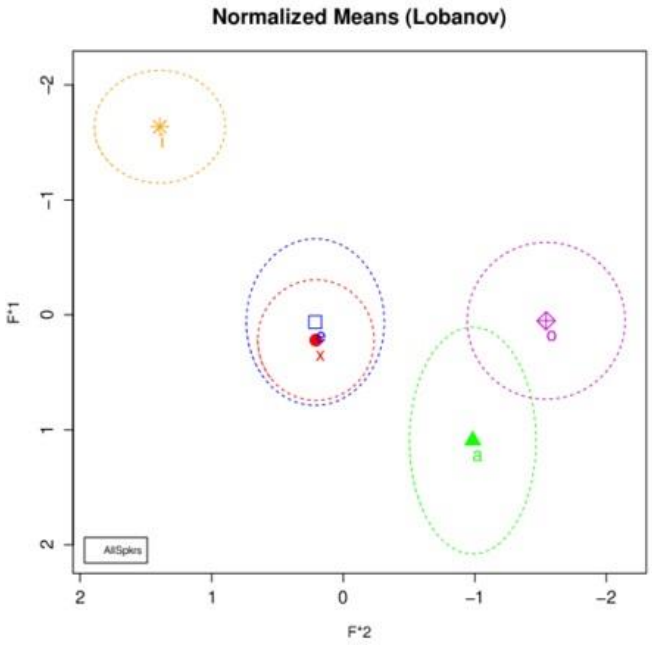

Figure 6. Mean vowel space (sd = 1) for normalized vowels (HMs in red with an 'x')

As in Figure 5, once again there is considerable overlap between HMs and [ɛ]; if normalized vowels are plotted, the degree of overlap is even greater. Lexical vowels in the sample appear to have larger standard deviations than HMs in the sample, indicating that they are more variable in their formants; it is unclear why this might be the case. It appears that HM vowels tend somewhat more to the central region than [ɛ], with [ɛ] tending to be higher. However, the degree of overlap is such that we must conclude that on average in the sample, HM vowels have the same quality as mid-front [ɛ].

4.3 Fundamental Frequency

Results for F0 were not as consistent as those for F1-F2, but are also generally supported by the predictions, with a low average across speakers. However, the exact tonal contour could not be precisely identified, either by average F0 or by pitch track. Figure (7)[7] summarizes the overarching trend, and compares average F0 of HMs and the three lexical level tones in the sample, ignoring context, contour, and syllable duration.

[7] This and all other figures in this section were created with R (R Core Team 2020)

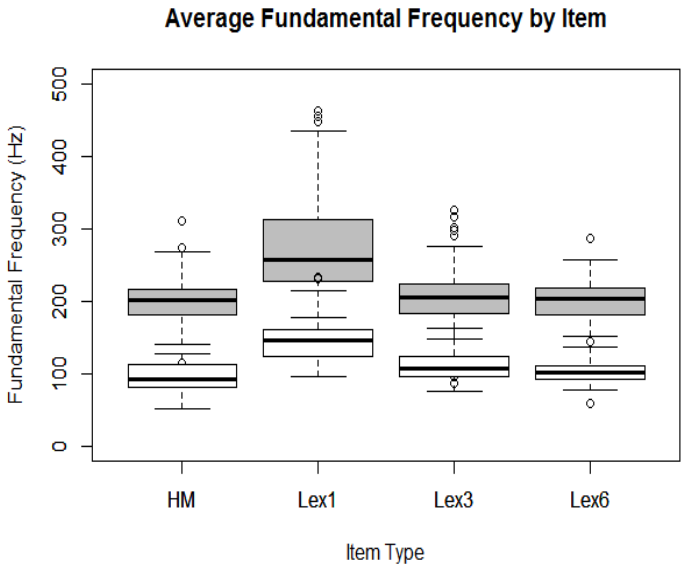

Figure 7. Boxplot for average F0 by Item Type, Gray = Female, White = Male

Generally speaking, Lex1 (the high level tone of HKCT) is the highest in terms of F0 (avg. 274-148Hz), and is therefore easily distinguishable from HM (200-92Hz; $p < 0.0001$). While average HM F0 is low compared to all other items, these items are not usually found to be below Lex6 (with one male speaker providing an exception), and are generally within the normal pitch range of a speaker. There is little appreciable difference between HM and Lex3-Lex6; in fact, Lex3 and Lex6 are statistically indistinguishable in terms of mean F0 (207-109Hz and 201-103Hz; $p = 0.15, 0.24$), and are not clearly distinguished in terms of pitch tracks (see Figure 8). This is not surprising considering ongoing sound change in HKCT, in which these two tones are in the process of merging (Mok, Zuo, & Wong 2013). Figure 8 shows this general trend in terms of F0 contour (focusing on the Female speakers); the high degree of similarity between HM-Lex3-Lex6 can be more clearly appreciated.

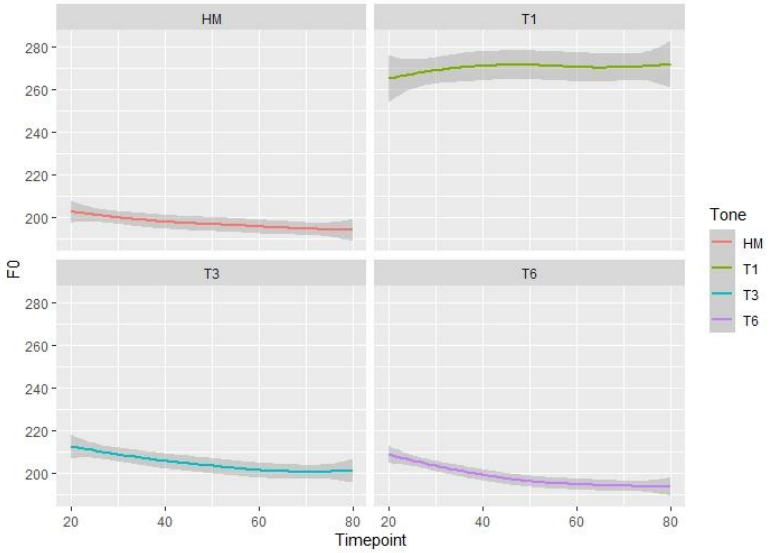

Figure 8. Average Pitch Tracks (female)

Within individual speakers, the same overall trend is observed, with HM being comparable to Lex3 and Lex6, and with Lex1 much higher than all three (see Table II, Figures I and II in Appendix). There are three exceptions to this trend: speakers m1, f2, and f4. Speaker m1 has an average HM F0 much lower (61Hz) than any of the other lexical tones (often accompanied by creak, causing measurement difficulties), and the mean is therefore not comparable. Speaker f2 shows an exaggerated contextual effect on HM F0, caused by a high incidence (0.47) of HMs occurring in the sequence [*gam2 eh*…] 'So uh…', which raises the pitch substantially given a preceding high rising tone leading into the HM; this most likely relates to her role as interviewer in the recordings. Finally, only 4 HMs were identified for f4, leading to higher than normal variation. These outliers were removed in subsequent analyses.

Results of a linear mixed-effects model in the appendix (Table VI), looking at F0 by position with speakers as a random effect, show a slightly significant decrease of F0 in Final position [-12.54673, SE = 4.989950, t = -2.514400, p = 0.0128*]. It is worth pointing out that this effect disappears if separate linear models are conducted on female and male speakers. Average F0 for the Initial and Medial positions is higher relative to the Final and Isolated positions (Table V), which could also be attributed to the effects of context/intonation. Possible influencing factors for these effects include expressions of uncertainty in initial position, question intonation at the end of an utterance, falling intonation at the end of an utterance, etc. However, none of these positions are found to have a significant effect on F0 except for the Final position.

If the mean values are compared across HMs and Lexical Tones, there is a highly significant ($p < 0.0001$) difference in means between HM and Lex1, but an inconclusive difference between HM and Lex3-Lex6; given t-tests of the results (see Table III in Appendix), it is not possible to reject the null hypothesis that these three variables have the same average F0. However, if we remove outliers, namely m1, f2, and f4, we get closer to the predicted result: the difference in means between HM and Lex3 becomes significant ($p < 0.05$). This difference is not very substantial, however, and the difference between Lex3 and Lex6 remains insignificant, which complicates interpretation. This same result holds when the effect of Item Type on F0 is analyzed through a linear mixed-effects model, the results of which can be observed in the Appendix (Table VII).

5. DISCUSSION

The acoustic features of Hong Kong Cantonese HMs are largely in line with cross-linguistic trends in terms of duration (Shriberg 2001; Yuan et al. 2016), F0 (Shriberg 2001; Braun & Rosin 2015), and vowel quality (Candea et al. 2005), and are less variable than expected given their sub-lexical status. The assumption that HMs lack a consistent F0 contour, or have no consistent vocalic quality, does not hold. It was shown that HKCT [ɛ] and HMs in the sample overlap to a considerable degree, and that HMs are consistently in the lower pitch ranges (even lower than the average pitch of the phrases in which they appear, see §4.1), approaching Tone 6. This is again unsurprising as the usual written form for the majority of these items is 欸 *e6* or 嗯 *em6*. This can be attributed to the fact that HMs have a specific form in the languages in which they appear, agreeing with findings on vowel quality in Candea et al. (2005). The degree of indeterminacy in distinguishing between Tones 3 and 6 means that the Omnisyllabic Tone Hypothesis cannot be established conclusively, although evidence suggests that HMs are most closely related to one of these two.

In the two sections that follow, two major areas will be discussed: results and their connection to crosslinguistic trends, with particular reference made to findings and claims in Candea et al. (2005), Dingemanse & Woensdregt (2020) and Liesenfeld (2019), and Matisoff (1995).

5.1 The Variability and Regularity of HMs

The data above shows that HMs in HKCT, as in other languages, have a conventionalized form, which is in some ways specific to this language. We can see this most clearly from the results for F1 and F2, where the vast majority of tokens converge on mid-front [ɛ]. Rather than simply reducing to schwa[8] or [ɐ], the choice of vocalic quality is licensed by the phonology of HKCT, where [ɛ] can occur elongated in open syllables. This is not to say that there aren't variant pronunciations; a common alternative to [ɛ] is low-front [a], although in the sample it is much less common (see §4.2). This regularity contradicts the findings in Liesenfeld (2019), where a wide variety of vocalic qualities, including high-back [u], where more common than the mid-front quality.

In this sense, having a conventionalized or arbitrary form, HMs do not seem to differ from other vocabulary. However, following Dingemanse & Woensdregt (2020) it is here held that this form is not entirely arbitrary, and is in some way shaped by interactional pressures. This is why, for instance, it is rare to see any instances of high-front [i] or high-back [u] as HM qualities; if HMs were entirely arbitrary, as most vocabulary is, any quality should be able to stand-in. Consistently, however, we find this is not the case. Only a limited set of vowel qualities occur consistently with HMs, ranging from the mid-front to central to low.

In terms of F0, conclusions are somewhat murkier. The results do not unambiguously support a Tone 6 equivalence, which was the predicted result. Although there was some evidence supporting this, the degree of variation across speakers, the high number of outliers, and the similarity between Tone 3 and Tone 6 complicated interpretation. High variability in F0 as well as low F0 is still in line with crosslinguistic trends for HMs (Shriberg 2001; Candea et al. 2005; Zhao & Jurafsky 2005; Braun & Rosin 2015), in line with claims of ease of articulation and low disruption to communication, but inability to effectively categorize HM tone makes it difficult to claim that HMs are incorporated into HKCT tonal phonology. This additionally raises questions about the specificity of tone compared to vowel quality: HM vowel quality seems to be conditioned both by ease of articulation (among other factors) and language-specific vocalic inventories, but whether the same can be said for tone is uncertain; this will be further discussed in §5.2.

Inter-speaker variation in terms of F0 appears to be larger than in the general vocabulary; while it is still low for each speaker, HM F0 appears to be incorporated into the phonology in different ways. This 'speaker-specificity' is a feature of HMs in other languages (Brown & Rosin 2015; McDougall & Duckworth 2017). While HMs are predicted to be constrained by universal properties, solutions to these constraints can be speaker specific. Consider the case of speaker 'm1' in the sample (see §4.3). This speaker manages the constraint for low F0 not by assimilating it to Tone 6, but rather has an F0 far below any lexical tone in their inventory. A case like this could be interpreted as giving more weight to cross-linguistic trends in HM well-formedness, at the expense of fitting the HM within the HKCT tonal system. A potential reverse situation can be observed in Mandarin (see §2.3); the choice of alveolar over bilabial place of articulation for nasal-final HMs (in some cases at least) could be interpreted as prioritizing language-specific phonology. Different prioritization schemes could be at the heart of the high level of idiolectal variation.

Returning to the place of HMs within the HKCT grammar, it seems that they occupy a unique place, sandwiched between communicative pressures for regularity and language-specific pressures to conform to the relevant phonological inventory. While in some respects they are quite regular (vowel quality), in others they are quite variable (F0), although even in this regularity a central tendency towards 'lowness' can be observed.

[8] Schwa does occur in some phonologically reduced elements in CT, such as the SFP combination *ga-maa* [kə.ma].

5.2 The Omnisyllabic Tone Hypothesis

Evidence for the Omnisyllabic Tone Hypothesis (Matisoff 1995) is open to interpretation. While HMs tend toward the lower end of the F0 range for each speaker, comparable to both Tones 3 and 6 (see Figure 8), a clearly discernible lexical level-tone equivalent for HKCT hesitation markers remains elusive. This is not entirely unexpected given Candea et al.'s (2005) finding that HM F0 was more variable than F0 in lexical words, but as mentioned, Cantonese offers a unique case due to its omnisyllabic tonal system[9]. The similarities between Tone 6 and HM-F0 contour are more substantial than for Tone 3, but the difference is minor. This ambiguity questions the integrity of the omnisyllabic tone hypothesis, although further research, including perceptual research (crucially), is necessary before coming to any clear conclusions.

It is quite possible that the apparent similarity between hesitation markers and the mid- and low tones of Cantonese is due to the adoption of a 'hesitatory intonation', with F0 lower than the overall phrase (note that HKCT HMs are consistently lower in F0 than the phrases in which they appear, see §4.1). This is the case for English and German (Shriberg 2001; Braun & Rosin 2015) and presumably for other non-tonal languages (Candea et al. 2005). Under this analysis, there would be no grounds for adopting the Omnisyllabic hypothesis, and HM F0 would fall entirely under intonation and be independent from larger HKCT tonal phonology. This analysis raises questions about the special status of lexical tone relative to segmental vowels (which occur unambiguously in hesitations), which are outside the scope of this paper.

It was originally hypothesized (§2.4) that one of these two tones would be the default for HMs; the mid-level tone due to evidence from SFPs, which default to Tone 3 when they convey no additional meaning, and the low-level tone due to its being the lowest level tone, and due to its appearance in loanwords as the equivalent to unstressed syllables. Both of these provide strong support for the omnisyllabic tone hypothesis; therefore, assuming the tenability of this hypothesis, the failure of HMs to unambiguously attach to one of the level tones can be due to any or all of the factors listed below:

1) Merging of mid-level T3 and low-level T6
2) Hesitation markers as peripheral vocabulary
3) Contextual/Intonational effects on HMs

Beginning with point (1), there is some degree of acoustic indeterminacy between the mid-level Tone 3 and the low-level Tone 6 of HKCT; as mentioned previously, this may have something to do with these being the target of an ongoing phonological merger (Mok, Zuo & Wong 2013). It has also been noted (Khouw & Ciocca 2007) that the primary means speakers use to distinguish the level tones is F0; no other cues are suggested. Going back to the data, it is not possible to reliably distinguish these two tones in the sample; although T6 does end at a lower point by about 10Hz on average, average F0 across the syllable should be enough to reliably distinguish the two. Therefore, even if HMs clustered around one of these underlyingly, the acoustic indeterminacy between the two would make identification impossible without a perceptual test.

An additional issue which may muddy the relationship between HMs and HKCT tonal phonology is the lexical status of HMs as peripheral vocabulary; as is the case with sentence-final particles, peripheral items in HKCT may have non-arbitrary tonal profiles (see §2.4). For instance, F0 in SFPs can be seen to interact with intonation more than arbitrary vocabulary, with certain question types taking certain tones (Sybesma & Li 2007; Tsang 2020). Since HMs may have varied discourse moderating uses aside from marking a

[9] None of the languages in Candea et al. (2005) are tonal except for Mandarin, which as mentioned in §2.3 has a default neutral tone for peripheral items like HMs.

pause (Tottie 2016), these may influence tonal contour. This being the case, it is possible that HMs have no consistent tonal profile, and should be seen as independent of tonal phonology owing to their lexical status.

Closely related to this point is the possible interference of context or intonation on HM F0. Tone in HKCT interacts with intonation in several ways, particularly at the edges of an utterance (Ma, Ciocca, & Whitehill 2011; Matthews and Yip 2013: 28). As mentioned previously, linear models run on the effects of position on F0 turned up significant results for final position (see Appendix Table VI). It is possible that intonational effects, such as natural falls in F0 utterance finally, or contextual effects, such as speaker f2's high-rising Tone 2, influence overall HM F0. This may necessitate normalization prior to analysis, such as what was done for F1-F2. However, normalization of tone, such as the procedure outlined in Stanford (2016) for naturalistic connected speech, is not recommended due to the lack of acoustic differentiation between the mid-level (used as a reference for normalization) and low-level tones. It should also be noted that the actual effects of position within the phrase were relatively minor, and that the effect disappears when tests are run individually on female and male speakers.

6. CONCLUSIONS

A natural extension of the present research would be a perceptual study of HKCT hesitation markers, along the lines of Vasilescu et al. (2005). This would be particularly illuminating for issues related to HM tone, where analysis of acoustic data alone does not seem to suffice. For instance, a forced-choice task classifying tonal contour, equating it with one of the lexical tones, would be potentially enlightening. Future work on HMs could also focus on the effects of bilingualism, either in English or in Mandarin, on HKCT. Both languages have different HM profiles and are widely spoken in Hong Kong (Liang 2015), and (usually younger) speakers often codeswitch between English and Cantonese (Barton & Lee 2017). Hong Kong English combines aspects of both, which potentially leads to compelling sociolinguistic patterning of HM acoustics.

The purpose of this research has been to shed additional light on the acoustic manifestation of hesitation markers in the Hong Kong variety of Cantonese, as well as to theorize on how these items are incorporated into the tonal and vocalic phonological system of HKCT. An ancillary goal has been to explore how these items are incorporated into HM typology in a cross-linguistic sense, through an exploration of the theories and findings of Candea et al. (2005), Dingemanse (2017), and Dingemanse & Woesndregt (2020). It is hoped that a clearer definition of these items will lead to further research on Cantonese, as well as on peripheral items generally.

REFERENCES

Barton, D. & Lee, C. (2017). 'Methodologies for researching multilingual online texts and practices'. In *Researching Multilingualism: Critical and Ethnographic Perspectives* (Martin-Jones and Martin, eds.). London: Routledge.

Bauer, R. & Matthews, S. (2017). 'Cantonese'. In *The Sino-Tibetan Languages (2nd Ed.)* (Thurgood and LaPolla, eds.). Routledge: London.

Boersma, P., & Weenink, D. (2021). *Praat: Doing Phonetics by Computer* [Computer program]. Version

6.1.38, retrieved 2 January 2021 from <http://www.praat.org/>.

Braun, A., & Rosin, A. (2015). 'On the speaker specificity of hesitation markers'. The Scientific Committee for ICPhS (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow 10-14 August 2015*, The University of Glasgow, pp. 731-734

Candea, M., Vasilescu, I., & Adda-Decker, M. (2005). 'Inter-and intra-language acoustic analysis of autonomous fillers'. *Disfluency in Spontaneous Speech* (DISS) 4. 47–52.

Clark, H. H., & Fox Tree, J. E. (2002). 'Using *uh* and *um* in spontaneous speaking'. *Cognition*, 84(1).

Cheung, H. (1972). 香港語法研究 [A study of Hong Kong grammar]. Hong Kong: Chinese University of Hong Kong.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chor, W. (2014). 'Epistemic modulations and speakers stance in Cantonese conversations'. In Harvey, M. & Antonia, A. (eds), *The 45th Australian Linguistic Society Conference Proceedings – 2014* Newcastle: NOVA Open Access Repository <http://nova.newcastle.edu.au>.

De Leeuw, E. (2007). 'Hesitation markers in English, German, and Dutch'. In *Journal of Germanic Linguistics*, 19(2), 85-114.

Dingemanse, M., Torreira, F., & Enfield, N. J. (2013). 'Is "Huh" a universal word? Conversational infrastructure and the convergent evolution of linguistic items'. *PloS one*, 8(11), e78273.

Dingemanse, M. (2017). 'On the margins of language: Ideophones, interjections and dependencies in linguistic theory'. In Enfield, N. J. (ed.), *Dependencies in language*, 195–202. Berlin: Language Science Press.

Dingemanse, M. & Woensdregt, M. (2020). 'The cultural evolution of collateral signals'. In Motamedi, Schouwstra, & Filippi (eds.), *Redrawing the boundaries of language: Evolution of Language Proceedings of the 13th International Conference*.

Erbaugh, M. S. (2001). *The Chinese Pear Stories: Narratives Across Seven Chinese Dialects*. Retrieved January 2021, from <http://pearstories.org/docu/biblio.htm>.

Fox Tree, J. (2001). 'Listeners' uses of *um* and *uh* in speech comprehension'. *Memory & cognition*, 29(2), 320-326.

Fox, B., Hayashi, M., & Jasperson, R. (1996). "Resources and Repair: A Cross-Linguistic Study of Syntax and Repair." In *Interaction and Grammar*, 185–237. Cambridge: Cambridge University Press.

Kiu, K. L. (1977). 'Tonal rules for English loan words in Cantonese'. In *Journal of the International Phonetic Association*, 7(1), 17-22.

Kobayashi, S., Yamamoto, M., & Nakagawa, S. (1993). 'Acoustic characteristics concerning the occurrences of interjections, repairs etc'. In *The Technical Report of the Institute of Electronics, Information and Communication Engineers, SLP* 93-1-2, 7-10.

Khouw, E., & Ciocca, V. (2007). 'Perceptual correlates of Cantonese tones'. *Journal of phonetics*, 35(1), 104-117.

Kwong, L. & Wong, M. (2015). 'The Hong Kong Cantonese Corpus: Design and Uses'. In *Journal of Chinese Linguistics*. <http://compling.hss.ntu.edu.sg/hkcancor/>.

Law, S. P. (1990). *The Syntax and Phonology of Cantonese Sentence-final Particles*. (Doctoral dissertation, Boston University).

Lee, J. (2015). 'PyCantonese: Cantonese linguistic research in the age of big data'. Talk at the Childhood Bilingualism Research Centre, Chinese University of Hong Kong. September 15 2015.

Lee, W.-S. (2004). 'The effect of intonation on the citation tones in Cantonese'. in *Proceedings of the 1st International Symposium on the Tonal Aspects of Languages*. Beijing, 107–110.

Levelt, W. J. M. (1983). 'Monitoring and self-repair in speech'. *Cognition*, 14, 41–104.

Li, C. & Thompson, S. (1981). *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: University of California Press.

Liang, S. (2015). *Language Attitudes and Identities in Multilingual China: A Linguistic Ethnography*. London: Springer.

Liesenfeld, A. (2019). 'Cantonese turn-initial minimal particles: Annotation of discourse-interactional functions in dialog corpora'. In *Proceedings of the 33rd Pacific Asia Conference on Language, Information and Computation (PACLIC 33)*, 471-479.

Lobanov, B. (1971). 'Classification of Russian vowels spoken by different listeners'. *Journal of the Acoustical Society of America* 49:606-08.

Ma, J. K. Y., Ciocca, V., & Whitehill, T. L. (2011). 'The perception of intonation questions and statements in Cantonese'. *The Journal of the Acoustical Society of America*, 129(2), 1012-1023.

Matisoff, J. A. (1995). 'Tone, intonation, and sound symbolism in Lahu: loading the syllable canon'. *Sound Symbolism*, 115-129.

Matthews, S., & Yip, V. (2013). *Cantonese: A Comprehensive Grammar*. Routledge.

McDougall, K., & Duckworth, M. (2017). 'Profiling fluency: An analysis of individual variation in disfluencies in adult males'. *Speech Communication*, 95, 16-27.

Mok, P. P., Zuo, D., & Wong, P. W. (2013). 'Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese'. *Language Variation and Change*, 25(3), 341.

PolyU Corpus of Spoken Chinese, Department of English, Hong Kong Polytechnic University, Modified 4 June 2015, Retrieved 05/12/2020 from <http://asianlang.engl.polyu.edu.hk/> .

R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.

Sheik, A. (2021). *CantoDict.* URL: <http://www.cantonese.sheik.co.uk/dictionary/>.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). 'The preference for self-correction in the organization of repair in conversation'. *Language*, 53, 361–382.

Shriberg, E. E., & Lickley, R. J. (1993). 'Intonation of clause-internal filled pauses'. *Phonetica*, 50(3), 172-179.

Shriberg, E. (2001). 'To *errrr* is human: ecology and acoustics of speech disfluencies'. *Journal of the International Phonetic Association*, 153-169.

Silverman, D. (1992). 'Multiple scansions in loanword phonology: Evidence from Cantonese'. *Phonology*, 9(2), 289-328.

Stanford, J. N. (2016). 'Sociotonetics using connected speech: A study of Sui tone variation in free-speech style'. *Asia-Pacific Language Variation*, 2(1), 48-82.

Strassel, S., Kolář, J., Song, Z., Barclay, L. & Glenn, M. (2005). 'Structural metadata annotation: Moving beyond English'. *Interspeech*, 1545-48.

Sybesma, R., & Li, B. (2007). 'The dissection and structural mapping of Cantonese sentence final particles'. *Lingua*, 117(10), 1739-1783.

Thomas, E. & Kendall, T. (2007). *NORM: The vowel normalization and plotting suite*. <http://lingtools.uoregon.edu/norm/about_norm1.php>.

Tottie G. (2016). 'Planning what to say: *Uh* and *um* among the pragmatic markers'. In *Outside the Clause*, 97-122. John Benjamins.

Tsang, Y. (2020). *Basic Sentence Final Particles in Hong Kong Cantonese*. Hong Kong: Greenwood Press.

Vasilescu, I., Candea, M., & Adda-Decker, M. (2005). 'Perceptual salience of language-specific acoustic differences in autonomous fillers across eight languages'. *Interspeech*.

Wiedenhof, J. (2015). *A Grammar of Mandarin*. Amsterdam: John Benjamins.

Wieling, M., Grieve, J., Bouma, G., Fruehwald, J., Coleman, J., & Liberman, M. (2016). 'Variation and change in the use of Hesitation markers in Germanic languages'. In *Language Dynamics and Change*, 6(2), 199-234.

Yuan, J., Xu, X., Lai, W., & Liberman, M. (2016). 'Pauses and pause fillers in Mandarin monologue speech: The effects of sex and proficiency'. *Proceedings of Speech Prosody 2016*, 1167-1170.

Zhao, Y., & Jurafsky, D. (2005). 'A preliminary study of Mandarin filled pauses'. *Proc. DiSS'05, Disfluency in Spontaneous Speech Workshop*, Aix-en-Provence, France, 179-182.