Strength of syllabic influences on articulation in Mandarin Chinese and French: Insights from a motor control approach

Liang Ma, Pascal Perrier, Jianwu Dang

Strength of syllabic influences on articulation in Mandarin Chinese and French: Insights from a motor control approach.

Liang Ma[a*], Pascal Perrier[b,c] & Jianwu Dang[d]

[a] *Department of Chinese Language and Literature, Fudan University, Shanghai, China*

[b] *Univ. Grenoble Alpes, GIPSA-Lab, F-38000 Grenoble, France*

[c] *CNRS, GIPSA-Lab, F-38000 Grenoble, France*

[d] *Japanese Advanced Institute of Sciences and Technology, Ishikawa, Japan*

liangma@fudan.edu.cn, pascal.perrier@gipsa-lab.grenoble-inp.fr, jdang@jaist.ac.jp

Running title: Strength of syllabic influences on articulation

*Corresponding author. Tel.: +8618221172202

Post address: Room 701, Guanghua (west) Building, Fudan University

220, Handan road, Shanghai, China, 200433

E-mail address: liangma@fudan.edu.cn (Liang Ma)

Abstract

This paper investigates how strong the influences of syllables are on the organisation of speech movements in Mandarin Chinese and French. More specifically, we evaluate the strength of these influences by measuring the extent of anticipatory coarticulation in V_1CV_2 sequences, within and across the boundaries of the CV_2 syllable. In line with motor control studies that have investigated serial-order motor tasks such as speech production, we consider that the extent of anticipatory motor planning reveals information about the units of planning, their boundaries and their strengths. For $V_1/t/V_2$ and $V_1/k/V_2$ sequences, where V_1 and V_2 are either /i/, /a/ or /u/, we used an electromagnetic magnetometer to record articulatory data from three native speakers of Mandarin Chinese and three native speakers of French. We labelled key articulatory configurations for V_1, C and V_2 based on general articulatory criteria. Our two major findings are:

(1) Within CV_2 boundaries, coarticulation is strong for both groups of speakers; in French, the coarticulation patterns can be fully explained by anticipatory behaviour, while in Mandarin Chinese other strategies might also be involved.

(2) Across CV_2 boundaries, anticipatory coarticulation is observed in numerous cases in French, while no significant anticipation can be noted in Mandarin Chinese. We conclude that these measures provide evidence for the hypothesis that the strength of the syllable as a unit of speech planning is more important in Mandarin Chinese than in French. Finally, we propose potential explanations for these differences.

Keywords: Speech production, Syllable, Coarticulation, Speech Units，Cross-linguistic study

1. Introduction

The syllable is considered by a large majority of psycholinguists, phonologists and phoneticians to be the most universal representational unit of speech production. According to the main streams in the literature, the syllable structures how the speaker plans to produce a speech utterance, and also influences how the speaker organizes and coordinates articulatory movements during speech production (cf, Shattuck-Hufnagel, 2011 and Cholin, 2011, for a review of the relevant literature, or Hyman, 2010, for a clear refutation of this hypothesis). Some of these suggestions are summarized in the sections below.

1.1 Psycholinguistic models: the syllable at the core of speech production planning.

The most famous psycholinguistic models of speech production give the syllable a crucial role in the planning of an utterance. For Dell and his colleagues (Dell, 1986; Dell et al., 1999) the phonological encoding involved in speech planning is based on syllable structure (onset-nucleus-coda), and relies on the decomposition of each word into its syllable constituents and their stress patterns. This decomposition is stored for each word in the brain, in the *mental lexicon*. Phonological encoding consists then in placing the different phonemes of the word at their appropriate places within the pre-defined syllable-based structure of the utterance.

For Levelt et al. (1999), however, the syllabic structure is not given by the lexicon, and the input to the phonological encoding does not include any information about syllabic organization. Rather these authors suggest that syllabification which transforms a sequence of phonemes into a sequence of syllables is a fully integrated part of the phonological encoding itself, and depends on the context in which the word is pronounced. This “on the fly” processing makes adaptability and variability possible. It allows, among other things, for re-

syllabification at the frontiers of words. According to Levelt &Wheeldon (1994), for very frequent syllables the organization of the articulatory gestures is based on motor schemas stored in the brain in the so-called “mental syllabary”. These motor schemas would be systematically used for every repetition of these frequent syllables. For infrequent syllables the articulatory gestures would require to be redefined for each new repetition. This distinction opens the door for potentially different status in the *phonetic plan*, the output of the phonological encoding (Levelt, 1993), for frequent and infrequent syllables.

Although these psycholinguistic models are different, both of them suggest that the output of phonological encoding consists of a structure in which the syllable plays a major role. Hence, according to these models, the organization of articulatory movements is deeply influenced by the structuring of a sequence of phonemes into syllables.

1.2 In phonetic models the syllabic structure influences the articulatory organization.

Phonetic models, i.e. models that explain the relation between the phonological structure and the execution of the articulatory movements, also stress in the main the central role of the syllable.

In the context of the Articulatory Phonology (Browman & Goldstein, 1986, 1989), in which the phonological description of a speech sequence is based on articulatory gestures, Browman & Goldstein (1988, 2000) proposed that there should be a special cohesion between the different gestures underlying the constituents of the syllable. The authors then formally described this cohesion with a number of intergestural phasing rules known as *c-center theory*: at the onset of the syllable, the consonantal gesture is locked *in-phase* with the nucleus gesture; if the onset is a cluster of consonants, the individual consonantal gestures are locked *anti-phase* with each other; at the coda, the consonantal gesture is *anti-phase* with respect to

the nucleus gesture, and if the coda is a cluster of consonants, this phase locking only concerns the first consonant.

In the Converter/Distributor (C/D) model of Fujimura (1992, 2000) the generation of physical speech signals begins with the specification of the phonological input in terms of a string of syllables associated with prosodic structural information. The C/D model transforms this string of syllables into a time sequence of triangles in which each triangle specifies the time position, the duration and the magnitude of a syllable: the vertex of a triangle specifies the time position of the nucleus of the syllable, and the two extremities of the base specify the time positions of the onset and coda. The triangle height, correlated with the length of its base, represents the magnitude of the syllable. The magnitude determines both the amplitude and the duration of the jaw oscillation underlying the articulation of the syllable. Variations in the triangle's height account for variations in jaw amplitude observed when stress varies (Menezes et al., 2003).

In the DIVA model, the most sophisticated neural model of speech acquisition and speech production to date (Guenther et al., 1998, Guenther & Vladusicha, 2012), speech motor goals are specified by target formant trajectories at the level of the CV syllable (Guenther et al., 2006). DIVA learns and stores in the brain the articulatory patterns responsible for producing frequently used syllables into a "Speech Sound Map" in the left Ventral Premotor Cortex. In DIVA the production of an utterance made of frequent syllables activates the corresponding Speech Sound Map cells sequentially. This representation is inspired by the concept of the *mental syllabary* proposed by Levelt & Wheeldon (1994) (see also above in section 1.1).

1.3. Articulatory correlates of syllabic organization are phoneme- and language- dependent.

Behind these main streams in the various theoretical and modeling accounts of the organization of a speech utterance given by psycholinguists and phoneticians, the actual physical correlates of syllabic organization are far from being universal and consistent across languages. An example of this variability is the classical distinction made in speech rhythm characterization between syllable-timed languages (Cantonese, French, Spanish, Italian), stress-timed languages (English, German, Russian) and mora-timed languages (Japanese), although the efficacy of this distinction has recently been questioned (see for example Ramus et al, 1999, or Arvaniti, 2012)

Several studies that have tried to evaluate the c-center theory in various languages have also contributed to showing this variability (see Pouplier, 2012, for a review). In Romanian, Marin (2011) found that the c-center theory does not apply when the initial consonantal cluster includes a liquid or a nasal. In American English Marin & Pouplier (2010) found that the phase coordination between nucleus and coda differs from the c-center theory when the first consonant is a [l]. For German, Pouplier (2012) found exceptions to the in-phase onset-nucleus coordination predicted by the c-center theory, when the onset is [kl]. Hermes (2013) has shown that in Italian the c-center theory does not apply to word-initial syllables when the onset is a cluster with an initial [s] such as [sp] or [sf]. Pastätter & Pouplier (2014) found for Polish that the onset-nucleus phasing with initial clusters containing a sibilant varies with the position of the sibilant in the cluster. Pouplier & Benus (2011) found that in Slovak syllabic liquids, i.e. syllables in which the nucleus is not a vowel but a liquid consonant (such as [smrt]), the initial consonantal cluster ([sm]) is not in-phase with the nucleus ([r]). These different exceptions to the c-center framework suggest first that the segmental content of the syllable influences the articulatory organization within the syllable. This has been emphasized by Pouplier (2012) and carefully investigated for English and

German by Brunner et al (2014). In addition, the fact that exceptions to the c-center rules vary across languages in place (onset-nucleus or nucleus-coda phasing), in their number, and in the segmental nature of the clusters, suggests that the influence of the syllabic structure on articulation is also largely language-dependent.

1.4. Is the syllable the most basic functional unit of speech planning in Mandarin Chinese?

In this context the paper published by O'Seaghdha et al. (2010) about Mandarin Chinese is particularly interesting. O'Seaghdha and his colleagues suggested that the syllable should be the most basic functional unit of speech planning in Mandarin Chinese, in contrast to Indo-European languages that should rely much more heavily on phonemes. Their hypothesis was motivated by a literary review about the priming effects of phonemes as opposed to syllables in the production of words in Indo-European languages such as Dutch or English, and by a comparison of these results with their own observations of priming effects in the production of Mandarin Chinese. For example, O'Seaghdha et al. (2010) compared studies carried out by Meyer (1990, 1991) with speakers of Dutch, with their own investigations of speakers of Mandarin Chinese (Chen et al., 2002a). Meyer showed that, for the production of monosyllabic and bisyllabic words, speakers of Dutch had shorter reaction times when they had implicit information about the first phoneme of the word to be pronounced than when they did not have this information. Implicit information about the first syllable in its entirety also reduced the reaction time. These results support the idea that both the syllable in its whole and the segments that build the syllable play a role in speech planning for native speakers of Dutch. Using a similar implicit priming paradigm and a similar experimental procedure with native speakers of Mandarin Chinese, Chen et al. (2002a) did not find that implicit knowledge of the first phoneme influenced the production of bi-syllabic words, but they found that the implicit knowledge of the first syllable had a strong influence.

Interestingly they observed a priming effect even if the priming syllable and the target syllable did not have the same tone and shared only the same segmental constituents.

To provide further support to their hypothesis, O'Seaghdha et al. (2010) designed a set of new experiments for mono-syllabic words with native speakers of Mandarin Chinese and native speakers of English. All these experiments relied on the same implicit priming paradigm as the studies of Meyer (1990, 1991) and Chen et al. (2002a): in these experimental conditions, both the priming object, which delivers the implicit knowledge, and the prompt, which elicits the pronunciation of the target syllable, vary in form (bi-syllabic words, mono-syllabic words, non-orthographic symbols). In all cases, these experiments have produced two important results: (1) implicit knowledge of the onset phoneme of the syllable does not have any influence on the reaction time for Mandarin Chinese speakers, even though this effect is clear for English speakers; and (2) syllables sharing the same segmental constituents but not the same tone as the target syllable have a priming effect.

On the basis of this information, O'Seaghdha et al. (2010) concluded that "atonal syllables" (i.e. abstract syllables with unspecified tone) are the most basic functional units of speech production planning in Mandarin Chinese (see Qu et al., 2012, O'Seaghdha et al., 2013, Qu et al., 2013, for further discussions about these findings).

1.5. Our contribution: a motor control approach to investigate the strength of the syllable in speech planning.

In this paper, we investigate possible differences between Mandarin Chinese and French with respect to the strength of the syllable in speech planning. Our approach is radically different from those of Meyer and O'Seaghdha (see section 1.4) since we do not consider the planning phase itself. Instead we look for the consequences of this planning in the articulatory signals of speech production. We observe to what extent Mandarin Chinese and French differ, and whether these differences are consistent with the idea that syllable

structure will constrain the organization of articulation to a greater or lesser degree depending on the language. Thus our approach is more inspired by methodologies developed in experimental phonetics than by experimental paradigms classically used in psycholinguistics. Nevertheless we consider our work to be complementary to those studies cited above that were carried out from a psycholinguistic perspective. In line with many other researchers (see for example Ramus et al., 1999), we believe that high level phonological structure is reflected in motor control strategies, which are in turn observable at the articulation level. We therefore look at articulation in order to find evidence which could support differences in the phonological status of the syllable. In order to avoid in this investigation any possible bias associated with hypotheses linked with a phonetic model or a phonetic theory, we adopted a methodology based on general motor control principles that are not speech-specific.

Our methodology is mainly based on generally well-accepted theoretical points in motor control research. Since speech production is simultaneously a semiotic task and a serial-order motor task, it can be analyzed with a methodology used in motor control research (see Grimme et al., 2011 for a review), and by taking into account that linguistic rules and communication purposes apply constraints to the emergence of the motor control strategies. The word *planning* belongs to the key-words in motor control research: *motor planning* refers to the phase of preparation for forthcoming movement. During this phase, in serial-order motor tasks, the brain elaborates a sequence of motor commands that will be sent to the peripheral motor system, in the aim of reaching a certain number of specified motor goals at the right time. One feature of serial-order human motor planning is *anticipation*: the Central Nervous System takes into account the forthcoming goals when it specifies the motor commands at a certain time (Schmidt, 1968; Orliaguet et al., 1997). Schmidt (1968) considered that anticipatory strategies in serial-order motor planning are "*concerned with the movement of the various body segments in the proper direction and at the proper time so that*

the movement is "coordinated," and the resulting movement is smooth and efficient" (p.631). Thus, anticipation in motor planning is particularly important in tasks that are achieved under time constraints, such as piano playing (Engel et al., 1997) or speaking (Daniloff & Hammarberg, 1973; Benguerel & Cowan, 1974; Lubker, 1981; Bell-Berti & Harris, 1982; Perkell, 1990; Abry & Lallouache, 1995; Vaxelaire et al., 2007; Noiray et al., 2011). Another generally well-accepted idea is that serial-order motor planning decomposes a sequence of movements into a series of chunks, or motor planning units. These chunks work on two levels: on a lower level, they are planned individually, and on a higher level, they are grouped into bigger planned chunks (see, e.g., Klapp, 2003, for taping or speech production, and Hulstijn & Van Galen, 1983, for handwriting). This idea is very similar to the general hypothesis made in the most successful psycholinguistic models (see in section 1.1 above), in which speech planning also structures an utterance into chunks, or phonological units, namely syllables, words, sentences, etc. (see also Kent & Minifie, 1977)

Thus the basic principles underlying our methodology are as follows: (1) since speaking is a serial-order motor task constrained by linguistic rules, speech planning and speech motor planning should interact in such a way that speech motor planning units should reflect phonological units; and (2) since anticipation is one of the key features of speech motor planning, we should be able to learn more about the boundaries of the phonological units and their strengths by looking at the consequences of anticipation in speech signals. Hence, we have recorded articulatory data from Mandarin Chinese speakers and from French speakers. In order to contribute to the debate about possible differences in the cohesive strength of the syllable between Mandarin Chinese and Indo-European languages, we have chosen French as our Indo-European language because it is one of the most often cited examples of "syllable-timed" languages. When analyzing the articulatory data, we have looked for correlates of anticipation in motor control. Acoustic and articulatory correlates of

anticipation in speech have been called *anticipatory coarticulation*, a term which we will adopt henceforth. In order to estimate the strength of the syllable, we have observed how syllable boundaries influence anticipatory coarticulation, and we have statistically assessed whether these influences differ across the two languages under investigation. Our approach shares some similarities with that of Kozhevnikov & Chistovith (1965), some 50 years ago, in their investigation of the influence of the syllable on articulation in Russian.

In section 2, a summary of the main phonological and phonetic properties of French and Mandarin Chinese is provided, with a special emphasis on the syllable characteristics in these two languages. In section 3, the methodology is described. In section 4, quantitative measures of anticipatory coarticulation are presented and analyzed. In the last section, the findings are discussed, and we show how our results contribute to the debate about the strength of the syllable in speech planning.

2. Some phonological and phonetic basics in Mandarin Chinese and in French.

The role of the syllable in lexical access, word spotting, as well as its link with the lexicon has been already largely studied in Mandarin Chinese and in French. The short summary of some important findings presented below sheds light on the focus of our study, namely potential differences in syllable strength across both languages.

Our methodology is based on the study of anticipatory coarticulation in articulatory movements. Our scope of investigation is articulation, rather than acoustics, because non-linearities in the articulatory-acoustic relations might minimize in the acoustic signals some important aspects of anticipation. Evidence has revealed that patterns of anticipatory coarticulation might be influenced by the segmental properties of the language. For example, it has been suggested in the literature that the density of the vowel system of a language could influence vowel-to-vowel coarticulation patterns (Manuel, 1990). In addition, constraints

involved in the production of the consonant have also been shown to influence coarticulation in Vowel-Consonant-Vowel (VCV) sequences (Recasens, 1997, Fowler & Brancazio, 2000). In the design of our experimental protocol and the interpretation of our results, we have taken into account a number of important basic phonological and phonetic properties of each language, which could interact with the impact of the syllable on anticipatory coarticulation. These basic phonological and phonetic properties are also summarized below.

2.1 Status of the syllable

a. Mandarin Chinese

In Mandarin Chinese, the syllable is classically considered to be the most important structuring phonological unit (Chao, 1968; Cheng, 1973; Wu & Lin, 1989; Chen et al., 2002a; O'Seaghdha et al., 2010, 2013; see also section 1.4 above). This point of view has strongly influenced works on speech recognition, synthesis and information retrieval in Mandarin Chinese, since these works use mostly syllable-based features (see for example Chen et al., 2002b).

The Chinese syllable consists of four parts, an onset, a nucleus, a coda and a tone. The syllable structure is defined by strict rules: there are no consonantal clusters in a syllable; only a single consonant can exist in the onset positions; and only two consonants, the nasal /n/ and /ŋ/, can be located in the coda position. One consequence of these constraints is that Mandarin Chinese contains a very small number of syllables. There are only about 400 syllables when tones are not taken into account and about 1300 syllables when tones are included (Deng & Dang, 2007). In addition, most Mandarin Chinese words are monosyllabic, with the consequence that Mandarin Chinese contains a large number of homophones, some of them occurring with a very high frequency (Duanmu, 2002). Furthermore, the orthography of Chinese is based on characters of syllable size, since a written character (an ideogram) corresponds to a single syllable.

Finally Mandarin Chinese does not have any case of re-syllabification, a phenomenon that occurs at the boundaries between a word ending with a consonant and a word starting with a vowel, and that associates the final consonant and the following initial vowel within a single syllable in spite of the word boundaries (Delattre, 1969). This goes well with the fact that the syllable is strongly associated with the lexical content in Mandarin Chinese.

b. French

The French syllable consists of an onset, a nucleus and a coda. Onsets and codas in French can be clusters of two or three consonants, resulting in many possible syllables. French has re-syllabification phenomena. The word-final consonant may emerge as an onset of a syllable which extends into the next word. Syllables may cross word boundaries: for example the word “chaque” (=each) corresponds to the syllable [ʃak] in “chaque train” (=each train), while in the expression “chaque avion” (each airplane) it is distributed over two syllables [ʃaka] (Gaskell et al., 2002). Thus depending on the context the stop consonant /k/ in [ʃak] is either the coda or the onset of a syllable. Consequently, the predominant syllable structure in French is the open CV syllable, but other structures like CVC, VC and CCV are also common (Delattre, 1969).

Syllables in French seem to play a crucial role in speech perception for the segmentation of the speech signal (the so-called *syllable effect* proposed by Mehler et al., 1981). Mehler et al. hypothesized a top-down mechanism to explain their results: to retrieve the phonological structure from the acoustic continuum, listeners would project onto this continuum syllables stored as units in the brain. However, these findings were limited to $C_1/a/C_2$ sequences in which C_1 was a plosive and C_2 a liquid. Content et al. (2001a), studying a larger corpus in which C_2 was either a liquid, or a fricative or a stop consonant, did not succeed in their attempts to replicate Mehler et al.'s (1981) results. They suggested that the

syllable effect observed by Mehler et al. (1981) might have been due primarily to some fine "allophonic and/or subphonetic" acoustic differences, specific to the sequences in which C_2 is a liquid and that would reflect the underlying organization of speech segments, and not to a systematic top-down process involving syllabic representations. Furthermore, Content et al. (2001b) demonstrated that French listeners were not consistent with each other in the syllabification process: in a simple CVCV word, the second syllable is nearly always considered as starting with the consonant, while the first syllable is associated either to CV or to CVC.

2.2 Vowel inventories

The description of the vowel inventory in Mandarin Chinese is quite controversial. Mandarin Chinese vowels are mainly associated with three places of articulation (front, mid, back) and three degrees of aperture (high, mid, low). The strongest controversy concerns the mid-vowel, since the mid-vowel in Mandarin Chinese has many variations, which can be classified as a mid central /ə/, a mid front /e/, a mid back /ɤ/ or a rounded /o/ (Chao, 1968, Duanmu, 2002). However, most studies agree that Mandarin Chinese has only five vowels: three high vowels with a front unrounded /i/, a front rounded /y/ and a back rounded /u/; a low vowel /a/; a mid central vowel /ə/ (Chao, 1968, Duanmu, 2002, Deng & Dang, 2007).

The French vowel inventory includes 11 oral vowels and four nasal vowels (Calliope, 1989). Distinction among the 11 oral vowels is realized in a three-dimensional space, which includes the place of articulation (three places along the front/back direction), the degree of aperture in the vocal tract (four degrees) and the rounded/spreading contrast of the lips for front vowels.

2.3 Consonant inventories

Mandarin Chinese has 22 consonants: six stops /t^h, k^h, p^h, t, k, p/, five fricatives /s, f, ʂ, x, ɕ /, six affricates / ts, ts^h, tʂ, $tʂ^h$, tɕ, $tɕ^h$/, three nasals /m, n, ŋ/, one liquid /l/ and one approximant /r/. Consonants can be categorized as labial, dental, palatal, retroflex, and velar sounds, depending on the place of articulation.

In French, there are 17 consonants: six stops /t, k, p, d, g, b/, six fricatives /v, f, z, s, ʒ, ʃ/, three nasals /m, n, ɲ/, one liquid /l/ and one trill /R/ (Calliope, 1989). Depending on the place of articulation, French consonants can be classified into labial, dental, alveolar, palatal or velar sounds.

Looking at the stop consonants for Mandarin Chinese and for French, one can see that there are the same numbers of stops in the two languages. However the contrast in Mandarin Chinese stop consonants depends on whether they are aspirated or non-aspirated while in French it depends on whether they are voiced or non-voiced. Nevertheless, the stop consonants in the two languages are associated with the same place of articulation (labial, dental and velar).

2.4 Tones in Mandarin

Mandarin Chinese is a tonal language, in that its tones are lexically distinctive. It has four stressed lexical tones and one neutral tone. The tone and the syllable are inseparable in Mandarin since every syllable is assigned a tone in the lexicon. For example, 妈 *ma1* (Tone1: high level tone, “mother”), 麻 *ma2* (Tone2: low rising tone, “numb”), 马 *ma3* (Tone3: falling-rising tone, “horse”), and 骂 *ma4* (Tone4: high falling tone, “scold”) represent four different words by four different tones (the words are written as *ma1*, *ma2*, *ma3*, and *ma4* in the Pinyin transcription system).

In Mandarin each tone always concerns a syllable in its entirety. It is generally assumed that tones are associated and aligned with the syllable (Xu, 1998). A tone does not extend across syllable boundaries.

A study of slip errors in Mandarin Chinese has shown that errors can affect the segmental constituents of the syllable independently of the tone. This suggests that within a syllable, tones and segmental constituents are represented separately (Chen, 1999). Gao has shown that the c-center theory developed in the context of the Articulatory Phonology (see section 1.2 above) applies to the understanding of the coordination between the tones and the supralaryngeal articulation characterizing the segmental constituents (Gao, 2008)

2.5 Summary

In summary, results from the literature draw the following picture for the status of the syllable in Mandarin Chinese and in French: (1) for both languages, the syllable plays an important role in structuring a spoken utterance and it influences the physical signals of speech; (2) in Mandarin Chinese, the syllable is strongly linked to the lexicon and syllable structure is defined by strict rules; (3) in French, more variability is allowed in syllable structure and the link with the lexicon is less strong; (4) in Mandarin Chinese and in French, the group of the segmental constituents of the syllable is a unit and this group can be considered independently of the tone in Mandarin Chinese.

As far as segmental properties are concerned, some differences exist between both languages; in particular, in consonant production. However, they are limited and both languages do share a number of segmental properties.

All these factors lead us to believe that comparing French and Mandarin Chinese is a valid approach to the evaluation of the strength of syllable influences on articulation. Because of the dissociation between tones and segmental constituents in the syllable representation of

Mandarin Chinese and to the proximity in the segmental properties of both languages, a common corpus based on logatoms, including comparable phonemes in both languages seems to be relevant. The absence of tone in French is not a drawback in this comparison. On the basis of these statements, we designed the experiment described below.

3. Methodology

Our methodology (Ma, 2008) is based on the concept of anticipation in serial-order motor control. Hence, we implicitly assume that a speech utterance can be decomposed into a series of discrete goals, and that forthcoming goals are likely to influence the planning and the execution of current goals. Importantly, we consider these goals to be related to phonemes in both languages. The existence of phonemes is controversial. Hence it is crucial to understand that assuming a role for the phonemes at the lowest level of speech motor control does not bias our study. This is essentially for two reasons. First, proponents of the syllable as basic unit of speech production do not systematically deny the possibility that phonemes exist. Syllables are often considered to be groups of phonemes with a specific cohesion. For example, even though O'Seaghdha et al. (2010) consider the syllable to be the most basic functional unit in Mandarin Chinese, the model of the production of a CV monosyllable described on their Figure 1 (p. 286) clearly gives a place to the phoneme. Second, our methodology provides also a way to test the phoneme hypothesis itself: if anticipation exists across units having the size of a phoneme, it will support the concept of phoneme-sized segments.

More specifically, our corpus consists of Vowel1-Consonant-Vowel2 (V_1CV_2) sequences. Anticipatory coarticulation is measured on both sides of the boundary of the CV_2 syllable. We observe how vowel V_2 influences both the production of vowel V_1 (anticipatory

coarticulation passing through the syllable boundary) and the production of consonant C (anticipatory coarticulation remaining inside the syllable boundaries). Expected results and their interpretations in the context of a serial-order representation of the speech production task can be summarized as follows:

- Anticipation of V_2 is observed in V_1 and C. This result would suggest that V_1, C, and V_2 are individual goals on which the whole sequence is built. This would support the hypothesis that phoneme is one unit of the representation of the speech production task, and that syllable boundaries have no significant influence on anticipatory strategies.
- Anticipation of V_2 is observed in C, but not in V_1. This result would suggest that C and V2 are two individual goals on which the syllable is built, and that V1 is controlled independently from the syllable. This would support the hypothesis that the phoneme is one unit of the representation of the speech production task, and that the syllable boundaries have a strong influence in the organization of articulatory movement.
- No anticipation is observed. This result would suggest that the proposed representation of speech motor control as a serial-order task based on phoneme-sized segments is not valid.

Combinations of these results can be also found, which will be interpreted accordingly.

3. 1. Corpus

Speech material consisted of V_1CV_2 nonsense words: /ak/V_2, /uk/V_2, /ik/V_2, /at/V_2, /it/V_2 and /ut/V_2 where vowel V_2 was /a/, /i/ or /u/. The words were uttered at a normal speech rate by three native speakers of French (FR1, FR2 and FR3) and three native speakers of Mandarin (CH1, CH2 and CH3). Each target word was embedded in a carrier sentence: “C’est V_1CV_2 ça?” (/sɛ/ V_1CV_2 /sa/ ʔ) in French and “这是 V_1CV_2 吗?” (/tʂə ʂʅ/ V_1CV_2 /ma/?) in

Mandarin Chinese. Each carrier sentence was repeated 10 times, except for subjects FR1 and CH2, who were pilot subjects and only produced 4 repetitions.

In spite of the differences in phoneme inventories and in linguistic structures, which have been listed above, an attempt was made to ensure that the corpora of both languages contained similar sequences and were similar in size. The *high level* tone was used for all the sequences in Mandarin Chinese, because it is most similar to the characteristics of the same sequence when it is uttered in French. The pitch and the intensity contours of each sentence were controlled *a posteriori* and they were found to be similar in both languages. Additionally, the phonotactic rules of both languages were respected. In Mandarin Chinese, the velar series consonant /k, k^h/ cannot be combined with the high front vowels /i/ and /y/. The corpora did not include the sequences /aki/, /iki/ and /uki/ in either language. The /ta/, /ti/, /tu/, /ka/ /ku/ syllables are frequent syllables in both languages.

The corpus was presented to the subjects in written form on several sheets. The V_1CV_2 sequence was written with phonetic symbols that were interpreted by the speakers of Mandarin Chinese as *pinyin* transcriptions. In the *pinyin* transcription, symbols “t” and “k” correspond to the aspirated consonants /t^h/ and /k^h/. Mandarin Chinese speakers therefore produced aspirated stops.

Together with the acoustic speech signals, the articulatory data were collected with an electromagnetic midsagittal articulograph (EMMA; AG100 Carstens Electronics). Four sensors were placed on the tongue at distances ranging from 0.5cm to 5cm from the tongue tip. Sensors were also attached to the upper lip, the lower lip, and the lower incisor. Two reference sensors were placed on the upper incisor and on the bridge of the nose. All the sensors were carefully located in the midsagittal plane, in order to ensure optimal measurement accuracy. The sensors on the tongue were named T_1, T_2, T_3 and T_4, from the tongue tip to the tongue back. A special effort was made to attach the sensors for all subjects

at the same distance from the tongue tip: T_1 at 0.5cm, T_2 at 2cm, T_3 at 3.5cm and T_4 at 5cm. The sensors' signals were sampled at 200Hz, and the acoustic signal at 20kHz. The orientation in the mid-sagittal plane of the subject-specific bite plane was measured before speech recordings started by asking the subject to maintain a Plexiglas slab firmly between his/her upper and lower teeth.

Before the start of the analysis of the data, using a combination of rotations and translations in the mid-sagittal plane, the data were aligned in reference to the two reference sensors that were attached to the incisor and to the bridge of the nose. These rigid transformations were made in order to correct for possible head movements in the mid-sagittal plane. In a final step, the coordinates were transformed in the mid-sagittal plane in order to localize the origin of the space at the position of the upper incisor sensor and to align the X-axis with the direction of the bite plane. Head corrected articulatory data were low-pass filtered (0dB at 18Hz; -20dB at 20.1 Hz) before they were analyzed. The articulatory data were labeled by hand with a Matlab® software that we developed specifically for this purpose (see below, section 3.2) for more details.

3.2. Labeling

The aim of the labeling procedure was to achieve accurate detection of the vocal tract configuration that is representative of the context dependent articulation of each elementary sound of the V_1CV_2 sequence. This was done interactively using the following guidelines:

- For stop consonants, the configuration at release was chosen. For this to happen the burst onset was manually located on the acoustic signal based on both the spectrographic and the temporal representations. The nearest articulatory configuration in time (sampled at much slower sampling rate than acoustics) was then selected.

----------Insert figure 1 about here-------

- More specifically for vowels the labeling procedure consisted of two successive steps (see Figure 1 for an /aka/ example recorded from the French subject FR1). In *Step 1* the time interval within which the first three formants were quasi stable was defined manually on the spectrogram. The middle point of this interval was taken as a first approximation of the vowel position (Circles on Figure 1). In *Step 2*, starting from this first approximation, a more accurate label was automatically determined by looking for the closest most characteristic articulatory configuration of the vowel in the sagittal plane. The most characteristic articulatory configuration was specified in the midsagittal plane based on the classical phonetic description of the articulations of vowels: for vowel /a/, it was considered to be the lowest back position; for vowel /i/ it was considered to be the most anterior high position; for vowel /u/ it was considered to be the most posterior high position. We based this labeling on tongue back sensor T_4 (Stars on Figure1). Sensor T_4 was chosen because in the great majority of tongue positions it corresponds to the highest part of the tongue. Accordingly, sensor T_4 gives the most reliable information about the displacement of that part of the tongue which has been classically used for vowels to describe the global back/front and low/high positioning of the articulation (Straka, 1965). In the example depicted in Figure 1, it can be seen that a difference as big as 2mm can exist between the sensor's positions observed at these two labels (stars *versus* circles).

3. 3. Data analysis

The data were analyzed in order to study anticipatory strategies in articulation. For each V_1CV_2 sequence, the influences of the second vowel V_2 on the articulation of the first vowel V_1 and on the consonant C were statistically analyzed.

The tongue position was characterized with the three sensors T_2, T_3 and T_4. The tongue tip sensor T_1 was not included in the analysis, since the tip is often considered to be less constrained than other parts of the tongue during vowel articulation (see for example Harshman et al., 1977, p.702). Hence the variability of sensor T_1 can be due to factors other than those we controlled for the purpose of the study.

Because of hardware problems during the experiment, a large amount of data was missing on sensor T_3 for subject CH_3. Sensor T_3 was not therefore taken into consideration in the analysis of the results for CH_3.

The anticipatory strategies of the six speakers were analyzed and compared using z-score normalized data. The z-score transformation was used with the objective of reducing the variability due to intrinsic speaker-specific factors. The z-scores were calculated separately for each speaker using formulas (1) and (2) below, where x and y represent the horizontal and vertical positions of the tongue, respectively. μ_x and μ_y are the mean position values for all tokens, and σ_x and σ_y are the standard deviations across those tokens. z_x and z_y are the normalized horizontal and vertical values of the z-scores.

$$z_x = \frac{(x-\mu_x)}{\sigma_x} \quad (1) \qquad z_y = \frac{(y-\mu_y)}{\sigma_y} \quad (2)$$

A variance analysis ANOVA (Repeated Measures) was carried out as an across-speaker analysis for /ak/V_2, /uk/V_2, /ik/V_2, /at/V_2, /it/V_2 and /ut/V_2, separately. The dependent variables were z_x and z_y of the sensors for vowel V_1 and for the consonant C. The independent

factors were V_2 (3 levels for V_1/t/V_2, namely /a/, /u/, or /i/, and 2 levels for V_1/k/V_2, namely /a/ or /u/), and the Language (2 levels, French and Mandarin Chinese). The main effect of V_2 and the interaction between Language and V_2 were measured. Statistical analyses were carried out for the effect of V_2 on V_1, and the effect of V_2 on C separately. Differences are considered to be statistically significant for $p<0.05$.

We first looked at the significance level of the Language x V_2 interaction. If it was significant ($p<0.05$), post hoc t-tests (with Bonferroni correction) were conducted to assess the simple main effects of V_2 in French speakers and in Chinese speakers separately. If the Language x V_2 interaction was not significant, we considered the level of significance of the main effect of V_2 to evaluate the anticipatory strategies globally across both groups of speakers. SPSS for Windows was used for these statistical analyses.

4. Results

In this study, anticipatory coarticulation was first analyzed by considering the amount of variability in V_1 and in C associated with changes of V_2. This variability was then further analyzed, in order to assess whether it was due to an anticipation of V_2 in V_1 or C. Variability was assumed to be the consequence of anticipatory strategies, if it corresponded to a clear trend of the articulatory configuration of V_1 or C to move from a reference non coarticulated configuration towards the articulatory configuration of V_2. This evaluation required a quantitative determination of the key characteristics of the reference non coarticulated articulatory configuration of each vowel of the corpus.

4. 1. Key characteristics of sensors' positions in vowels /a,u,i/

To determine the key characteristics of the sensor positions for the three vowels of the corpus, symmetrical V_1/t/V_2 sequences in which V_1=V_2 were used (Fig 2). The apical consonantal context was chosen in order to reduce the impact of the consonantal articulation on the main

part of the tongue. The symmetrical context was selected to avoid any form of vowel-to-vowel influence.

-------insert figure2 about here-------

The key characteristics can be described as follows:

- For vowel /a/, all the sensors have a relatively low height for all the subjects in both languages; this is associated with an essentially flat tongue shape; hence, the vocal tract constriction (i.e. the region of minimum cross-sectional area in the vocal tract) is not located in the palatal region described by the sensors.
- For vowel /u/, the sensors describe an arch with its highest point located in the back of the palatal region; T_3 (when it is available) and T_4 are located in the highest part of the arch, in its back part close to the palate for all the subjects in both languages.
- For vowel /i/, for all the subjects in both languages, except for FR_2, the sensors describe an arch with its highest point located in the front of the palatal region; for these subjects, T_2 and T_3 are located in the highest part of the arch, in its frontal part close to the palate while T_4 is lower. For subject FR_2, T_3 and T_4 are in the highest part of the arch.

Considering these key characteristics of the positions of the sensors for /a, u, i/, anticipatory maneuvers in V_1 or in C should be associated for all subjects with the following properties:

- If V_2 is /a/: the T_3 and T_4 positions should be lower than if V_2 is /u/; the T_3 position should be lower than if V_2 is /i/.
- If V_2 is /u/: the T_4 position should be higher (except for subject FR_2) and more posterior than if V_2 is /a/ or /i/; the T_3 position should be higher than if V_2 is /a/.

- If V_2 is vowel /i/: the T_2 and T_3 positions (T_3 and T_4 positions for subject FR_2) should be higher , and T_2, T_3 and T_4 positions should be more anterior than if V_2 is /a/ or /u/.

If our observations confirm these anticipatory maneuvers, we consider that the speakers have anticipatory strategies. If the observations do not confirm them, speakers are not considered as having anticipatory strategies. Henceforth, when we state that the results are or are not consistent with an anticipatory behavior, we will refer to the "end of section 4.1".

4. 2. Effects of V_2 on C (intrasyllabic coarticulation)

a) V_1/t/V_2 sequences

Figure 3 shows the dispersion ellipses calculated from the measurements of all repetitions of /at/V_2 sequences for each speaker, corresponding to a 2σ variation. The tongue positions (in cm) of C=/t/ are described by sensors T_2, T_3 and T_4. Different lines used to draw ellipses represent different V_2 contexts. It can be observed that the scatters of T_2, T_3 and T_4 show large differences depending on V_2. The clearest and the most robust differences are in the vertical direction both for the speakers of French and the speakers of Mandarin Chinese, while some variability also exists in the horizontal direction. Similar phenomena are observed for /ut/V_2 and /it/V_2 sequences

-------Insert figure3 about here---------

-------Insert table1 about here---------

In Table 1, the p-values of the main effect of V_2 and of the Language × V_2 interaction on the sensor positions of consonant /t/ in the V_1/t/V_2 sequences are listed for each vowel V_1. The horizontal (X) and vertical (Y) positions of the tongue were analyzed separately. In the Y

direction there is no significant difference ($p > 0.05$) between Mandarin Chinese and French, and the influence of V_2 on C is systematically significant. In the X direction, differences exist between the languages when V_1 is /i/ or /u/.

-------Insert table2 about here---------

Table 2 provides more details about these global results. It lists the significant average differences in sensor positions (post-hoc t-tests with Bonferroni correction were conducted on the z-scored data) measured for consonant /t/ between two different V_2 contexts. In the Y direction significant differences are found for all the sensors in the two language groups. All these differences are consistent with anticipatory strategies (see end of section 4.1), except in the Mandarin speakers' group for T_4 in the V_1/tu/ *versus* V_1/ti/ pairs, for which a significant opposite effect is observed for the three possible vowels V_1, /a,i,u/. In the X direction, significant differences occur less often than in the Y direction. These significant differences are all consistent with anticipatory strategies for the French speakers. For the Chinese speakers, the significant differences are mainly (6 cases among a total of 10 significant cases) in contradiction with anticipatory strategies. This last observation should however be taken with caution, since the 6 cases were observed for sensor T3, for which data are only available for two Chinese subjects (CH1 and CH2).

b) V_1/k/V_2 sequences

As mentioned in section 3.1, the corpora did not include /ki/ sequences for either language. Tongue sensor positions of /k/ in /ak/V_2 sequences are depicted in Figure 4. It can be seen that the differences of /k/ position related to the V_2 variation are important in both the horizontal and vertical directions. Similar phenomena were observed for the /ukV_2/ and the /ikV_2 /sequences.

In Table 3, the p-values of the main effect of V_2 and of the Language × V_2 interaction on the sensor positions of consonant /k/ in the V_1/k/V_2 sequences are listed for each vowel V_1. The horizontal (X) and vertical (Y) positions of the tongue were analyzed separately. There is no significant difference ($p>0.05$) between Mandarin Chinese and French. The influence of V_2 on C is systematically significant both for the French and the Mandarin speaker groups.

-------Insert figure4 about here---------

-------Insert table3 about here---------

-------Insert table4 about here---------

Table 4 provides more details about these global results. It lists the average difference in sensor positions (post-hoc t-tests with Bonferroni correction conducted on the z-scored data) measured for consonant /k/ between two different V_2 contexts. For both languages there are many significant differences. In the horizontal direction these significant differences are all consistent with anticipatory strategies (see end of section 4.1); however, all significant differences in the vertical direction are, for both speaker groups, in contradiction with the anticipation strategies (in boldface). This result will be discussed below in section 5.

4. 3. Effects of V_2 on V_1 (Extra-syllabic anticipation)

a . V_1/t/V_2 sequences

In Table 5, the p-values of the main effect of V_2 and of the Language × V_2 interaction on the sensor positions of vowel V_1 are listed for the V_1/t/V_2 sequences. The horizontal (X) and vertical (Y) positions of the tongue were analyzed separately. It can be stated that significant differences exist between the French and Mandarin speakers groups when V_1=[a]

and V_1=[u]. For V_1=[i], V_2 has no significant influence on V_1 for either language group. Figures 5 and 6 show the tongue positions (cm) for V_1 for /at/V_2 and /ut/V_2 sequences, respectively. Table 6 gives the details of the results for each of these sequences, split by language groups.

-------Insert figure5 about here---------

-------Insert figure6 about here---------

-------Insert table5 about here---------

-------Insert table6 about here---------

Looking at Table 6 it is striking to observe that among the 36 pairwise comparisons (2 V_1 x 3 V_2 x 2 directions x 3 sensors) in each group of speakers 24 are significant for the French speakers and only 4 are significant for the Mandarin speakers. Moreover for the French speakers all the significant comparisons are compatible with anticipatory strategies (see end of section 4.1) while for the Mandarin speakers 3 comparisons are in opposition with anticipatory strategies (boldfaces) and only one is in agreement with them. In particular, we observed that for the Mandarin speakers V_1 is more posterior in /ati/ than in both /ata/ and /atu/.

b . V_1/k/V_2 sequences

-------Insert figure7 about here---------

-------Insert table7 about here---------

As mentioned above, the V_1/k/V_2 analysis only involves V_1/ka/ and V_1/ku/ sequences. Figure 7 shows the tongue sensors positions for /u/ in /uk/V_2 sequences. The variations of the articulatory characteristics of V_1 associated with V_2 are weak compared to those observed in V_1/t/V_2 sequences (see Figures 5 and 6).

In Table 7, the p-values of the main effect of V_2 and of the Language × V_2 interaction on the sensor positions of vowel V_1 are listed for the V_1/k/V_2 sequences. The horizontal (X) and vertical (Y) positions of the tongue were analyzed separately. Significant differences between the language groups are observed for /ak/V_2 (in Y-direction only) and /uk/V_2 (in X-direction only), but not for /ik/V_2 for which no significant V_2-to-V_1 effect is observed. Table 8 gives the details of the significant results, split by language groups.

-------Insert table8 about here---------

Significant differences associated with changes in V_2 are observed for speakers of French only, and all these differences are compatible with the predictions of anticipatory coarticulation. There is no significant difference for the speakers of Mandarin.

4. 4. Summary

All the results taken together can be summarized as follows:

a. Intra-syllabic anticipation

In all the sequences under investigation a clear influence of vowel V_2 on C is observed both for French and Mandarin speakers.

For the French speakers, this kind of influence is fully compatible with anticipatory strategies except for consonant /k/ in the vertical direction; however, a careful analysis of the articulatory patterns of consonant /k/ for the French speakers shows that the positions of the

sensors in the vertical direction are the mechanical consequence of the anticipatory strategies in the X-direction. For this consonant, the elevation of the tongue is limited by the shape of the palate and the soft-palate because the tongue is in contact with the palate. Since the palate contour goes down towards the back of the vocal tract, and since anticipating consonant /k/ means moving the tongue backwards, anticipating will cause the tongue to lower, independently of the influence of V_2 in the vertical direction. Consequently our data shows clear evidence that in French speakers the influence of V_2 on C is entirely due to their anticipation of producing phoneme V_2 in phoneme C within the CV_2 syllable.

For the Mandarin speakers we also observe that V_2 tends to influence C, which is compatible with anticipatory strategies. However, the exceptions are more numerous for the Mandarin speakers than for the French speakers. Just as with the French speakers (see above), we find exceptions for consonant /k/, which can be explained by a combination of anticipatory strategies in the X-direction and the physical limitations of the shape of the palate where the tongue and palate come into contact. The exceptions observed for consonant /t/ are more intriguing: for all the vowels V_1, the tongue back sensor of the tongue is lower when /t/ is followed by the back high vowel /u/ than when it is followed by the front high vowel /i/; the tongue sensor T_3 is further back when /t/ is followed by the front vowel /i/ than when it is followed by the back vowels /a/ and /u/ (note though that data on T3 were available for two Mandarin speakers only). Thus, in the Mandarin Chinese speakers, the consonant C is influenced by the forthcoming vowel V_2 within a syllable. However in contrast to the French speakers this influence seems not to be systematically consistent with anticipatory maneuvers.

b. Extrasyllabic anticipation

We observe clear differences between the French and the Mandarin Chinese speakers in the extra-syllabic anticipation, especially in the influence of V_2 on V_1.

When V_1 is /a/ or /u/, the French speakers clearly and systematically anticipate the articulation of V_2 in V_1 when the consonant is either /t/ or /k/. When V_1 is /i/, V_2 does not affect V_1. Results for /i/ can be explained by the fact that the correct production of vowel /i/ allows very little freedom in the positioning of the tongue (Gay et al, 1991, 1993). Hence, anticipating V_2 in /i/ (i.e., moving the tongue away from the very accurate position required for a good perception of the vowel) would endanger the efficacy of the communication. Our data show for the French speakers clear evidence of systematic extra-syllabic anticipation in the V_1CV_2 sequences, when it does not endanger the perceptual quality of the utterance (Lindblom, 1990).

For Mandarin Chinese speakers, we find essentially no influence of V_2 on V_1 except in two /at/V_2 sequences. For only one sensor and one comparison, this influence is compatible with anticipation.

c. About the representativeness of our data

We are aware of the fact that our investigation includes only three speakers for each language. The intra syllabic anticipation of V_2 in C is quasi-systematic in all of the six speakers so that its statistical validity should not be questioned; however, given the small number of subjects in each language group, we cannot reject the possibility that the differences in extra-syllabic anticipation observed between the two groups are the result of speaker-specific factors rather than language-specific factors.

To clarify this issue we ran further statistical analyses of the anticipation of vowel V_2 in V_1. We did not reconsider anticipations of V_2 on C, since they were very clear for our data. First, we defined subgroups for which we randomly selected 2 of the 3 speakers in each language. We then used the same statistical method to compare extra-syllabic anticipation patterns of the French subgroups to those of the Mandarin subgroups (see section 3.3). Following this we carried out a variance analysis ANOVA (Repeated Measures) on the z-

scored normalized data. The Language-subgroup (two French speakers, two Mandarin speakers) × V_2 (/a/, /u/, or /i/) interactions were carried out for all sequences.

Second, nine new “virtual language" groups of two subjects were created by associating a French and a Mandarin speaker in each of the groups (see Appendix). For each pair of these “Virtual Language” groups (18 pairs in total, see Appendix) an ANOVA (Repeated Measures) was again carried out on the z-scored normalized data to assess the potential effect of the factor “Group” on the anticipation of vowel V_2 in V_1. The “Virtual Language” group × V_2 (/a/, /u/, or /i/) interactions were carried out for all sequences.

In the first set of statistical tests (subgroups of French speakers as opposed to subgroups of Mandarin Chinese speakers), the Language-subgroups ×V_2 interaction is systematically significant in both the X and Y directions. This indicates that the contribution of three subjects in each language is balanced in our experiment, and we can discard the possibility that a dominant subject in one of the groups could have inadvertently caused our results. In contrast, in the second set of statistical tests (mixed subgroups of French and Mandarin Chinese speakers), the “Virtual Language” subgroup ×V_2 interaction is essentially insignificant. While a few significant cases did exist, they were very rare and not systematic enough to be relevant (see Appendix). This indicates that if we combine a French and a Mandarin speaker in a “virtual language” group, the “virtual language” does not show any reliable influence on the extra-syllabic anticipation in V_1CV_2 sequences.

These additional statistical analyses support our initial results that French speakers show extra-syllabic anticipation within V_1CV_2 sequences, while Mandarin speakers show no such anticipation.

4. 5. Temporal measures for V_1CV_2 sequences

Our data suggest that French speakers anticipate vowel V_2 during the production of V_1 but Mandarin speakers do not. In contrast, both French and Mandarin speakers anticipate V_2

while producing C, although the anticipation is somewhat weaker and the coarticulation patterns are more complex for Mandarin speakers than for French speakers. As mentioned in section 3.1, the Mandarin speakers in our experiment produced aspirated stops. Consequently, stop consonants are not strictly the same for the speakers of the two languages. Their duration could be longer in Mandarin Chinese, which might be a cause for the differences observed in anticipatory behaviors in V_1. It has been shown that the amount of anticipation decreases when the time left to reach the target increases (Abry & Lallouache, 1995; Perkell, 1990). In order to assess this possibility carefully, we analyzed on two types of duration, namely the vowel-to-vowel V_1-V_2 interval and the duration of the opening gesture from the consonant release to the vowel V_2. The former is defined as the time interval separating labels V_1 and V_2 as determined in *step* 2 of the labeling procedure (see section 3.2), while the latter is defined as the duration going from the release of consonant C to the label V_2.

-------Insert figure 8 about here---------

-------Insert figure 9 about here---------

Figures 8 and 9 show the average duration and the standard deviations of all the sequences of each speaker respectively from label V_1 to label V_2 (V_1-V_2 duration), and from the consonant release to label V_2 (opening gesture, C-V_2 duration). ANOVAs were carried out for this temporal analysis. It shows that speakers FR1 and CH2 have significantly longer durations than the other two speakers, and that they cannot be clearly distinguished from each other. Speakers FR3, CH1 and CH3 feature intermediate durations, and Speaker FR2 has significantly shorter durations than the other speakers. In terms of variability, all the speakers display comparable standard deviations.

Observed differences in extra-syllabic anticipation between the French and the Mandarin Chinese speakers cannot be explained by the fact that aspirated consonants in Mandarin Chinese are longer than unvoiced consonants in French. There is no separation in

the measured durations that can be associated with the languages, while the results suggest that anticipation is clearly stronger for French. Our results show that there is no relationship between the amount of anticipation of V_2 in either V_1 or C and the time left to reach V_2.

5. Discussion and Conclusion

The development of motor skills in the achievement of serial-order motor tasks goes through “chunking”; structuring the tasks into chunks that are handled as single units of planning (see among many others, Verwey, 1996). Within each of these chunks, specific cohesive motor strategies take place, such as anticipation (Schmidt, 1968). Anticipatory coarticulation in the production of speech is a phenomenon that many have observed (Daniloff & Hammarberg, 1973; Benguerel & Cowan, 1974; Lubker, 1981; Bell-Berti & Harris, 1982; Perkell et al., 1990; Abry and Lallouache, 1995; Vaxelaire et al., 2007; Noiray et al., 2011). On the other hand, the phonological structure underlying the production of speech is made up of units that can have many different sizes: from the phoneme to the Intonational Phrase, by way of the syllable (Fougeron & Keating, 1997). On the basis of these different statements we have proposed a “motor control based” approach to compare the strength of the syllable as a unit organizing articulation in Mandarin Chinese and French. Our underlying hypothesis is that the strength of the syllable as a phonological unit should be reflected in the strength of the cohesive motor control strategies within the syllable, where the syllable is a way of “chunking” motor planning.

We have approached the topic by studying anticipatory coarticulation within and across syllable boundaries. We consider the strength of the syllable as a unit of motor planning, and then as one unit of phonological planning, to be stronger if coarticulation remains within the syllable boundaries and does not extend across them. We intentionally did not consider carry-over coarticulation because we believe it to be influenced by many factors:

motor planning similar to the one determining anticipatory coarticulation, physical factors associated with inertial properties and causality principles inherent to any physical system such as the biomechanical speech production system (see Whalen, 1990, for convincing experimental results along these directions). Note that our approach has some similarities to that of Kozhevnikov & Chistovich in 1965 for the study of the influence of the syllable on articulation in Russian.

We have studied three speakers of each language and have obtained clear statistical results showing that anticipatory coarticulation exists in both languages in V_1CV_2 sequences, but it remains strictly within the CV_2 syllable boundaries in Mandarin Chinese, while it goes beyond them and extends from vowel V_2 to vowel V_1 in French. The measure and the statistical comparison of CV_2 and V_1-V_2 durations in both languages demonstrated that the observed differences in V_1-V_2 anticipatory coarticulation are not the result of timing differences. The design of the corpus also eliminates possible influences of differences in the prosodic structure of the sequences across languages (this has been confirmed by a global comparison of the F0 and energy time patterns of the utterances) or in syllable frequency. The difference in the density of the vowel systems of the two languages is not strong, given the large contextual allophonic variability in Mandarin Chinese. This slight difference is in favour of a smaller number of vowels in Mandarin Chinese than in French, which would suggest the existence of a larger coarticulation for Mandarin. Hence, based on the evaluation criteria listed above in the introductory part of the Methodology section, our conclusions are as follows

- In both languages, the production of V_1CV_2 sequences can be seen as the production of a sequence of phonemes, and the production of a phoneme is anticipated in the preceding phonemes. Thus, phonemes are one unit of the speaking task representation in Mandarin Chinese and in French.

- Because the extent of anticipation reflects the strength of the syllabic structure in the organization of articulatory movement and then in phonological planning, we suggest that the strength of the syllable is greater in Mandarin Chinese than in French, in that sense that it has a larger impact on the organization of the motor command patterns.

The only case of a clear influence of vowel V_2 on vowel V_1 in Mandarin Chinese was found in the /ati/ versus /ata/ comparison (Table 6). The influence was the opposite of what we would expect from an anticipatory behaviour: the /a/ articulation in /ata/ is more anterior than the /a/ articulation in /ati/. Such coarticulatory behaviour in opposition to a phoneme-based anticipatory strategy is consistent with the observation that in Mandarin Chinese /t/ in V_1/ti/ is more posterior than /t/ in V_1/ta/ or V_1/tu/. This is consistent with the idea that in sequences longer than a syllable the anticipatory strategies in Mandarin Chinese could involve the forthcoming syllable as a whole. However, a significant influence of V_2 on /t/ in Mandarin Chinese was observed only for sensor T_3, for which data are available for only two subjects. Hence, further investigations are needed before drawing any definitive conclusion about potential differences between French and Mandarin Chinese in the internal structure of the syllable and its role in coarticulation in pluri-syllabic sequences.

Our findings support the hypothesis that the syllable has more influence on motor planning and thus on the phonological planning, in Mandarin Chinese than in French. Mok (2010) carried out a similar comparative study of the vowel-to-vowel coarticulation across the syllable boundaries in Thai and Southern British English. Her analysis focused on the acoustic spectral properties (Formants) of the vowels, and she investigated both anticipatory and carry-over coarticulation. Thai and Southern British English were selected because of the differences in the complexity of syllable structure. According to Mok (2010) in comparison with English, Thai has quite a simple syllable structure, in particular because of a very small number of clusters (if any) at the syllable onset and a reduced number of consonants at the

coda, many of them being weakly articulated. Additionally, Thai is a syllable-timed language while English is a stress-timed language. Mok's (2010) hypothesis was that "*The complexity of the* [syllable] *boundary and the coordination of the boundary with the vowels may affect the coordination between vowels across the boundary.*" (p. 1347). Mok (2010) found that the vowel-to-vowel coarticulation is larger in English than in Thai, mainly in the carry-over direction. She concluded that "*languages with simple syllable structure and unambiguous boundary may allow less V-to-V coarticulation than languages with complex syllable structure.*" (p. 1353). She suggested that an explanation for this difference could be that in languages with complex syllable structures "*more cues are carried by the consonants, so more vowel variation is allowed*" (p. 1353).

As regards the link between syllable structure complexity and vowel-to-vowel coarticulation, our observations are convergent with Mok's (2010) hypothesis. As shown above in section 2.1, the syllable structure in French is more complex than in Mandarin Chinese. Our results are stronger than Mok's (2010): while she observed vowel-to-vowel coarticulation in both languages, the one in English being more important than the one in Thai, we observed in Mandarin Chinese significant influence of V_2 on V_1 in only one case, a case which was in opposition with anticipatory behaviour. The influence of syllable boundaries in Mandarin Chinese therefore seems to be even stronger than in Thai.

Overall our results are compatible with the hypothesis that in Mandarin Chinese the syllable is an important unit of motor planning in speech production. Hence, we may hypothesize that this strength could result in the storage in the brain of specific global motor schemas for each syllable in Mandarin Chinese. This would explain why, in spite of the fact that the phoneme is one unit of representation in speaking production of Mandarin Chinese, it has no priming effect, while the syllable as a whole has a clear priming effect (O'Seaghdha et al., 2010) . This hypothesis could also explain why syllable structure is simpler in Mandarin

Chinese: this simplicity would reduce the number of motor schemas to be stored in the brain, and then facilitate both the storage itself and the access to these schemas in speech production (consistent with O'Seaghdha et al.'s (2010) observations). Syllable strength would not be the consequence of the simplicity of the structure, but the simplicity would be the consequence of the fact that syllable is a strong unit of motor planning. The causality link between simplicity of the syllable structure and the strength of the syllable would be precisely the inverse of the one proposed by Mok's (2010) hypothesis. The relation between syllable structure complexity and vowel-to-vowel coarticulation could be due to general motor control principles rather than to more abstract linguistic properties (plurality of the "cues"). Obviously this explanation must be further investigated and the debate is still open.

In summary, our data support the hypothesis that the strength of the syllable on speech articulation is stronger in Mandarin Chinese than in French. However our experimental study involved only 3 speakers in each language, and was based on V_1CV_2 sequences. Obviously further evaluations of this hypothesis are required on a larger corpus and a larger number of subjects before drawing definitive conclusions.

Acknowledgments

Special thanks are due to Noël Nguyen for insightful suggestions and comments during the conduct of the study. This work was in large part supported by a Grant from the French Ministery for Education and Research to Liang Ma. This work was also supported by a grant from the French "Agence Nationale de la Recherche" (DOCVACIM - ANR-07-CORP-0018 - PI: Rudolph Sock, Strasbourg University), by the National Basic Research Program of China (No. 2013CB329301), and by the national natural science foundation of China under contract No. 61233009.

References

Abry, C. & Lallouache, T. M. (1995). Modelling lip constriction anticipatory behavior for rounding in French with the MEM (Movement Expansion Model). In *Proceedings of the 13th International Congress of Phonetic Sciences, Vol. 4* (pp.152-155), Stockholm, Sweden.

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics, 40*(3), 351-373.

Bell-Berti, F., & Harris, K. S. (1982). Temporal patterns of coarticulation: Lip rounding. *Journal of the Acoustical Society of America, 71*, 449–459.

Benguerel, A. P., & Cowan, H. A. (1974). Coarticulation of upper lip protrusion in French. *Phonetica, 30*, 41–55.

Browman, C. P., & Goldstein, L. (1986). Towards an Articulatory Phonology. *Phonology Yearbook, 3*, 219–252.

Browman, C. P., & Goldstein, L. (1988). Some notes on syllable structure in Articulatory Phonology. *Phonetica, 45(2-4)*, 140–155.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6*, 201–251.

Browman, C. P., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée, 5*, 25–34.

Brunner, J., Geng, C., & Gafos, A. (2014). Phonetic effects in the timing of onset clusters. *Proceedings of the 10th International Seminar on Speech Production* (pp. 61–64) Cologne, Germany.

Calliope (1989). *La parole et son traitement automatique.* Paris: Masson.

Chao, Y. R. (1968). *A Grammar of Spoken Chinese.* Berkeley, CA: University of California Press.

Chen, J.-Y. (1999). The representation and processing of tone in Mandarin Chinese: Evidence from slips of the tongue. *Applied Psycholinguistics,20,* 289–301.

Chen, J.-Y., Chen, T.-M., & Dell, G. S. (2002a). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language, 46,* 751–781.

Chen, B., Wang, H-M, & Lee, L.-S. (2002b). Discriminating capabilities of syllable-based features and approaches of utilizing them for voice retrieval of speech information in Mandarin Chinese. *IEEE Trans. on Speech and Audio Processing, 10,* 303–314.

Cheng, C. C. (1973). *A synchronic phonology of Mandarin Chinese*. The Hague, The Netherlands: Walter de Gruyter, Mouton & Co.

Cholin, J. (2011). Do syllable exist? Psycholinguistic evidence for the retrieval of syllabic units in speech production. In C.E. Cairns and E. Raimy (Eds.), *Handbook of the Syllable* (pp. 225–254). Leiden, The Netherlands: Koninklijke Brill NV.

Content, A., Meunier, C., Kearns, R. K., Frauenfelder, U. H. (2001a). Sequence detection in pseudowords in French: Where is the syllable effect? *Language and Cognitive Processes, 16,* 609–636.

Content, A., Kearns, R. K., & Frauenfelder, U. H. (2001b). Boundaries versus onsets in syllabic segmentation, *Journal of Memory and Language. 45*, 177–199.

Daniloff, R. G., & Hammarberg, R. E. (1073). On defining coarticulation. *Journal of Phonetics, 1,* 239–248.

Delattre, P. (1969). Syllabic features and phonic impression in English, German, French and Spanish. *Lingua, 22,* 160-175.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

Dell, G. S., Chang, F., & Griffin, Z. M. (1999). Connectionist Models of Language Production: Lexical Access and Grammatical Encoding. *Cognitive Science, 23(4),* 517–542.

Deng, L., & Dang, J. (2007). Speech analysis: The Production-Perception perspective. In Hai-Zhou Li and Chin-Hui Lee (Eds.), *Advances in Chinese Spoken Language Processing* (pp. 3–32). World Scientific: New Jersey.

Duanmu, S. (2002). *The Phonology of Standard Chinese.* Oxford: Oxford University Press.

Engel, K., Flanders, M., & Soechting, J. (1997). Anticipatory and sequential motor control in piano playing. *Experimental Brain Research, 113,* 189–199.

Fougeron, C., & Keating, P. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America, 101,* 3728–3740.

Fowler, C. A., & Brancazio, L. (2000). Coarticulation resistance of American English consonants and its effects on transconsonantal vowel-to-vowel coarticulation. *Language and speech, 43*(1), 1-41.

Fujimura, O. (1992). Phonology and phonetics – A syllable-based model of articulatory organization. *Journal of the Acoustical Society of Japan (E), 13(1),* 39–48.

Fujimura, O. (2000). The C/D model and prosodic control of articulatory behavior. *Phonetica, 57,* 128–138.

Gao, M. (2008). *Tonal Alignment in Mandarin Chinese: An Articulatory Phonology Account.* Unpublished Doctoral Dissertation (Linguistics), Yale University, CT.

Gaskell, M. G., Spinelli, E., & Meunier, F. (2002). Perception of resyllabification in French. *Memory & cognition, 30*(5), 798-810.

Gay, T., Boë, L.-J., Perrier, P., Feng, G., & Swayne, E. (1991). The acoustic sensitivity of vocal tract constrictions: a preliminary report. *Journal of Phonetics, 19,* 445–452.

Gay, T., Boë, L.-J., & Perrier, P. (1993). Acoustic and perceptual effects of changes in vocal tract constrictions for vowels. *Journal of the Acoustical Society of America, 92*(3), 1301–1309.

Grimme, B., Fuchs, S., Perrier, P., & Schöner, G. (2011). Limb versus speech motor control: A conceptual review. *Motor control, 15(1),* 5–33.

Guenther, F. H., & Vladusicha, T. (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics, 25(5),* 408–422.

Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review, 105*, 611–633.

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language, 96,* 280–301.

Harshman, R., Ladefoged, P., & Goldstein, L. (1977). Factor analysis of tongue shapes. *Journal of the Acoustical Society of America, 62*(3), 693-707.

Hermes, A. (2013). *Articulatory Coordination and Syllable Structure in Italian*. In S. Fuchs & P. Perrier (Eds.). *Speech Production and Perception, Vol. 2*. Frankfurt am Main: Peter Lang Verlag.

Hulstijn, W., & Van Galen, G. P. (1983). Programming in handwriting: Reaction time and movement time as a function of required length. *Acta Psychologica, 54,* 23–49.

Hyman, L. M. (2010. Does Gokana really have no syllables? (or: What's so great about being universal?), *UC Berkeley Phonology Lab Annual Report*, 95–117.

Kent, R. D., & Minifie, F. D. (1977). Coarticulation in recent speech production models. *Journal of Phonetics, 5,* 115–133.

Klapp, S. T. (2003). Reaction times analysis of two types of motor preparation for speech articulation: Actions as a sequence of chunks. *Journal of Motor Behavior, 35(2)*, 135–150.

Kozhevnikov, V., & Chistovich, L. (1965). *Speech Articulation and Perception*, Joint Publication Research Service (pp. 104–118). Washington, D.C.

Levelt, W. J. (1993). *Speaking: From intention to articulation*. Cambridge, MA, USA: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–75.

Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition, 50*, 239–269.

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H&H theory. In W.J. Hardcastle and A. Marchal (Eds.) *Speech Production and Speech Modelling* (pp. 403–439). Kluwer Academic Publishers: The Netherlands.

Lubker, J. (1981). Temporal aspects of speech production: Anticipatory labial coarticulation. *Phonetica, 38*, 51–65.

Ma, L. (2008). *La coarticulation en français et en chinois : Étude expérimentale et modélisation.* Unpublished Doctoral Dissertation, Université Aix-Marseille I, Aix-en-Provence, France (212 pages).

Manuel, S. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America, 88*, 1286–1298.

Marin, S. (2011). Organization of complex onsets in Romanian. *Proceedings of the 9th International Seminar on Speech Production* (pp.179–186). Montréal, Canada.

Marin, S., & Pouplier, M. (2010). Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model. *Motor Control, 14(3)*, 380–407.

Mehler, J., Dommergues, J.-Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation, *Journal of Verbal Learning and Verbal Behavior, 20,* 298–305.

Menezes, C., Pardo, B., Erickson, D., & Fujimura, O. (2003). Changes in syllable magnitude and timing due to repeated correction. *Speech Communication, 40 (1-2),* 71–85.

Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language, 29*, 524–545.

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language, 30,* 69–89.

Mok, P.K.P. (2010). Language-specific realizations of syllable structure and vowel-to-vowel coarticulation. *Journal of the Acoustical Society of America,*

Noiray, A., Cathiard, M. A, Ménard, L., & Abry, C. (2011). Test of the movement expansion model: Anticipatory vowel lip protrusion and constriction in French and English speakers. *Journal of the Acoustical Society of America, 129*(1), 40–349.

Orliaguet, J.-P., Kandel, S., & Boë, L.-J. (1997). Visual perception of motor anticipation in cursive handwriting: Influence of spatial and movement information on the prediction of forthcoming letters. *Perception, 26,* 905-912.

O'Seaghdha, P. G., Chen, J.-Y., & Chen, T.-M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition, 115*, 282–302.

O'Seaghdha P. G, Chen J-Y, & Chen T-M (2013). Close but not proximate: The significance of phonological segments in speaking depends on their functional engagement. *Proc Natl Acad Sci. USA, 110(1),* E3.

Pastätter, M. & Pouplier, M. (2014). The temporal coordination of Polish onset and coda clusters containing sibilants. *Proceedings of the 10th International Seminar on Speech Production* (pp. 312–315) . Cologne, Germany.

Perkell, J. S. (1990). Testing theories of speech production: implications of some detailed analyses of variable articulatory data. In W.J. Hardcastle and A. Marchal (Eds.) *Speech Production and Speech Modelling* (pp. 263–288). Kluwer Academic Publishers: The Netherlands.

Pouplier, M. (2012). The gestural approach to syllable structure: Universal, language- and cluster-specific aspects. In Susanne Fuchs, Melanie Weirich, Daniel Pape, & Pascal Perrier (eds.) *Speech Planning and Dynamics* (pp. 63–96). Frankfurt am Main: Peter Lang Verlag.

Pouplier, M., & Beňuš, Š. (2011). On the phonetic status of syllabic consonants: Evidence from Slovak. *Journal of Laboratory Phonology, 2(2),* 243–273.

Qu, Q, Damian, MF, & Kazanina, N. (2012). Sound-sized segments are significant for Mandarin speakers. *Proc Natl Acad Sci USA, 109(35),* 14265–14270.

Qu, Q, Damian, MF, & Kazanina, N. (2013). Reply to O'Seaghdha et al.: Primary phonological planning units in Chinese are phonemically specified. *Proc Natl Acad Sci. USA, 110(1),* E4.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265-292.

Recasens, D., Pallares, M., & Fontdevila, J. (1997). A model of lingual coarticulation based on articulatory constraints. *Journal of the Acoustical Society of America, 102,* 544–561.

Schmidt, R. A. (1968). Anticipation and timing in human motor performance. *Psychological Bulletin, 70(6),* 631–646.

Shattuck-Hufnagel, S. (2011). The role of the syllable in speech production in American English: A fresh consideration of the evidence. In C.E. Cairns and E. Raimy (Eds.), *Handbook of the Syllable* (pp. 197–224). Leiden, The Netherlands: Koninklijke Brill NV.

Straka, G. (1965). *Album phonétique*. Québec, Canada: Les Presses de l'Université Laval.

Vaxelaire, B., Sock, R., Hirsch, F., & Roy, J.-P. (2007). Anticipatory laryngeal movements. An X-ray investigation. *Proceedings of the XVIth International Congress of Phonetic Sciences* (pp. 525–528). Saarbrücken, Germany.

Verwey, W.B. (1996). Buffer loading and chunking in sequential keypressing. *Journal of Experimental Psychology: Human Perception and Performance, 22(3)*. 554-562

Whalen, D. H. (1990). Coarticulation is largely planned. *Journal of Phonetics, 18*, 3–35.

Wu, Z.J, & Lin, M.C. (1989). *A general introduction to experimental phonetics.* Higher education press. (in Chinese).

Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica, 55*, 179–203.

FIGURE CAPTIONS

Figure 1: Articulatory labeling: Trajectory of the tongue sensors T_1, T_2, T_3 and T_4 (from tongue tip to tongue back) for the first vowel /a/ in the /aka/ sequence within a 100ms long duration around the middle point of the time interval of vowel formant stability. Circles and stars represent the sensors' positions respectively at the middle point of the formant stability interval and at the final label position (see text). The palate contour is plotted on the top of the figure as a spatial reference

Figure 2: Dispersion ellipses, corresponding to a 2σ variation, of the tongue sensor positions (in cm) for: (a) V_1=/a/ in /ata/ sequence, (b) V_1=/u/ in /utu/ sequence, (c) V_1=/i/ in /iti/ sequence. The solid line at the top of each plot represents the palate contour. Native speakers of French are represented in the top row and native speakers of Mandarin are in the bottom row.

Figure 3 Distribution and 2σ dispersion ellipses of the tongue sensors positions (cm) for /t/ in /atV_2/ sequences. Sensors T_2, T_3 and T_4 are presented from left to right. The solid line at the top of each plot represents the palate contour. Native speakers of French are represented in the top row and native speakers of Mandarin are in the bottom row. Black line: V_2=/a/; Gray line: V_2=/i/; boldface line: V_2=/u/.

Figure 4: Distribution and 2σ dispersion ellipses of the tongue sensors positions (in cm) for /k/ in /ak/V_2 sequences. Sensors T_2, T_3 and T_4 are presented from left to right. Black line: V_2=/a/; boldface line: V_2=/u/.

Figure 5: Distribution and 2σ dispersion ellipses of the tongue sensors positions (in cm) for V_1=/a/ in /at/V_2 sequences (see Figure 3 for details).

Figure 6: Distribution and 2σ dispersion ellipses of the tongue sensors positions (cm) for V_1=/u/ in /ut/V_2 sequences (see Figure 3 for details).

Figure 7: Distribution and 2σ dispersion ellipses of the tongue sensors positions (in cm) for V_1=/u/ in /uk/V_2 sequences. Black line: V_2=/a/; boldface line: V_2=/u/ (see Figure 4 for details).

Figure 8: V_1-V_2 time interval (in s) for French and Mandarin Chinese speakers

Figure 9: Duration (in s) of the opening gesture C-V_2 for French and Mandarin Chinese speakers

	sequences	*P* for V_2 -to-C effect	*p* for Language×V_2 interaction
Y direction	/at/V_2	0.000 *	0.097
	/ut/V_2	0.000 *	0.158
	/it/V_2	0.000 *	0.655
X direction	/at/V_2	0.588	0.065
	/ut/V_2	0.982	0.013*
	/it/V_2	0.060	0.000*

Table 1: p-values of the ANOVAs run on the sensors' positions (z-score data) of consonant /t/ in the V_1/t/V_2 sequences. The main effect of V_2-to-C and the Language ×V_2 interaction are presented. Horizontal (X) and vertical (Y) directions are analyzed separately. The symbol (*) represents the significant cases of ANOVA ($p<0.05$).

	French speakers			Chinese speakers		
/t/ in /at/V_2 sequences	/atu/ ***vs*** /ata/	/ati/ ***vs*** /ata/	/atu/ ***vs*** /ati/	/atu/ ***vs*** /ata/	/ati/ ***vs*** /ata/	/atu/ ***vs*** /ati/
T_{2x}		-0.841	1.330		-0.672	0.841
T_{3x}					**1.119**	**-1.381**
T_{4x}		-0.761	1.234		-0.715	0.441
T_{2y}		2.007	-1.801		1.857	-1.708
T_{3y}	1.210	1.706	-0.496	0.895	2.296	-1.402
T_{4y}	2.121	0.513	1.708	1.567	2.120	**-0.553**
/t/ in /ut/V_2 sequences	/utu/ ***vs*** /uta/	/uti/ ***vs*** /uta/	/utu/ ***vs*** /uti/	/utu/ ***vs*** /uta/	/uti/ ***vs*** /uta/	/utu/ ***vs*** /uti/
T_{2x}		-0.972	1.184			
T_{3x}	-0.689				**1.137**	**-1.051**
T_{4x}		-0.853	1.077			
T_{2y}		1.702	-1.840		1.951	-1.731
T_{3y}	0.429	1.695	-1.266	0.615	2.171	-1.556
T_{4y}	1.445	0.926	0.519	1.442	2.115	**-0.673**
/t/ in /it/V_2 sequences	/itu/ ***vs*** /ita/	/iti/ ***vs*** /ita/	/itu/ ***vs*** /iti/	/itu/ ***vs*** /ita/	/iti/ ***vs*** /ita/	/itu/ ***vs*** /iti/
T_{2x}	0.919		1.153			
T_{3x}	0.816				**0.911**	**-1.470**
T_{4x}	0.794		1.312			
T_{2y}		1.857	-1.534		1.650	-1.562
T_{3y}	0.872	1.947	-1.075	0.715	2.111	-1.396
T_{4y}	1.874	1.055	0.819	1.800	1.901	**-0.524**

Table 2: Average differences in the sensors' positions (z-scored data) of consonant /t/ between two different V_2 contexts in V_1/t/V_2 sequences. Only the significant results of the post-hoc t-tests conducted on the z-scored data of French and Chinese speakers' groups separately (see text) are shown. T_{2x}, T_{3x} and T_{4x} represent differences in the X direction, and T_{2y}, T_{3y} and T_{4y} in the Y direction. Positive differences along the X-axis indicate that the sensor is more posterior for the first logatom of the pair. Positive differences along the Y-axis indicate that the sensor is higher. The boldfaces are for the cases where the observations are in contradiction with predictions based on anticipatory strategies.

	sequences	*p* for V_2 -to-C Effect	*p* for Language×V_2 Interaction
X direction	/akV_2/	0.000 *	0.080
	/ukV_2/	0.000 *	0.072
	/ikV_2/	0.000 *	0.192
Y direction	/akV_2/	0.000 *	0.075
	/ukV_2/	0.000 *	0.563
	/ikV_2/	0.000 *	0.069

Table 3: p-values of the ANOVAs run on the sensors' positions (z-score data) of consonant /k/ in the V_1/k/V_2 sequences (see Table 1 for details)

	French speakers	Chinese speakers
/k/ in /akV_2/ Sequences	/aku/ ***vs*** /aka/	/aku/ ***vs*** /aka/
T_{2x}	1.729	1.236
T_{3x}	1.566	
T_{4x}	1.486	0.866
T_{2y}	**-1.309**	**-0.637**
T_{3y}	**-1.319**	**-0.779**
T_{4y}	**-1.367**	
/k/ in /ukV_2/ Sequences	/uku/***vs*** /uka/	/uku/ ***vs*** /uka/
T_{2x}	1.489	0.958
T_{3x}	1.311	
T_{4x}	1.197	0.914
T_{2y}	**-1.318**	**-0.948**
T_{3y}	**-1.455**	**-1.437**
T_{4y}	**-1.255**	**-1.002**
/k/ in /ikV_2/ Sequences	/iku/ ***vs*** /ika/	/iku/ ***vs*** /ika/
T_{2x}	1.816	0.973
T_{3x}	1.726	0.712
T_{4x}	1.694	0.866
T_{2y}	**-1.245**	
T_{3y}	**-1.422**	**-0.654**
T_{4y}	**-1.183**	**-0.549**

Table 4: Average differences in the sensors' positions (z-scored data) of consonant /k/ between two different V_2 contexts in V_1/k/V_2 sequences. (see Table 2 for details)

	sequences	*p* for V_2 -to- V_1 effect	*p* for Language×V_2 interaction
X direction	/at/V_2	0.013*	0.000*
	/ut/V_2	0.028*	0.047*
	/it/V_2	0.279	0.934
Y direction	/at/V_2	0.000*	0.006*
	/ut/V_2	0.002*	0.037*
	/it/V_2	0.181	0.411

Table 5: p-values of the ANOVAs run on the sensors' positions (z-score data) of V_1 in the V_1/t/V_2 sequences (see Table 1 for details)

		French speakers			Chinese speakers	
/a/ in /atV_2/ sequence	/atu/ ***vs*** /ata/	/ati/ ***vs*** /ata/	/atu/ ***vs*** /ati/	/atu/ ***vs*** /ata/	/ati/ ***vs*** /ata/	/atu/ ***vs*** /ati/
T_{2x}		-0.543	0.742		**0.502**	
T_{3x}		-0.639	0.711	0.538	**0.842**	
T_{4x}		-0.839	0.710		**0.478**	
T_{2y}	0.871	1.463	-0.592			
T_{3y}	0.893	1.481	-0.575			
T_{4y}		0.493				
/u/ in /utV_2/ sequence	/utu/ ***vs*** /uta/	/uti/ ***vs*** /uta/	/utu/ ***vs*** /uti/	/utu/ ***vs*** /uta/	/uti/ ***vs*** /uta/	/utu/ ***vs*** /uti/
T_{2x}		-0.621	0.697			
T_{3x}		-0.557	0.566			
T_{4x}		-0.667	0.744			
T_{2y}		1.267	-1.391			
T_{3y}		0.918	-1.023			
T_{4y}			-0.645			

Table 6: Average differences in the sensors' positions (z-scored data) of V_1 between two different V_2 contexts for /at/V_2 and /ut/V_2 sequences. (see Table 2 for details)

	Sequences	*p* for V_2 -to- V_1 effect	*p* for Language×V2 interaction
X direction	/akV_2/	0.150	0.836
	/ukV_2/	0.002*	0.045*
	/ikV_2/	0.079	0.068
Y direction	/akV_2/	0.010*	0.021*
	/ukV_2/	0.601	0.844
	/ikV_2/	0.098	0.799

Table 7: p-values of the ANOVAs run on the sensors' positions (z-score data) of V_1 in V_1/k/V_2 sequences. (see Table 1 for details)

	French speakers	Chinese speakers
/a/ in /akV_2 / sequence	/aku/ ***vs*** /aka/	/aku/ ***vs*** /aka/
T_{2y}	0.767	
T_{3y}	0.655	
T_{4y}	0.648	
/u/ in /ukV_2/ sequence	/uku/ ***vs*** /uka/	/uku/ ***vs*** /uka/
T_{2x}	-0.973	
T_{3x}	-0.897	
T_{4x}	-0.920	

Table 8: Average differences in the sensors' positions (z-scored data) of V_1 between two different V_2 contexts for /ak/V_2 and /uk/V_2 sequences. (see Table 2 for details)

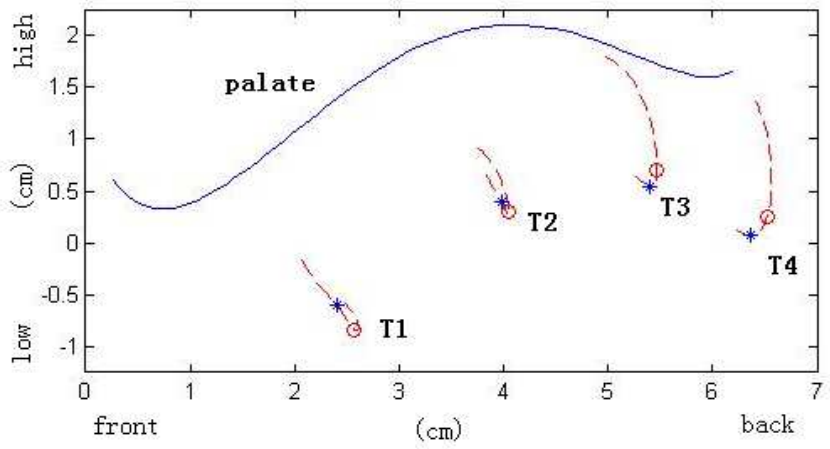
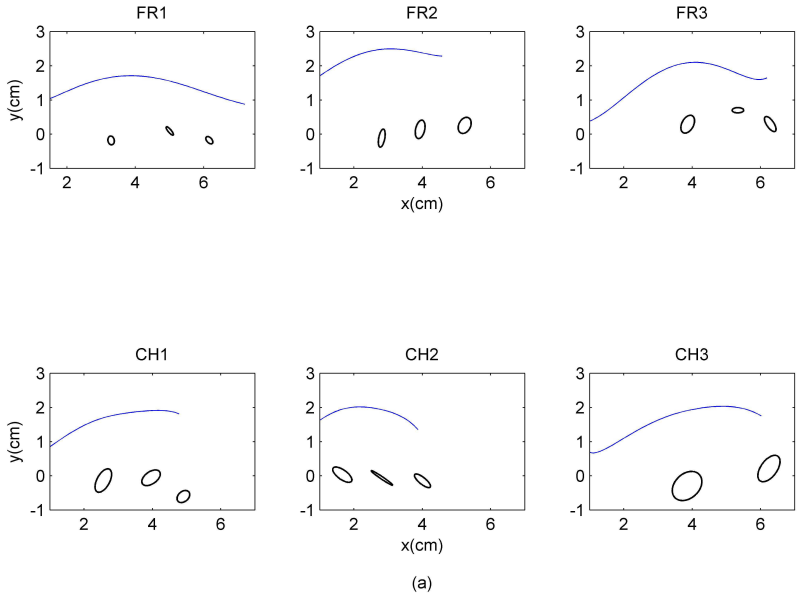
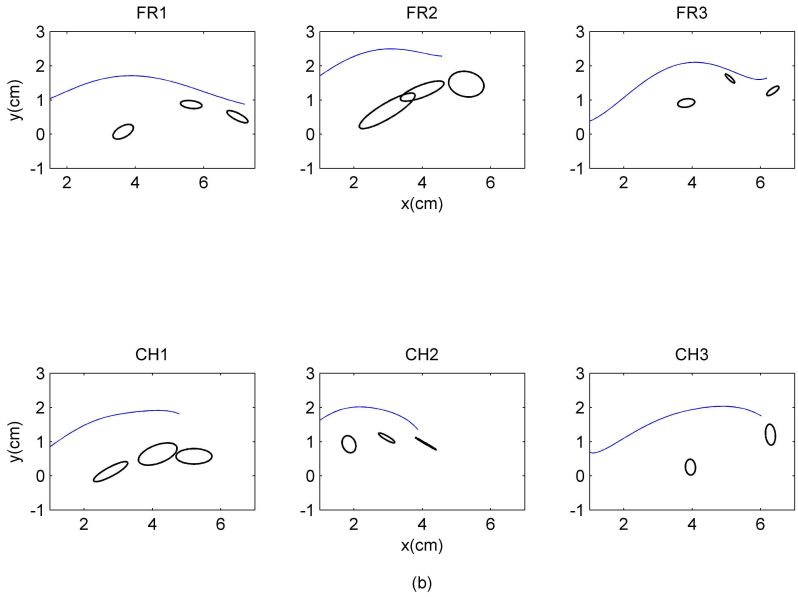

high
low
(cm)
palate
T1
T2
T3
T4
front
(cm)
back
2
1.5
1
0.5
0
-0.5
-1
0
1
2
3
4
5
6
7

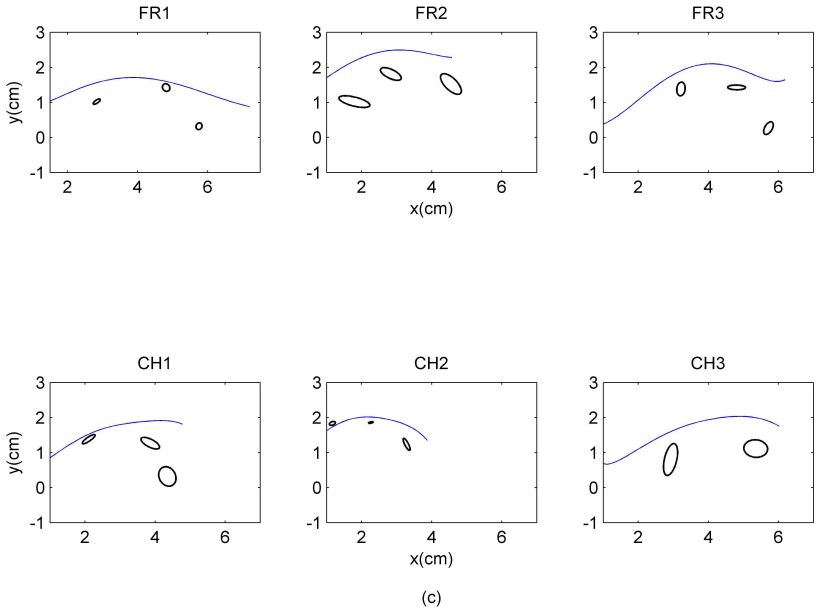
Figure 1



(a) V_1=/a/ in /ata/ sequence



(b) V_1=/u/ in /utu/ sequence



(c): V_1=/i/ in /iti/ sequence

Figure 2

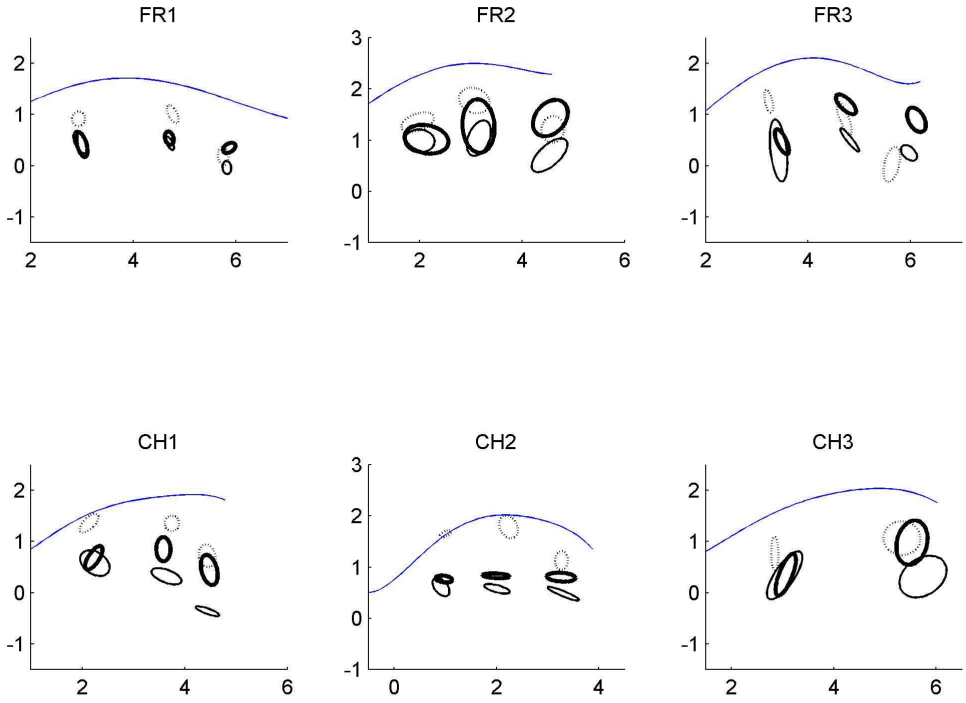


Figure 3

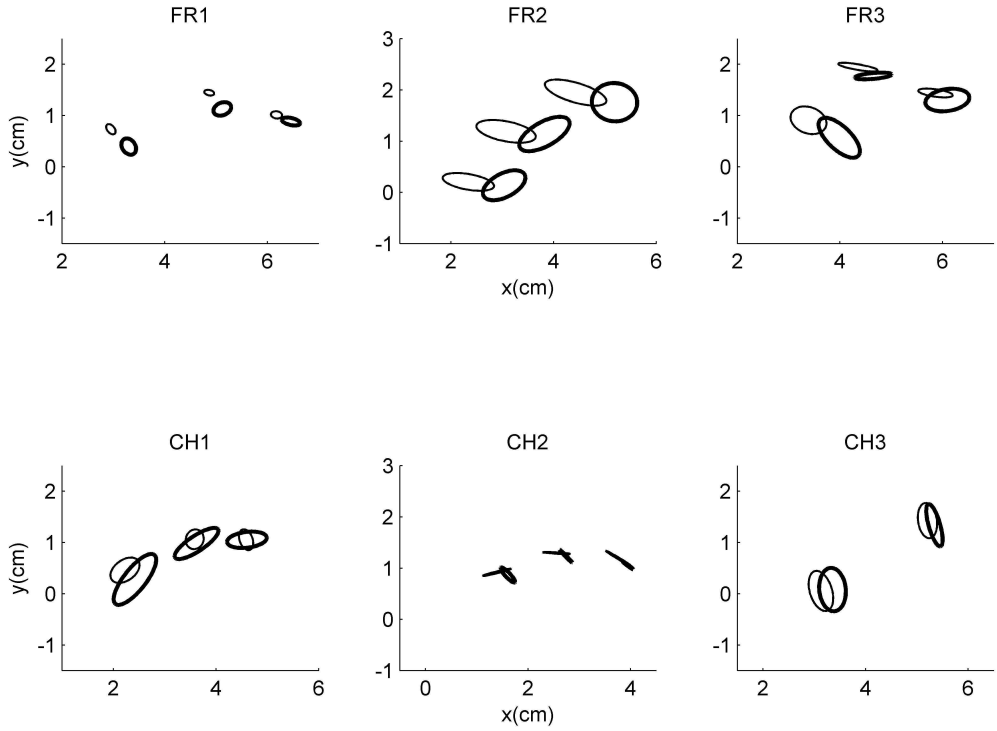


Figure 4

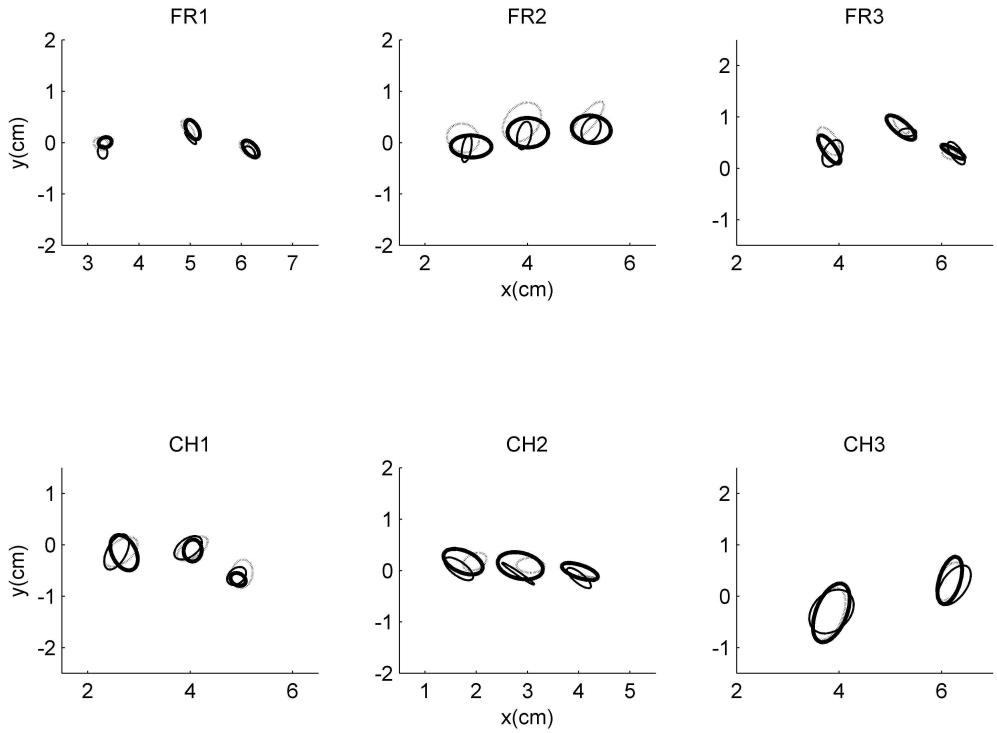


Figure 5

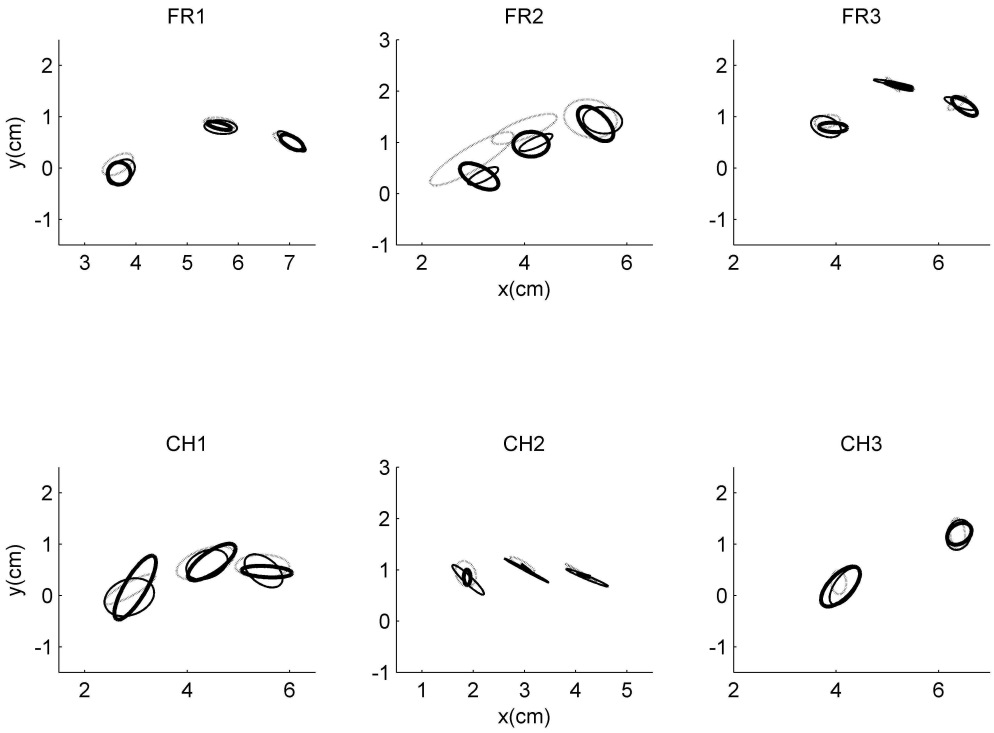


Figure 6

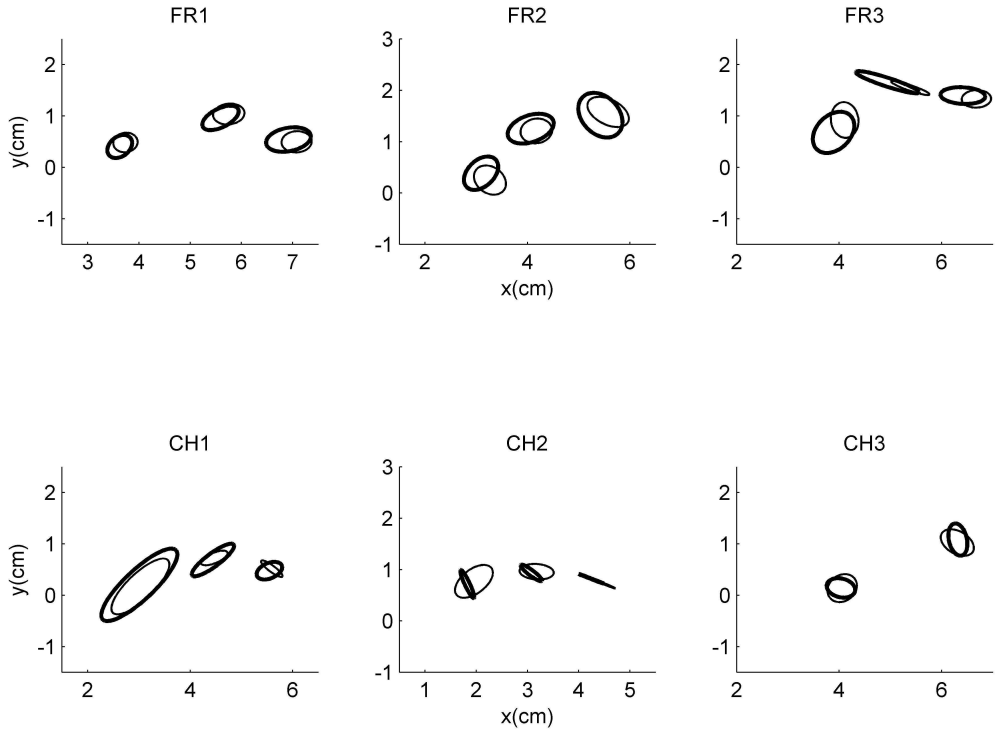


Figure 7

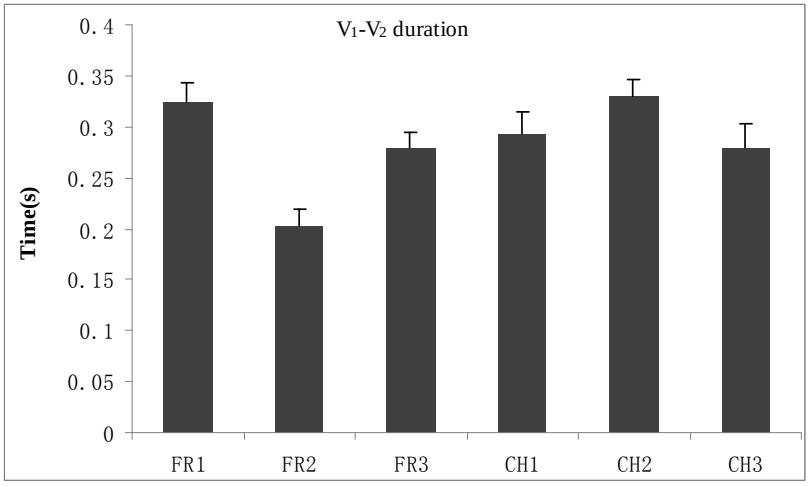

V1-V2 duration
Time(s)
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
FR1
FR2
FR3
CH1
CH2
CH3

Figure. 8

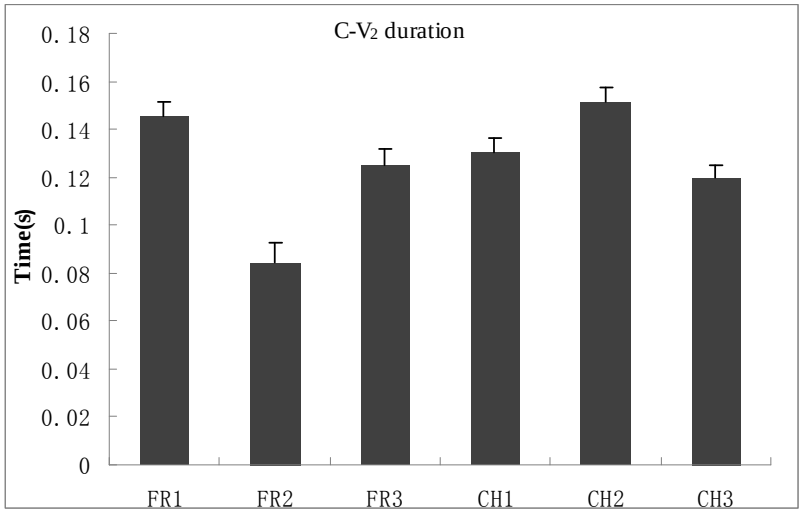

C-V2 duration
Time(s)
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
FR1
FR2
FR3
CH1
CH2
CH3

Figure 9

Appendix: ANOVAS on the “virtual language” groups.

This Appendix presents the results of the statistical analysis of the influence of the “virtual language" groups on the *extrasyllabic* anticipatory coarticulation. Nine “virtual language” groups were created by combining a French speaker (FRi, i=1, 2 or 3) and a Chinese speaker (CHj, j=1, 2 or 3) within a group. The analysis was strictly the same as the ANOVA that was carried out on the French and Mandarin Chinese speakers’ groups. The tables below present for each pair of “virtual language” groups (18 pairs in total) the p-values of the main effect of V_2 and of the Language × V_2 interaction on the sensor positions of consonant /t/ in the V_1/t/V_2 sequences, computed for each vowel V_1 separately. The p-values computed for the French and Mandarin Chinese groups are presented in Table 5.

Significant results are indicated by the symbol (*) ($p<0.05$). These tables do not reveal any significant Language x V_2 interaction except in two pair-comparisons for only one direction and one vowel V_1 (along the X-direction for /at/V_2 in the FR1 CH3 – FR3 CH1 comparison; along the Y-direction for /ut/V*2* in the *FR1 CH3 – FR3 CH2* comparison). This makes very unlikely the possibility that the Language x V_2 interaction observed in the comparison of the French and Mandarin Chinese speakers’ groups could be in fact the consequence of speaker-specific behaviors independent from the language.

FR1 CH1 – FR2 CH2

	sequences	*p* for V_2 -to- V_1 effect	*p* for Language×V_2 interaction
X direction	/at/V_2	0.011*	0.433
	/ut/V_2	0.028*	0.213
	/it/V_2	0.177	0.267
Y direction	/at/V_2	0.002*	0.155
	/ut/V_2	0.000*	0.356
	/it/V_2	0.231	0.245

FR1 CH1 – FR2 CH3

	sequences	*p* for V_2 –to- V_1 effect	*p* for Language×V_2 interaction
X direction	/at/V_2	0.034*	0.532
	/ut/V_2	0.029*	0.324
	/it/V_2	0.322	0.317
Y direction	/at/V_2	0.000*	0.433
	/ut/V_2	0.000*	0.231
	/it/V_2	0.345	0.190

FR1 CH1 – FR3 CH2

	sequences	*p* for V_2 -to- V_1 effect	*p* for Language×V_2 interaction
X direction	/at/V_2	0.035*	0.189
	/ut/V_2	0.000*	0.068
	/it/V_2	0.092	0.139
Y direction	/at/V_2	0.000*	0.075
	/ut/V_2	0.001*	0.577
	/it/V_2	0.149	0.623

FR1 CH1 – FR3 CH3

	sequences	*p* for V_2 -to- V_1 effect	*p* for Language×V_2 interaction
X direction	/at/V_2	0.005*	0.376
	/ut/V_2	0.000*	0.258
	/it/V_2	0.102	0.280
Y direction	/at/V_2	0.000*	0.560
	/ut/V_2	0.000*	0.341
	/it/V_2	0.098	0.278

FR1 CH2 – FR2 CH1

	sequences	*p* for V_2 -to- V_1 effect	*p* for Language×V_2 interaction
X direction	/at/V_2	0.011*	0.220
	/ut/V_2	0.028*	0.319
	/it/V_2	0.177	0.388
Y direction	/at/V_2	0.002*	0.240
	/ut/V_2	0.000*	0.711
	/it/V_2	0.231	0.098

FR1 CH2 – FR3 CH1

	sequences	*p* for V_2 -to- V_1 effect	*p* for Language×V_2 interaction
X direction	/at/V_2	0.035*	0.425
	/ut/V_2	0.000*	0.362
	/it/V_2	0.092	0.239
Y direction	/at/V_2	0.000*	0.141
	/ut/V_2	0.001*	0.325
	/it/V_2	0.149	0.788

FR1 CH2 – FR2 CH3

	sequences	*p* for V_2 -to- V_1 effect	*p* for Language×V_2 interaction
X direction	/at/V_2	0.025*	0.278
	/ut/V_2	0.005*	0.159
	/it/V_2	0.176	0.254
Y direction	/at/V_2	0.000*	0.340
	/ut/V_2	0.014*	0.289
	/it/V_2	0.255	0.107

FR1 CH2 – FR3 CH3

	sequences	*p* for V_2 -to- V_1 effect	*p* for Language×V_2 interaction
X direction	/at/V_2	0.017*	0.078
	/ut/V_2	0.000*	0.561
	/it/V_2	0.568	0.356
Y direction	/at/V_2	0.000*	0.728
	/ut/V_2	0.000*	0.467
	/it/V_2	0.412	0.593

FR1 CH3 – FR2 CH1

	sequences	*p* for V_2 -to- V_1 effect	*p* for Language×V_2 interaction
X direction	/at/V_2	0.034*	0.152
	/ut/V_2	0.029*	0.377
	/it/V_2	0.322	0.276
Y direction	/at/V_2	0.000*	0.198
	/ut/V_2	0.000*	0.099
	/it/V_2	0.345	0.176

FR1 CH3 – FR3 CH1

	sequences	*p* for V_2 -to- V_1 effect	*p* for Language×V_2 interaction
X direction	/at/V_2	0.005*	0.040*
	/ut/V_2	0.000*	0.143
	/it/V_2	0.102	0.560
Y direction	/at/V_2	0.000*	0.312
	/ut/V_2	0.000*	0.578
	/it/V_2	0.098	0.455

FR1 CH3 – FR2 CH2

	sequences	*p* for V_2 -to- V_1 effect	*p* for Language×V_2 interaction
X direction	/at/V_2	0.025*	0.314
	/ut/V_2	0.005*	0.222
	/it/V_2	0.176	0.189
Y direction	/at/V_2	0.000*	0.232
	/ut/V_2	0.014*	0.145
	/it/V_2	0.255	0.150

FR1 CH3 – FR3 CH2

	sequences	*p* for V_2 -to- V_1 effect	*p* for Language×V_2 interaction
X direction	/at/V_2	0.017*	0.240
	/ut/V_2	0.000*	0.506
	/it/V_2	0.568	0.310
Y direction	/at/V_2	0.000*	0.091
	/ut/V_2	0.000*	0.034*
	/it/V_2	0.412	0.709

FR2 CH1 – FR3 CH2

	Sequences	*p* for V_2 -to- V_1 effect	*p* for Language×V_2 interaction
X direction	/at/V_2	0.009	0.213
	/ut/V_2	0.005*	0.324
	/it/V_2	0.232	0.511
Y direction	/at/V_2	0.000*	0.692
	/ut/V_2	0.001*	0.751
	/it/V_2	0.063	0.437

FR2 CH1 – FR3 CH3

	sequences	*p* for V_2 –to- V_1 effect	*p* for Language×V_2 interaction
X direction	/at/V_2	0.000*	0.322
	/ut/V_2	0.000*	0.784
	/it/V_2	0.108	0.213
Y direction	/at/V_2	0.000*	0.279
	/ut/V_2	0.005*	0.656
	/it/V_2	0.063	0.096

FR2 CH2 – FR3 CH1

	Sequences	*p* for V_2 -to- V_1 effect	*p* for Language×V_2 interaction
X direction	/at/V_2	0.009*	0.122
	/ut/V_2	0.005*	0.224
	/it/V_2	0.232	0.343
Y direction	/at/V_2	0.000*	0.098
	/ut/V_2	0.001*	0.206
	/it/V_2	0.063	0.268

FR2CH2 – FR3 CH3

	sequences	*p* for V_2 -to- V_1 effect	*p* for Language×V_2 interaction
X direction	/at/V_2	0.001*	0.311
	/ut/V_2	0.000*	0.738
	/it/V_2	0.409	0.263
Y direction	/at/V_2	0.000*	0.827
	/ut/V_2	0.000*	0.287
	/it/V_2	0.320	0.467

FR2 CH3 – FR3 CH1

	sequences	*p* for V_2 –to- V_1 effect	*p* for Language×V_2 interaction
X direction	/at/V_2	0.000*	0.232
	/ut/V_2	0.000*	0.765
	/it/V_2	0.108	0.198
Y direction	/at/V_2	0.000*	0.535
	/ut/V_2	0.005*	0.426
	/it/V_2	0.063	0.166

FR2 CH3 – FR3 CH2

	sequences	*p* for V_2 -to- V_1 effect	*p* for Language×V_2 interaction
X direction	/at/V_2	0.001*	0.102
	/ut/V_2	0.000*	0.560
	/it/V_2	0.409	0.435
Y direction	/at/V_2	0.000*	0.320
	/ut/V_2	0.000*	0.156
	/it/V_2	0.320	0.726