Startling Auditory Stimulus as a Window into Speech Motor Planning

by

Cheng-hao Chiu

B.A., National Chengchi University, Taiwan, 2002
M.A., National Chung Cheng University, Taiwan, 2005

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY in

The Faculty of Graduate and Postdoctoral Studies (Linguistics)

THE UNIVERSITY OF BRITISH COLUMBIA (Vancouver)

December 2015
© Cheng-hao Chiu 2015
Abstract

While speech planning has long been a topic of discussion in the literature, the specific content of speech plans has remained largely conjectural. The present dissertation brings to this problem a methodology using startling auditory stimulus (SAS) to examine the contents of prepared movement plans unaltered by feedback regulation.

The startling auditory stimulus (SAS, > 120dB) has been found to elicit rapid release of prepared movements with high accuracy and largely unaltered EMG muscle activity patterns. Because the response latency of these SAS-triggered movements is too short to allow for feedback or correction processes, the executed movements have been used to reveal the contents of the movement plans with little or no feedback information influencing the prepared motor behaviours.

In the present dissertation, the first experiment applied this methodology to CV syllable production to test whether English CV syllables can be elicited in the same manner as other limb movements. Results show that a SAS can trigger an early release of a well-formed prepared English CV syllable, including intact lip kinematics and vowel formants. The second experiment investigated whether the observed short latency and additional lip compression are speech-specific or generic to any oral movement. Results
show that while prepared speech-like and non-speech movements are subject
to early release by SAS, lip compression does not occur as frequently as it
does in Spoken speech, suggesting that this preparatory compression may be
speech-specific, likely relating to aerodynamic factors. The third experiment
further tests whether lip compression that is independent of aerodynamic
factors is observed in all speech-related tasks and is elicited at a short latency
by SAS. Results show that comparable lip compression resulting from move-
ment overshoot was observed for both Spoken and Mouthed speech. The
fourth experiment looked into the level of suprasegmental gestures in speech
planning. The results show that while both pitch contour and formants were
maintained in the SAS-induced responses, pitch levels were compromised,
suggesting that a prepared syllable ought minimally to include phonemic
contrasts. SAS provides a useful tool for observing the contents of speech
plans.
Preface

This dissertation contains one chapter of introduction, four content chapters, and one chapter of conclusion. Three papers out of these chapters have been published in different journals and one of the chapter has been accepted for a conference proceedings. Detailed descriptions of collaboration and contribution are as follows.

Chapter 1, Section 1.2 “Neural pathways for normal and SAS-induced speech” was written by Cheng-hao Chiu, in collaboration with Dr. Bryan Gick, the dissertation supervisor. Experiments reported in Chapters 2 – 4 were designed and run in collaboration with the Motor Control and Learning Laboratories at UBC, directed by Dr. Ian Franks. Experiment design was generated by Dr. Bryan Gick, Cheng-hao Chiu, and the group from the Motor Control and Learning Laboratories, including Dr. Romeo Chua, Dr. Dana Maslovat, and Andrew Stevenson. Cheng-hao Chiu was responsible for the writing of these chapters, with collaboration from Dr. Bryan Gick, Dr. Ian Franks, and Dr. Eric Vatikiotis-Bateson. Chapter 2, “The StartReact effect in syllable production”, was included as Experiment 2 of 2 in Stevenson et al. (2014). Chapter 3, “Startling spoken, mouthed, and non-speech movements”, has been submitted to the Annual Meeting of Canadian Acoustical Association and will be published in Canadian Acoustics. Chap-
Chapter 5. “Pitch planning in English and Taiwanese Mandarin” was designed and conducted by Cheng-hao Chiu, with help from Dr. Yu-an Lu at the department of Foreign Language and Literature at National Chiao Tung University, Hsinchu, Taiwan.

Published papers included in this dissertation are listed here.


The experiments included in this dissertation were conducted with the approval of the UBC Behavioural Research Ethics Board: Preparation and Control of Movement (H09-00632) and Processing Complex Speech Motor Tasks (H04-80337). I have completed the Interagency Advisory Panel on Research Ethics Introductory Tutorial for the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS). The certificate was issued on August 22, 2008.
Table of Contents

Abstract ................................................. ii

Preface .................................................... iv

Table of Contents ........................................ vi

List of Tables ............................................. x

List of Figures ............................................ xii

Acknowledgements ........................................ xxi

Dedication ................................................ xxv

1 Introduction ............................................. 1
  1.1 Startle paradigm ...................................... 2
    1.1.1 SAS and its applications ......................... 2
    1.1.2 Simple RT tasks and the StartReact effect .... 4
  1.2 Neural pathways for normal and SAS-induced speech . 8
  1.3 Organization of this dissertation ..................... 15

2 The StartReact effect in syllable production ............ 17
3.3.1 Lip compression ........................................... 52
3.3.2 Startle indicator ........................................ 59
3.3.3 Kinematic markers, reaction times, and timing .... 61
3.4 Discussion ...................................................... 67

4 Startling syllables with pre-speech movements .......... 72
  4.1 Introduction ................................................. 72
  4.2 Methods ....................................................... 74
    4.2.1 Participants ............................................ 74
    4.2.2 Apparatus, task, and procedures .................... 74
    4.2.3 Data reduction, dependent measures, and analyses ... 75
  4.3 Results ....................................................... 82
    4.3.1 Startle indicator ...................................... 82
    4.3.2 Kinematic markers, reaction times, and timing ..... 84
    4.3.3 Lip kinematics ......................................... 88
  4.4 Discussion .................................................... 92

5 Pitch planning in English and Taiwanese Mandarin .... 94
  5.1 Introduction ................................................ 94
  5.2 Methods ..................................................... 100
    5.2.1 Participants ............................................ 100
    5.2.2 Apparatus, task, and procedures .................... 101
    5.2.3 Recording equipment .................................. 102
    5.2.4 Data preparation and statistical analyses .......... 103
  5.3 Results ..................................................... 104
  5.4 Discussion .................................................. 108

viii
6 General discussion .................................................. 112
6.1 Summaries .......................................................... 112
6.2 Theoretical implications ........................................... 114
  6.2.1 Startle paradigm as a window into speech motor plan-
       ning ............................................................... 114
  6.2.2 Forward control ................................................ 118
6.3 Conclusion and future work ....................................... 119

Bibliography .......................................................... 121

Appendices

A Chapter 3 additional figures ............................... 138
### List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Experimental results of Exp. 2, showing mean and standard deviations (in brackets)</td>
<td>36</td>
</tr>
<tr>
<td>3.1</td>
<td>Numbers of trials with lip compression across conditions. Percentages were calculated from the trials included in analyses.</td>
<td>53</td>
</tr>
<tr>
<td>3.2</td>
<td>Mean compression displacement across conditions (mm); standard deviations in parentheses.</td>
<td>54</td>
</tr>
<tr>
<td>3.3</td>
<td>Mean reaction times (in ms) to dependent markers across conditions; standard deviations in parentheses. The voluntary movement onsets of Spoken responses from Chapter 2 are included for comparison.</td>
<td>64</td>
</tr>
<tr>
<td>4.1</td>
<td>Mean reaction time (in ms) at dependent markers across conditions; standard deviations are in parentheses.</td>
<td>84</td>
</tr>
<tr>
<td>4.2</td>
<td>Mean lip compression onset latency (in ms), lip compression duration (in ms) and lip compression displacement (in mm); standard deviations are in parentheses.</td>
<td>88</td>
</tr>
<tr>
<td>5.1</td>
<td>The four lexical tones in Mandarin Chinese and possible words</td>
<td>99</td>
</tr>
</tbody>
</table>
5.2 Average acoustic burst release times and syllable durations (in ms) across 4 Taiwanese Mandarin tones. Standard deviations are shown in parentheses.
List of Figures

1.1 Information processing for simple RT tasks .......................... 5
1.2 Proposed neural circuits for control and SAS-induced speech production. The grey arrows show the pathways for control responses and the black arrows present the pathways for SAS-induced responses. The solid lines are stimulus-triggered pathways in forward control and the dashed lines represent the origin of the feedback information input. .......................... 13

2.1 Relative positions of the monitor, participant, and loud speaker. .......................... 23
2.2 Sternocleidomastoid (SCM) and orbicularis oris muscles for the attachment of surface EMG electrodes. Note the SCM and orbicularis oris figures were adapted and modified from Gray’s Anatomy (20th ed.) .......................... 25
2.3 Experimental apparatus and set-up. .......................... 26
2.4 Average rectified raw EMG traces of baseline trials, including (from top) right SCM, left SCM, upper lip, and lower lip. EMG activity was plotted with respect to the imperative stimulus (the vertical grey line) .......................... 28
2.5 EMG activity, lip displacement and response acoustic waveforms of control exemplar. The EMG measure includes (from top) right SCM, left SCM, upper lip, and lower lip. EMG activity was rectified and plotted with respect to the imperative stimulus (the vertical grey line). Lip displacement and response acoustic waveforms are also plotted with respect to the imperative stimulus. Point A marks the beginning of voluntary movement; point B marks the time to the lower lip opening onset; point C marks the lowest position of the lower lip. The intervals across kinematic markers were labeled from $t_1$ to $t_5$. See text for details.

2.6 Mean EMG activity and kinematic displacement of control (grey) and startle (black) trials in the vocalization condition with respect to the imperative stimulus. The top four channels represent the mean EMG across four muscles (from top to bottom): left SCM, right SCM, upper lip, and lower lip. The bottom two channels depict the mean upper lip and lower lip displacement trajectories, respectively.

2.7 Mean EMG activity and kinematic displacement of control (grey) and startle (black) trials in the vocalization condition with respect to the lower lip opening onset. The top four channels represent the mean EMG across four muscles (from top to bottom): left SCM, right SCM, upper lip, and lower lip. The bottom two channels depict the mean upper lip and lower lip displacement trajectories, respectively.
2.8 SS ANOVA comparison of formant frequencies F1 and F2 over syllable /ba/ durations. The black line denotes the predicted fit for control responses and the white line denotes the startle responses. The dark grey and light grey bands surrounding the predicted fits represent a 95% confidence interval. Any white space between the transition lines for each condition represents a statistically significant difference between the given measures.

3.1 Lip displacement and response acoustic waveforms are also plotted with respect to the imperative stimulus. Point A marks the beginning of voluntary movement; point B marks the time to the lower lip opening onset; point C marks the lowest position of the lower lip. The intervals across these markers were labeled as $t_1$, $t_2$, and $t_3$. See texts for details.

3.2 Mean upper lip (left) and lower lip (right) EMG activity and displacement trajectories in the Spoken condition across participants. Data in all channels were normalized to the lower lip opening onset (vertical line). The distance of lip displacement between control and startle responses does not reflect the absolute distance between control and startle responses.
3.3 Mean upper lip (left) and lower lip (right) EMG activity and displacement trajectories in the Mouthed condition across participants. Data in all channels were normalized to the lower lip opening onset (vertical line). The distance of lip displacement between control and startle responses does not reflect the absolute distance between control and startle responses.

3.4 Mean upper lip (left) and lower lip (right) EMG activity and displacement trajectories in the Non-speech condition across participants. Data in all channels were normalized to the lower lip opening onset (vertical line). The distance of lip displacement between control and startle responses does not reflect the absolute distance between control and startle responses.

3.5 Mean EMG activity and kinematic displacement of control (grey) and startle (black) trials in the Spoken (repeated from Chapter 2), Mouthed, and Non-speech condition. All data were normalized to the imperative stimulus. The top four channels represent the mean EMG across four muscles (from top to bottom): left SCM, right SCM, upper lip, and lower lip. Each channel is of the same scale but vertically arranged for visualization. The bottom two channels depict the mean upper lip and lower lip displacement trajectories, respectively. Red boxes mark the window of startle reflex activity, ranging from 30 ms to 120 ms after the onset of the imperative stimulus.
3.6 Relative timing ratio (non-transformed) of various kinematic time markers for startle and control trials. The whiskers for each bar are standard error bars. The time frame for these events is the time between lower lip voluntary movement onset and lower lip displacement trough \( t_3 \). Time intervals include the time between voluntary movement onset and lower lip opening onset \( t_1 \) and between lower lip opening onset and lowest lower displacement trough \( t_2 \). 66

4.1 Schematic Spoken response with lip displacement and response acoustic waveforms plotted with respect to the imperative stimulus. Point A marks the beginning of voluntary movement; point B marks the time to the lower lip opening onset; point C marks the acoustic onset; point D marks the lowest position of the lower lip. The intervals across these markers were labeled as \( t_1, t_2, t_3, t_4, \) and \( t_5 \). See text for details. 76

4.2 Lip compression markers. Points M and N mark the onset and offset of lip compression, respectively (see text for details). Lip compression is here defined as the latency between Points M and N \( (t_c) \). Lip compression displacement measures the change in lower lip displacement within \( t_c \). 78
4.3 Schematic lip displacement trajectories (top) from the upper and lower lips and correspondent lip aperture profile (bottom). Data was taken from one trial from one of the participants. The two lips reach an equilibrium point and lip aperture stabilizes. Data was taken from one example trial from a participant. The vertical grey line marks the onset of lip compression. See text for details.

4.4 Schematic lip displacement trajectories (top) from the upper and lower lips and correspondent lip aperture profile (bottom). Data was taken from one trial from one of the participants. This figure shows that lip aperture continues to decrease after the two lips make contact. The vertical grey line marks the onset of lip compression. An inflection of lip aperture profile is observed. See text for details.

4.5 Mean EMG activity and kinematic displacement of control (grey) and startle (black) trials in the Spoken, Mouthed, and Non-speech condition. All data were normalized to the imperative stimulus. The top four channels represent the mean EMG across four muscles (from top to bottom): left SCM, right SCM, upper lip, and lower lip. Each channel is of the same scale but vertically arranged for visualization. The bottom two channels depict the mean upper lip and lower lip displacement trajectories, respectively. Red boxes mark the window of startle reflex activity, ranging from 30 ms to 120 ms after the onset of the imperative stimulus.
4.6 Relative timing ratio (non-transformed) of various kinematic and acoustic time markers for startle and control trials. The time frame for these events is the time between lower lip voluntary movement onset and lower lip displacement trough \((t_5)\). Time intervals include the time between: voluntary movement onset and lower lip opening onset \((t_1)\), lower lip opening onset and voice onset \((t_2)\), voluntary movement onset and voice onset \((t_3)\), and lower lip opening onset and lower lip displacement trough \((t_4)\).

4.7 Upper lip (left) and lower lip (right) EMG activity and displacement trajectories in the *Spoken* condition. Data in all channels were normalized to the lower lip opening onset (vertical line).

4.8 Upper lip (left) and lower lip (right) EMG activity and displacement trajectories in the *Mouthed* condition. Data in all channels were normalized to the lower lip opening onset (vertical line).

4.9 Upper lip (left) and lower lip (right) EMG activity and displacement trajectories in the *Non-speech* condition. Data in all channels were normalized to the lower lip opening onset (vertical line).
5.1 SS ANOVA comparison of formant frequencies F1 and F2 over syllable /ba/ durations. The black line denotes the predicted fit for control responses and the white line denotes the startle responses. The dark grey and light grey bands surrounding the predicted fits represent a 95% confidence interval. Any white space between the transition lines for each condition represents a statistically significant difference between the given measures. NB: The figure is a revised version of Figure 2.8.

5.2 Schematic timeline of potential feedback for SAS-elicited speech

5.3 Taiwanese Mandarin F1 and F2 range for control and startle responses. SS ANOVA results of F1 and F2 frequency range across 4 tones. Each shaded band represents a 95% confidence interval. The white space between the transition lines for each condition represents a statistically significant difference between the given measures.

5.4 SS ANOVA comparison of pitch over English /ba/ durations. Each shaded band represents a 95% confidence interval. The white space between the transition lines for each condition represents a statistically significant difference between the given measures.

5.5 SS ANOVA results of pitch profiles across 4 Taiwanese Mandarin tones. Other legends follow Figure 5.3.
A.1 Number of trials with lip compression in Mouthed condition. The bars in black represent trials occurring in the first half of the block; the grey bars represent trials occurring in the second half of the block. 139

A.2 Number of trials with lip compression in Non-speech condition. The bars in black represent trials occurring in the first half of the block; the grey bars represent trials occurring in the second half of the block. 139
Acknowledgements

I am immensely grateful to get this far. A moment like this would not be possible were it not for all the support that I received from people around me during my Ph.D. years. Here I am giving my sincere thanks.

First and foremost, my deepest gratitude goes to my advisor, Bryan Gick, for his guidance through every step of this dissertation work and for his being a mentor for my academic career as well as an inspiring friend for life. My perspectives on research and on life have been brought to a whole new level based on the fellowship we built together. Thank you, Bryan, for your unwavering support and encouragement, particularly during the most difficult times.

I am also enormously grateful to Dr. Ian Franks from UBC Kinesiology for all his patient guidance, responsive communication, and selfless support. Working with Dr. Franks truly broadened my view of science and of knowledge, particularly from a perspective other than linguistics. His profound understanding of motor control and his careful attitude to research are the two things that I look up to the most. Thank you, Dr. Franks, for showing me how to become a scientist. I expect myself to shout out “Eureka!” to you in return one day in the future.

I am also thankful to Eric Vatikiotis-Bateson, my other supervisory com-
mittee member. Thank you for sharing your thoughts and providing alternative perspectives to my research, especially through the last few years of my writing process. Your insights to data and research questions are more than valuable. I cannot thank you enough for all your input to my dissertation.

During the final stage of this dissertation write-up, precious comments and suggestions came from my university and external examiners: Todd Handy, Carla Kam, and Philip Hoole. Thank you all for your helpful comments and suggestions on my work.

Throughout my time at UBC, my home lab, the Interdisciplinary Speech Research Laboratory (ISRL), undoubtedly provided me the strongest supports. In particular, I would like to give my most sincere thanks to Donald Derrick and Mark Scott, for their inspiring research and insights. I was truly privileged to work with you. Wonderful people I met over the past few years, including Scott Moisik, Noriko Yamane, Murray Schellenberg, Jen Abel, and Alexis Black, thank you for all the supports and your faith in me.

My connection to the Department of Kinesiology at UBC was also an unforgettable experience. Romeo Chua, Dana Maslovat, Andrew Stevenson, Chris Forgaard, and Jarrod Blinch, you have all shown me how to approach a discipline outside linguistics. Here I would like to show my gratitude for you help and sharing. Thank you all!

Over these years at UBC, the Department of Linguistics not only ignited my enthusiasm toward research but also accommodated me both academically and personally. My special thanks to all of the UBC Linguistics faculty, staff, and graduate fellows. Molly Babel, Strang Burton, Gunnar Hansson, Lisa Matthewson, Amanda Miller, Douglas Pulleyblank, Hotze Rullmann,
Joseph Stemberger, and Martina Wiltschko, thank you for your direction and inspiration. Thank you, Sonja Thoma, Meagan Louie, Joel Dunham, Joash Johannes, for the supports and covers during my life trough. My gratitude also goes to Edna Dharmaratne and Shaine Meghji for their hearty help, supports and kindness over these years. Your efforts on administrative matters made my life at UBC home. Thank you!

I would also like to give my thanks to some scholars that I have worked with for the past few years: Sid Fels, Ian Stavness, Peter Anderson, Ling Tsou, Cormac Flynn, Andrew Ho, Jonathan Evans, Jackson T.-S. Sun, and Yuwen Lai. Thank you for including me in you projects and collaboration. My appreciation to all your supports and guidance cannot be overlooked. This dissertation and my research alongside involve great deal amount of laboratory work. This dissertation work wouldn’t be finished as smoothly as it was without the help from these laboratories. I would like to give my thanks to the laboratories: the ISRL (UBC), the Motor Control and Learning Laboratory (UBC), the Psycholinguistics Laboratory (NCTU, Taiwan), and the Experimental Phonology Laboratory (NCTU, Taiwan) and their directors who had kindly offered the space and help for my research. Thank you everyone.

Many thanks to the National Institute of Health Grant DC-02717 awarded to Haskins Laboratories and to a Discovery Grant from the Natural Sciences and Engineering Council of Canada (NSERC) awarded to Bryan Gick to support my research throughout these years.

My journey through Ph.D. cannot be as fruitful if there had not been friends, cohorts, and colleagues sharing all my highs and lows. Thank you,
Amélia Reis Silva, Analía Gutiérrez, Raphael Girard, Masaki Noguchi, Mark Scott, Anita Szakay, and Beth Stelle. Your support and the time we spent together have become an unforgettable chapter of my life. I will always remember our laughs and tears together. They made my Ph.D. stories. My thanks also go to my caring Taiwanese friends: Rae (林瑞屏), Sally (潘昭穎), Chiang (江佳陽), Justin (林昱廷), and John (吳宗聖). The true friendship we share together is one of the most valuable things I obtained from UBC. Thank you all from the bottom of my heart.

I am much blessed by the love and support from my beloved family. My gratitude to all of you can never be overemphasized.

All the remaining errors in this work are my full responsibility.
Dedication

To my mother in heaven

Mom, everything good about me is from you.
Hope this makes you proud.

媽，謝謝您成就我的夢想
希望您感到驕傲
Chapter 1

Introduction

When people are startled, a series of physiological reflexes can be observed: we shut our eyes, shrug our shoulders, and grimace. These startle reflexes are the result of a protective mechanism. Interestingly, if people are startled when they are engaged in a prepared movement, this prepared movement will be performed at a much shorter latency than a normal voluntary movement (e.g., Carlsen et al., 2004b; Siegmund et al., 2001; Valls-Solé et al., 1999). The rapid release reflex opens a window that permits researchers to investigate how people plan actions, and in particular complex actions such as speech. This dissertation employs the startle paradigm to explore speech motor planning.

While speech motor control has attracted a lot of attention over the past decades, only few studies have offered insight into specific motor content in a speech plan. The content of such a plan should presumably specify, at minimum, those aspects of speech that are essential in determining linguistic contrast, though it may presumably omit many aspects of a physical speech utterance that may be determined by production or feedback mechanisms. Recent speech research focuses on forward models (Pickering and Garrod, 2013; Scott et al., 2013; Tian and Poeppel, 2012), expanding on a
long tradition of work in programming of speech motor plans (e.g., Keele, 1981; Klapp, 2003; Lashley, 1951). These works have attempted to uncover the detailed content of forward plans by assessing behavioral performances and neuroimaging results. However, the plans revealed in these studies were not investigated using methods that allow the forward plans to be observed independently from feedback control. This dissertation uses an experimental paradigm with a startling auditory stimulus (the “startle” paradigm; see below for details) to elicit prepared speech production. This paradigm is considered a reliable method to elicit prepared responses while only limited or no feedback information is in play. The content of such plans should presumably specify, at minimum, those aspects of speech that are essential for determining linguistic contrasts. This paradigm would thus allow researchers to reveal some of the details in speech plans before any feedback adjustments.

This chapter will first introduce the startle paradigm and then provide an overview of neural pathways for SAS-induced responses. A series of experiments using this startle paradigm conducted to investigate speech motor control are described in the subsequent chapters. The overall structure of this dissertation is included in the final section of the chapter.

1.1 Startle paradigm

1.1.1 SAS and its applications

Studies over the past four decades have illustrated that the response in a simple reaction time (RT) task may be programmed prior to the in-
troduction of the imperative stimulus (Klapp, 1977, 1995, 2003; Wadman et al., 1979). The presentation of the imperative stimulus then triggers the initiation of the prepared response. In limb movement studies, a startling acoustic stimulus (SAS, >120 dB) has been shown to be an effective trigger of prepared movements (Carlsen et al., 2012; Valls-Solé et al., 1999). Prepared responses to a SAS are performed at shorter latencies than normally triggered responses, while the kinematic variables and EMG characteristics of prepared responses remain largely unchanged; for instance, a SAS was found to trigger a prepared arm movement after a time lapse as small as 70 ms after the stimulus onset (e.g., Carlsen et al., 2004b; Valls-Solé et al., 1999). This early release of prepared action in response to a SAS, termed the StartReact effect, has been shown to produce a high accuracy rate with unmodified muscle activity profiles (Valls-Solé et al., 2008, 1999). Valls-Solé et al. (2008) hypothesized that the StartReact effect bypasses feedback input because the response latencies produced are too short to accommodate the usual cortical processing for voluntary movements. Because of their very short onset latency, SAS-induced actions may be fully executed before they are affected by sensory feedback, thus offering the researcher a window onto the unmodified plan.

The largely unaltered muscle activity patterns observed in startle trials suggest that the SAS-induced release is not simply the result of a startle reflex superimposed on a faster voluntary movement. Rather, the early release of the voluntary movement indicates response triggering at a subcortical level, since the reaction times observed are too short to allow for any cortico-cortical transfer of activity (cf. Carlsen et al., 2004b; Valls-Solé
et al., 1999). Carlsen et al. (2009a) proposed that the voluntary and startle systems interact at the supra-spinal level (in the reticular formation in the brainstem), where information for the prepared movements is stored. A comparison of finger and arm movements in Carlsen et al. (2009a) shows that movements involving subcortical pathways are more susceptible to early initiation by a SAS, while movements that are more strongly mediated by cortico-spinal connections, such as finger lifting movements, are less sensitive to SAS triggering. These findings suggest that the StartReact response involves a rapid triggering of subcortically stored information.

Recent studies however, have provided evidence for cortical activation during startle, suggesting that not all programming is stored and triggered at subcortical levels. Transcranial magnetic stimulation (TMS) studies have demonstrated that, when stimulation is applied to the motor cortex together with a SAS, the resulting cortical silent period delays the StartReact response in startle trials (Alibiglou and MacKinnon, 2012; Stevenson et al., 2014). If only subcortical processes were involved, TMS should not affect the StartReact response; the observed delay in the StartReact response therefore suggests that the pathways involved in the StartReact response are mediated by, rather than bypass, cortical areas.

1.1.2 Simple RT tasks and the StartReact effect

In the motor control literature, a lot of attention has been paid to response preparation and execution. From the viewpoint of information pro-

\footnote{Exp. 2 of 2 in Stevenson et al. (2014) is included below as Chapter 2 of this dissertation.}
cessing, a movement response is invoked through a series of processes, from the input of the imperative stimulus to the output of motor behaviour. A simple reaction time (RT) task with a single target response can be used to test the StartReact effect. In this task, participants prepare the response in advance and only respond to the imperative stimulus, thus ensuring that the programming of the response occurs prior to the stimulus presentation (Figure 1.1).

Figure 1.1: Information processing for simple RT tasks

Klapp (1995, 2003) demonstrated that in a simple RT task, the reaction time needed to initiate the prepared speech act is only sensitive to the number of response elements (e.g., the number of repetitions), not to the complexity of the response (e.g., the number of syllables). When a speech response has been prepared in advance of a stimulus, the internal complexity of that response does not affect the reaction time. Klapp (1995, 2003) argued, based on these data, that speech motor plans are determined by two different processes: the internal (INT) structure and the sequence (SEQ) structure. In a simple RT task, the complexity of a prepared speech sequence (INT) does not affect reaction time, since the internal structure of the response (including duration) is specified before the stimulus. Only the number of chunks (SEQ) affects the RT in this type of task. In the context of the SAS design, it is assumed that in a simple RT task, the internal structure of a response is pre-specified before the stimulus, and these
pre-specified details, including kinematics and acoustics, are then subject to rapid release by the presentation of a SAS. However, Klapp only considers the INT and SEQ processes at the level of segments and syllables. It is not clear whether phonemic features, pre-speech postures, and suprasegmental information can also be pre-specified or organized in a speech plan. The present thesis intends to bridge this gap by using the startle paradigm to examine the planning of syllable production.

In a simple RT task using SAS, the target response is prepared in advance, while the imperative stimulus may be either a control tone or an unpredictable SAS. As the selection and programming of responses occurs before the stimulus presentation, the prepared upper limb response can be released by a SAS at a short response latency (Carlsen et al., 2004b; Valls-Solé et al., 1999). When participants are startled, reflexive EMG activity can be observed in the bilateral sternocleidomastoid (SCM) muscles; such activity is considered to be a reliable indicator of the startle reflex (Carlsen et al., 2011). Such reflex activity (pre-motor time shorter than 70 ms) is consistently observed when a SAS is presented to participants. In addition to this reflexive muscle activity, the StartReact effect is characterized by rapid release of a prepared response with unaltered kinematics (e.g., Carlsen et al., 2012; Valls-Solé et al., 1999).

The StartReact effect has also been observed in experiments on head-turning movements (Oude Nijhuis et al., 2007; Siegmund et al., 2001), eye movements (Castellote et al., 2007), and stepping (MacKinnon et al., 2007). In addition to the early release of the prepared response, the SAS can also cause the release of independently prepared motor events. For example,
Forgaard et al. (2013) used SAS to elicit prepared upper limb movements, which typically display an agonist ($AG_1$) – antagonist ($ANT$) – agonist ($AG_2$) tri-phasic EMG pattern. While the tri-phasic activation pattern in limb movements is traditionally understood to be a temporally bound movement series, the observed spatiotemporal complexity between agonist and antagonist muscles suggests that these targeted movements in fact result from coordination of muscle synergies (d’Avella and Bizzi, 2005): muscle synergies that are associated with the same functional task may rely on time-varied coordination, as opposed to a time-fixed clock. Forgaard et al. (2013) showed that, when a SAS is presented at the onset of $AG_1$ during a prepared a short elbow extension movement ($30^\circ$), the $ANT$ and $AG_2$ are released at a shorter latency. The fact that a SAS causes early elicitation of $ANT$ and $AG_2$ suggests that agonist and antagonist forces may be independently programmed by a SAS. These results indicate that the SAS paradigm can be used to determine whether serial/concurrent motor behaviours associated with a common functional purpose are independently programmed and whether or not their initiation is time-locked.

The SAS paradigm has been applied to both cortically and subcortically determined processes. As noted in the paragraphs above, an ongoing debate exists concerning whether a SAS triggers only subcortically stored programs (Carlsen et al., 2004a; Castellote et al., 2012; Nonnekes et al., 2014; Valls-Solé et al., 1999), or whether it also affects cortically processed motor behaviours (Alibiglou and MacKinnon, 2012; Stevenson et al., 2014). This dissertation seeks evidence to test first whether the StartReact effect does occur in cortically involved processes. SAS triggering may also be an effec-
tive method for probing motor behaviours with heavy cortical involvement (see Section 1.2), such as speech. In this case, the pathways for the StartReact response would be mediated by, rather than bypass, the cortical areas. In Section 1.2, I summarize current scholarship on neural communication for voluntary and SAS-induced speech production.

1.2 Neural pathways for normal and SAS-induced speech

Two types of voluntary vocalization can be identified, with differentiated neural pathways: low-level, innate (nonverbal or emotional) vocalization, and speech vocalization (Jürgens, 2002; Simonyan and Horwitz, 2011). As summarized in Jürgens (2002), nonverbal vocalization is initiated from the anterior cingulate cortex. The periaqueductal gray receives the projection from the anterior cingulate cortex and functions as a relay station by exerting activations on the reticular formation in the lower brainstem. Voluntary speech, by contrast, involves a wider and more complicated neural communication relay across cortical and subcortical areas. Speech motor preparation, storage, and initiation rely on transcortical communications, which are mediated by cortico-spinal connections.

One of the techniques most commonly used to identify cortical areas associated with a particular motor behaviour is transcranial magnetic stimulation (TMS). When TMS is applied through the scalp to the cortical area, cortical activity is paused for a very short period of time. A quiet period of EMG activity is observed after the presentation of TMS, resulting in a
delay of the intended motor response. Terao et al. (2001) performed TMS over the cortex during the preparation of vocalization in response to a visual cue. Their results show that RT of vocalization is delayed when TMS is delivered up to 150 ∼ 200 ms before the expected onset of vocalization. Similarly, Schuhmann et al. (2009) showed that object-naming utterances can be delayed when TMS is applied 300 ms after the presentation of a picture prompt.

While TMS-induced delay provides evidence for the involvement of cortical processing in speech production, fMRI studies also reveal strong evidence of cortical involvement for speech preparation and initiation. Riecker et al. (2005) investigated the role of the cortex in the production of syllable repetition. A blood-oxygen-level dependent (BOLD) contrast measurement revealed substantial activations in the left supplementary motor area (SMA), left dorsolateral prefrontal cortex (including Broca’s area), left anterior insula, and right superior cerebellum 3 ∼ 5 seconds after the onset of acoustic stimulation (i.e., clicks). On the other hand, the left sensorimotor cortex, left thalamus, left putamen/pallidum, left caudatum, and right inferior cerebellum achieved their peak activations 8 ∼ 9 seconds after the onset of speech production. The segregation of activations across these areas suggests two levels of speech motor control. The areas showing early BOLD signal changes are responsible for motor preparation, whereas the areas with late BOLD signal changes account for the execution process. Brendel et al. (2010) employed a similar design to investigate the time course of activation of the mesio-frontal cortex during the preparation and execution of syllable repetitions. A widespread network, including the brainstem, thalamus,
basal ganglia, SMA proper, inferior frontal gyrus (IFG; i.e., Broca’s area), and insula, showed significant activations for both motor preparation and execution (see Brendel et al., 2010, for details). In particular, extensive activations in SMA, IFG, and insula suggest that these areas are involved in both the preparation and execution of syllable repetitions.

Once prepared, speech motor commands are then delivered via a descending pathway through the putamen and substantia nigra to the parvo-cellular reticular formation, which is directly connected to the phonatory motoneurons in the medulla oblongata (Iwata et al., 1996; Jürgens, 2002; Simonyan and Horwitz, 2011). It should be noted that the initiation of prepared speech hinges on cortically determined processes and cannot be released simply by reticulo-spinal commands. When participants produce a pre-cued syllable sequence in response to a GO signal (as opposed to a NO GO signal), more significant activity is observed in the primary and somatosensory cortices, the superior temporal plane, the anterior insula, and the medial premotor areas, with particularly rich activity in the SMA near the superior convexity and portions of the pre-SMA and anterior cingulate sulcus (Bohland and Guenther, 2006). A main effect of overt speech (i.e., GO trials) suggests that these cortical areas are associated more with initiation than with planning or sequence buffering of the prepared speech.

The above studies suggest that the preparation of intended speech involves a wide range of cortical and subcortical components; in particular, the initiation and execution of prepared speech depends on the exertion of cortical projections to subcortical areas via a descending cortico-subcortical pathway.
The fact that presentation of a startling stimulus can cause rapid release of a response suggests that responses can be prepared ahead of time by the relevant systems and stored (or kept on hold) somewhere before the release. The debate about the storage location is still ongoing. While some SAS studies argue for subcortical storage based on the rapid release triggered by a SAS (Carlsen et al., 2004a; Castellote et al., 2012; Nonnekes et al., 2014; Valls-Solé et al., 1999), TMS data show that a SAS can elicit cortically prepared responses via a subcortically mediated pathway; therefore, the StartReact effect is also seen in cortically processed responses (Alibiglou and MacKinnon, 2012; Stevenson et al., 2014).

The storage of phonological representations was discussed by Baddeley et al. (1984), who demonstrated that the articulatory loop comprises subsystems for both storage and rehearsal. While initial storage may decay rapidly, rehearsal keeps providing updates (Baddeley, 1998; Baddeley et al., 1984; Jacquemot and Scott, 2006). Neuro-imaging data also support the notion of continuous updates provided by rehearsal (e.g., Chein and Fiez, 2001), which continually feed information to the storage. The encoding, storage, and rehearsal of phonological representations is believed to take place in a wide range of cortical areas, including the SMA, premotor cortex, and bilateral inferior frontal regions (see e.g., Chein and Fiez, 2001).

Now, turn to the neural pathways for SAS-induced responses. As summarized in Carlsen et al. (2012), the StartReact response is mediated via an ascending thalamo-cortical pathway, activated by exertion of the reticular formation on the thalamus. The neural pathways used during the StartReact response are considered to be separate from the normal pathways used
for voluntary movements. Increased activation in the thalamus provides
input to the primary motor cortex to initiate the cortically prepared move-
ment via a descending cortico-spinal pathway. It is noteworthy that the
StartReact pathways for upper limb movements largely overlap with the
pathways for speech production. Similarly, speech production may also rely
on thalamo-cortical circuits and a descending cortico-spinal pathway. Specif-
ically, after receiving input from the cerebellum, the thalamus projects to the
primary motor cortex and to Broca’s area; commands are mediated via the
putamen and reticular formation and sent down to the phonatory motoneu-
rons in the spine (Guenther et al., 2006; Iwata et al., 1996; Jürgens, 2002).
Given that speech production and upper limb movements both involve sim-
ilar thalamo-cortical pathways, it is likely that upper limb movements and
speech movements is hypothesized also share the same StartReact pathways
when elicited by a SAS. Shorter reaction times in StartReact responses are
accounted for by increased neuron activity, causing the response signal to
reach the initiation threshold more quickly (see Carlsen et al., 2012, for
details).

Figure 1.2 depicts the neural pathways for control and SAS-induced
speech production. The speech plan is encoded and prepared primarily
in the cortical areas (the grey circle), with the perceptual input and feed-
back information initiating from other cortical areas and the cerebellum (the
grey dashed lines inputting to the encoded program). While waiting for the
imperative stimulus, the encoded program can be rehearsed and constantly
updated (the black circle). When the imperative stimulus is presented, the
most updated motor program is released. As summarized in Carlsen et al.
Figure 1.2: Proposed neural circuits for control and SAS-induced speech production. The grey arrows show the pathways for control responses and the black arrows present the pathways for SAS-induced responses. The solid lines are stimulus-triggered pathways in forward control and the dashed lines represent the origin of the feedback information input.

(2012), the SAS-induced response traverses an ascending pathway from the cochlear nucleus, through the inferior colliculus, reticular formation, and thalamus, to the motor cortex (the black arrows). The released prepared program then travels along the descending motor pathways (the black arrows). For speech production, the commands of the prepared program can be released via the reticular formation and sent down to the spinal motoneu-
rons. If the speech performance can be initiated at a short latency with limited feedback information, this suggests that the triggering for speech and limb movements shares a similar subcortically mediated pathway.

We can conservatively calculate the time required for a speech response by following the same procedure used to calculate the required time span for voluntary limb movements. First, Schroeder and Foxe (2002) reported a response latency of $10 \sim 25$ ms from the onset of the auditory stimulus to the activation in the auditory cortex. Another $5 \sim 10$ ms are required for conduction between the lateral lemniscus and the thalamus (Stockard et al., 1977) for auditorily evoked responses. Transcortical and thalamus-primary motor cortex transmissions require $2 \sim 4$ ms for conduction (Carlsen et al., 2012; Guenther et al., 2006). Finally, the latency of the orofacial muscle EMG response to TMS on the face area of the motor cortex is about $11 \sim 12$ ms (Meyer et al., 1994), and the motor time for the muscle movement is delayed by $30$ ms. Adding these values together yields a minimum of $58 \sim 81$ ms in response to a SAS. As reported in Stevenson et al. (2014), the onset of SAS-induced responses occurs at about $75$ ms, suggesting the presence of a StartReact effect for speech movement. As the results reported in the following chapters will demonstrate, voluntary movement onsets fall within the predicted range for a StartReact response, suggesting that the release of the prepared syllable undergoes the proposed StartReact pathway.

To sum up this section: neural correlates and pathways for SAS-induced responses support the view that SAS-induced speech responses may contain unaltered details of speech plans, providing researchers with a window into speech motor planning that bypasses afferent feedback information.
1.3 Organization of this dissertation

The main purpose of this dissertation is to use the SAS paradigm to investigate how speakers construct their speech plans in speech motor control. Specifically, a series of experiments was designed to determine which details are included in the speech plan. The SAS paradigm elicits the prepared responses both rapidly and accurately. Therefore, these SAS-induced responses may reveal details that are pre-specified in the speech plan and are executed independently from feedback regulation.

Each chapter of this dissertation tackles a different aspect of the planning of spoken syllables. The dissertation content is structured as follows.

Chapter 2 begins by investigating whether the StartReact effect is observed in prepared speech elicited by a SAS. While a number of studies on limb movements suggest that the StartReact effect may be more associated with triggering from subcortical areas, it is not clear whether prepared movements that involve heavy cortical processing, such as speech production, are also subject to the same effect. An experiment was conducted that elicited a simple RT task under SAS. The upshot of this test is a straightforward confirmation that the startle paradigm is applicable to prepared speech. Specifically, the results show that the StartReact effect is observed for prepared syllable production, and that speech planning may govern a domain of syllables at least up to CV length.

Chapter 3 further tests whether multi-dimensional aspects of speech tasks are preserved in SAS-induced responses. In particular, this chapter tackles the question of whether SAS-induced effects can be speech-specific
or whether they are generic to any voluntary lip movements. To test this, Mouthed speech and Non-speech oral movements were employed in a SAS experiment. SAS-elicited responses reveal that lip compression did not occur in the Mouthed speech responses as frequently as in Spoken speech responses, and that lip compression in terms of lower lip vertical displacement from the Mouthed speech responses was not as large as those in Spoken speech responses. The results suggest that lip compression, potentially driven by aerodynamic factors, may be specific to speech, and that multi-dimensional information that is pre-specified in speech tasks ought to be preserved in response to SAS.

Chapter 4 investigates whether, in a simple RT task, lip compression that is independent of aerodynamics of the responses is observed in all speech-related tasks and ought to be elicited at a short latency by SAS. In particular, the experiment examines lip compression of the prepared sequence with a preceding mouth-closing movement and compares that with those without preparatory movements. While reaction time was not affected by additional preparatory movement, lip compression appears to be comparable in Spoken and Mouthed responses. The differences with regards to lip compression in the mouth-closed condition are not seen in this mouth-open condition.

Chapter 5 compares SAS responses in English and Taiwanese Mandarin, showing that acoustic details that create phonemic contrasts, such as formants and pitch contour profiles, may be pre-specified in the speech plan and thus be resistant to SAS-induced effects.

Finally, Chapter 6 discusses the implications of the findings and concludes the dissertation.
Chapter 2

The StartReact effect in syllable production

2.1 Introduction

In order to use the startle paradigm to examine different aspects of speech plans, it is first necessary to confirm that this experimental design is applicable to prepared speech production. In this chapter, an experiment using SAS was conducted to study prepared CV syllables. Positive results support the conclusion that the startle paradigm effectively elicits prepared speech, suggesting likely cortical involvement for startle effects, and thus providing a platform for further investigation.

2.1.1 The StartReact effect and cortically determined processes

Recently, research investigating the preparation and initiation of goal-directed actions (e.g., arm extension movements) has shown that preparation can be completed in advance, possibly stored in subcortical areas, and then triggered by a startling acoustic stimulus (i.e., SAS; > 120 dB)
When a participant is engaged in a reaction time task and is waiting for an imperative stimulus to produce a prepared movement, a startling auditory stimulus can trigger the specific pre-programmed response in significantly reduced reaction time without compromising response accuracy. The resulting response latency is so short (< 70 ms) that it appears unlikely the process of initiation could involve cortical or cognitive mechanisms. This early release of the intended movement, termed the “StartReact effect” (Valls-Solé et al., 1999), is accompanied by largely unaltered kinematic phasing and muscle activity patterns.

The StartReact effect has been reported in head-turning movements (Oude Nijhuis et al., 2007; Siegmund et al., 2001), eye movements (Castel-lote et al., 2007), and stepping (MacKinnon et al., 2007). Nevertheless, it is not clear whether speech movements, which presumably hinge on heavy cortical processing, can also be elicited in a similar fashion. In this thesis, I will explore whether a SAS can induce a StartReact effect in speech production. A simple reaction time (RT) task will be employed to examine the preplanning of various details of speech production.

Early in the literature on the StartReact effect, it was proposed that intended responses are programmed and stored in subcortical areas (e.g., Carlsen et al., 2004b; Valls-Solé et al., 1999), with the release of the prepared program bypassing normal cortical pathways for voluntary movements. However, while these studies emphasize the storage and triggering in subcortical areas, other studies propose that the subcortical storage and triggering hypothesis may not provide the only account for the StartRe-
act mechanism. For example, Siegmund et al. (2001) attributed the early release to enhanced perceptual processing and intersensory facilitation in cortical circuits. Evidence from transcranial magnetic stimulation (TMS) also provides support for cortical involvement in the initiation of prepared responses (e.g., Alibiglou and MacKinnon, 2012; Stevenson, 2011). More recent studies propose that the startle reflex induces increased activation in the thalamus, allowing the prepared response to be released as early as an involuntary response (e.g., Carlsen et al., 2012; Maslovat et al., 2011). Following Carlsen et al. (2012), it is hypothesized that “a SAS acts as a subcortically mediated trigger for a cortically stored motor program.” (Carlsen et al., 2012, p. 30).

Considering speech production, it is widely acknowledged that a wide range of cortical and subcortical areas are involved in the preparation and initiation of speech. In particular, the initiation of prepared speech hinges on cortically determined processes and cannot be released simply by reticulo-spinal commands (Bohland and Guenther, 2006). In their study, participants were instructed to produce a pre-cued syllable sequence when a GO signal stimulus was presented. No syllable production was elicited when the stimulus was a NO GO signal. Their results of GO trials showed more significant responses in the primary and somatosensory cortices, the superior temporal plane, the anterior insula, and the medial premotor areas, with particular focuses in the SMA near the superior convexity and portions of pre-SMA and anterior cingulate sulcus. A main effect of the overt speech (i.e., GO trials) suggests that these cortical areas are associated more with the initiation than with the planning or sequence buffering of the prepared
speech. Applying the startle paradigm to prepared speech not only allows us to examine the potential impact of the StartReact effect on speech production, but more generally can lend support to the hypothesis that cortically determined motor programs are subject to a rapid trigger by a SAS.

2.1.2 Syllables as speech units

For the purpose of the current study, CV syllables are chosen as the target syllable for a number of reasons. CV syllables are cross-linguistically common in terms of word inventory (Bell and Hooper, 1978). Earlier studies talked about the planning for different levels of “phonological/phonetic representations,” such as syllables (Levelt et al., 1999), feature (Bernhardt and Stemberger, 1988), and subphonemic representations (Derrick, 2011). In particular, syllables have attracted most attention and the role of mental syllabary has been emphasized through different experiment settings, such as naming task (Ferrand and Segui, 1998; Levelt and Wheeldon, 1994), priming (Cholin and Levelt, 2009), and fMRI (Brendel et al., 2011). Being fundamental units in speech production, mental syllabary is interpreted as output of the projections from premotor cortex, to primary motor cortex, and supplemented by cerebellum (Guenther, 2006).

Syllables being fundamental units of speech production is not only grounded by experimental results, but also supported by human biomechanical motor development. In language acquisition, it has been suggested that CV syllable (e.g., /ba/) may be a development emerged from mastication behaviors (Davis et al., 2002; MacNeilage, 1998; MacNeilage et al., 2000). In pre-speech babbling, CV syllables are commonly observed as it has a lot of in
common with chewing behaviors. Take /ba/ for example. The closure of the consonant and the opening of the vowel form a syllabic frame. The closure is completed between lips whereas no positional specification is required for the tongue. Thus, when the lips were burst open, the null specification of the tongue would naturally lead to a central vowel /a/, as opposed to other front vowels (e.g., /i/) or back vowels (e.g., /o/) (cf. MacNeilage et al., 2000). It is why /ba/ is commonly seen in early speech babbling.

If a syllable is a result of the composition of these motor command projections, the articulatory content should include details from these motor commands.

2.1.3 Proposal

The aim of this study is to investigate whether the StartReact effect is also observed in movements that involve heavy cortically determined commands, such as prepared speech. There exists detailed evidence to suggest that the preparation of intended speech involves a wide range of cortical and subcortical networks and the initiation and execution of prepared speech depends on exerting cortical projections to subcortical areas via a descending cortico-subcortical pathway (Iwata et al., 1996; Jürgens, 2002; Simonyan and Horwitz, 2011). I believe the use of speech provides a novel alternative to cortically dependent finger movements (i.e., Carlsen et al., 2009a; Honeycutt et al., 2013), as speech requires a high degree of cortical and cognitive involvement (Grimme et al., 2011) and has been shown to be controlled in a different manner than non-speech facial movements (Tremblay et al., 2003). Given the proximity between the mouth and the neo-cortex, it is assumed
that the threshold for the StartReact effect may not be as high as that in finger lifting tasks. Here we hypothesize that the initiation of prepared speech relying on cortical commands is also susceptible to the rapid release triggered by a SAS. I test this hypothesis by employing the SAS design. As revealed in previous studies, when paired with an imperative stimulus, a SAS can trigger an early release of the prepared response. If there is cortical involvement in the StartReact effect, we would predict that the syllable /ba/ would be subject to the StartReact effect and triggered early by the SAS.

As proposed by Carlsen et al. (2012), the StartReact circuit involves an ascending reticulo-thalamo-cortical pathway. The SAS invokes a rapid involuntary release of a prepared response. Given this account of the mechanism that underlies the StartReact effect I hypothesize that the initiation of prepared speech is also susceptible to rapid release which is triggered by a SAS.

2.2 Methods

2.2.1 Participants

Data were collected and analyzed from nine participants (3 male and 6 female; Mean = 23 years, SD = 4.2 years) who showed a consistent startle reflex in the sternocleidomastoid (SCM) muscle on baseline startle trials (a pretest startle trial prior to the testing session, no speech required) and more than 50% of the startle testing trials. Participants were all native speakers of North American English. Prior to the testing, participants signed
an informed consent form and were naïve to the hypothesis under investigation. The experiment was conducted by following the ethical guidelines established by the University of British Columbia.

2.2.2 Apparatus, task, and procedures

Participants sat in an upright chair facing a computer monitor (Acer, X223W, 22”, 60 Hz refresh rate) at a distance of approximately 1.5 meters and were instructed to look straight ahead at the monitor and respond to an acoustic stimulus by vocalizing the target syllable /ba/ as quickly as possible. A visual display of the syllable /ba/ was presented on the monitor concurrently with the acoustic stimulus. Throughout the testing session, participants were asked to start with their mouths closed in a relaxed posture without compressing their lips during the preparation. Prior to this testing session a baseline startle trial was introduced. Participants were seated and waiting for testing to begin when the unexpected baseline SAS (124 ± 2 dB, 40 ms, 1,000 Hz, < 1 ms rise time) was delivered. Figure 2.1 depicts the positioning of participants and experimental apparatus.

![Figure 2.1: Relative positions of the monitor, participant, and loud speaker.](image)

All testing trials began with a warning tone (100 ms, 1000 Hz, 80 dB)
played directly from the computer’s sound card. The acoustic imperative stimulus and visual /ba/ followed the warning tone by a random foreperiod of between 1500 and 2500 ms. This auditory signal was either a control stimulus (80 ± 2 dB, 100ms, 1,000 Hz) or startling stimulus (124 ± 2 dB, 40 ms, 1,000 Hz, < 1 ms rise time), generated by a customized computer program. The acoustic stimuli were amplified (HiFi stereo power audio amplifier A180W) and then presented via a loudspeaker placed directly behind the head of the participant. The acoustic stimulus intensities were measured using a sound level meter (Cirrus Research model CR:252B, “A” weighted scale, impulse response mode) at a distance of 30 cm from the loudspeaker (approximately the distance to the ears of the participant).

Participants performed a single testing session of approximately 20 min. The testing block consisted of twenty control trials and five startle trials. The five startle trials were presented pseudo-randomly such that the first trial was never a startle trial, nor were there two consecutive startle trials. Prior to this testing, participants also performed lip movement and mouthing articulation tasks in order to familiarize the testing procedure. Only results from the vocalized syllables were included in the analyses.

2.2.3 Recording equipment

Participants performed the tasks with three infrared light-emitting diodes placed on the center of the upper lip, the lower lip, and the bridge of the nose. 3D positions of these diodes were monitored using an OPTOTRAK (Northern Digital Inc., Waterloo, Ontario) motion analysis system (spatial resolution 0.01 mm). The data collected from the bridge of the nose were
considered as a reference marker for the other two landmarks. The OPTOTRAK camera unit was placed above the computer monitor that was used to display the syllable /ba/. The 3D positions of the upper and lower lips were sampled at 500 Hz. Raw data from the OPTOTRAK were converted into 3D coordinates and digitally filtered using a second-order dual-pass Butterworth filter with a low-pass cutoff frequency of 10 Hz.

Figure 2.2: Sternocleidomastoid (SCM) and orbicularis oris muscles for the attachment of surface EMG electrodes. Note the SCM and orbicularis oris figures were adapted and modified from Gray’s Anatomy (20th ed.) Figures 385 (http://www.bartleby.com/107/illus385.html) and 381 (http://www.bartleby.com/107/illus381.html), respectively.

Bipolar surface electromyography (EMG) electrodes (Therapeutics Unlimited Inc., Iowa City, IA) were attached to four different locations: the skin above the upper vermilion border (labeled as “upper lip”), the skin below the lower vermilion border (labeled as “lower lip”), and the left and right SCM muscles. The EMG electrodes were placed centered between the midline and the right corner of the mouth, parallel to the line of force of the
muscles. A ground electrode was placed on the right ulnar styloid process. A wired lapel microphone was pinned onto the collar of the participant in order to record the participant’s responses. Acoustic data were collected by the wired lapel microphone through a Preamp (USBPre Microphone Interface for Computer Audio, Sound Devices, LLC) before analyses. Figure 2.2 highlights the SCM and lip muscles to which the surface EMG electrodes are attached; Figure 2.3 illustrates the participants’ set-up for the current experiment. A customized LabView® computer program controlled the stimulus presentation and the collection of EMG and acoustic data at a rate of 4 kHz (National Instruments, PC-MIO-16E-1). Data collection began 500 ms
before the presentation of the stimulus and was terminated 2500 ms later.

2.2.4 Interpretation of EMG and lip displacement

In the current design, the surface EMG electrodes were placed on the skin above and below the vermilion border. The electrodes were positioned to capture activity from the orbicularis oris superior (OOS) and inferior (OOI), respectively. While it is acknowledged that the surface EMG may also pick up some activity from other surrounding muscles, such as levator labii superioris, depressor labii inferiroris, or depressor anguli oris. It is assumed in this study that the upper lip electrode indicates by and large EMG activity from OOS and the lower lip electrode shows primarily EMG activity from OOI.

Averaged rectified raw EMG traces of all baseline startle trials are displayed in Figure 2.4. EMG responses were observed in left and right SCM as well as the upper and lower lips at approximately the same time (upper lip Mean = 59 ms, SD = 9 ms; lower lip Mean = 54 ms, SD = 13 ms, left SCM Mean = 60 ms, SD = 16 ms; right SCM Mean = 62 ms, SD = 17 ms), all were considered as startle indicators in the following analyses. While it appears that all four muscles can be considered startle indicators, we chose to use SCM as our primary indicator for consistency with previous work (Carlsen et al., 2011). However, the involvement of the prime movers in the reflexive startle response did not allow for determination of voluntary EMG onset for vocalization during startle trials and thus our dependent measures included kinematic markers and acoustic burst onset (see below).

Typical profiles of EMG and lip displacement for control trials are dis-
Figure 2.4: Average rectified raw EMG traces of baseline trials, including (from top) right SCM, left SCM, upper lip, and lower lip. EMG activity was plotted with respect to the imperative stimulus (the vertical grey line).

played in Figure 2.5. In response to an imperative stimulus, the upper and lower lips first compress against each other (note the upward displacement of the lower lip from point A to B, Figure 2.5). This lip compression was evident in 91% of control trials and 87% of startle trials. Such compression between the lips is anticipated in order to contain the intraoral air pressure associated with the bilabial consonant [b]. As such, only the trials with lip compression were analyzed. Following lip compression, the opening phase of the vowel [a] was realized primarily through the downward movement of the lower lip, with only a minimal contribution from the upward movement of the upper lip (point B through C, as seen in Figure 2.5). SCM muscles also showed activity during this opening phase, presumably associated with
Figure 2.5: EMG activity, lip displacement and response acoustic waveforms of control exemplar. The EMG measure includes (from top) right SCM, left SCM, upper lip, and lower lip. EMG activity was rectified and plotted with respect to the imperative stimulus (the vertical grey line). Lip displacement and response acoustic waveforms are also plotted with respect to the imperative stimulus. Point A marks the beginning of voluntary movement; point B marks the time to the lower lip opening onset; point C marks the lowest position of the lower lip. The intervals across kinematic markers were labeled from $t_1$ to $t_5$. See text for details.
jaw lowering (see Uemura et al., 2008, for similar findings).

2.2.5 Data reduction, dependent measures, and statistical analyses

A total of 36 of the 225 trials (16%) were excluded from the analyses. Reasons for discarding trials included delayed lip movement onsets or voice onsets (defined as 2 SD away from the mean, 25 trials), startle trials in which no detectable SCM startle response or delayed response (>120 ms) was observed (10 trials), and trials with poor EMG data in which no obvious onsets could be identified (1 trial).

In each response, the initiation of lip compression was marked as the voluntary movement onset (point A on lower lip trace in Figure 2.5) which allowed us to measure the RT of the voluntary movement onset (time between the onset of IS to the beginning of lip compression). Differences between RTs of voluntary movements in both control and startle responses were measured and analyzed using a paired Student’s t-test. After the lower lip reached its highest point (point B in Figure 2.5), the opening phase of the movement began, here termed the lower lip opening onset. The lower lip continued to move downward during the production of /ba/, until it reached its lowest position (point C in Figure 2.5). Thus, two kinematic events can be identified: lip compression and lip opening. The acoustic burst release time was defined as the latency from the presentation of the IS to the onset of the acoustic burst of the bilabial consonant [b] (see acoustic signal, Figure 2.5). To measure the time of the initial acoustic burst release, the acoustic signals were first de-emphasized by applying a filter at a level of 10Hz.
using the PRAAT software (Boersma and Weenink, 2009). A narrow-band spectrogram (bandwidth of 43 Hz; wave-length of 30 ms) was displayed in PRAAT to mark the acoustic burst release. Syllable durations were also measured, marking the interval from the acoustic burst release to the end of periodicity of the vowel [a].

In order to examine whether the kinematic events in control responses are different from those in startle responses, we calculated the relative timing across kinematic markers A, B, and C by measuring the time between these markers, as well as the relative timing between the kinematic markers and acoustic burst onset (see Figure 2.5). The time between voluntary movement onset and lower lip opening onset is shown as $t_1$ (duration of lip compression) whereas $t_2$ marks the duration of the opening movement from the lower lip opening onset to the lower lip’s lowest position. The time frame for both of these events is $t_3$ (time between point A and C in Figure 2.5), and the relative timing of each kinematic event with respect to the entire time course was calculated. To examine the relative timing between the kinematic events and acoustic signal, we calculated the time from movement onset to acoustic burst release ($t_4$) and lower lip opening to acoustic burst release ($t_5$). The ratios of $t_1/t_3$, $t_2/t_3$, $t_4/t_3$, $t_5/t_3$ for each trial (both startle and control) were calculated with the ratio data transformed by trial using the arcsine square root transformation (this was used due to correlations between markers; McDonald, 2009). The transformed ratios were then aggregated by participant and analyzed via a paired Student’s $t$-test.

The peak-to-peak displacement of the lower lip was calculated as the vertical distance from the highest position (Figure 2.5, point B) to the lowest
position (Figure 2.5, point C). The difference in peak-to-peak displacement between control and startle trials was analyzed using a paired Student’s $t$-test. Acoustic burst release times and syllable durations were both similarly analyzed using paired Student’s $t$-tests.

In order to confirm that the target syllable /ba/ was successfully produced, the acoustics of the produced syllables in control and startle trials were analyzed. Acoustic formants F1 and F2 (the two lowest resonant frequencies of the vocal tract) are standard indicators of overall vocal tract shape for speech sounds, encoding information about articulator positions during vowels (Fant, 1960; Peterson and Barney, 1952) as well as about consonants and CV transitions (Delattre et al., 1955). Frequency values for F1 and F2 were extracted throughout the entire duration of produced syllables using LPC formant tracker in PRAAT and normalized to their respective $z$-scores within each participant. Data were analyzed across normalized durations. For statistical comparison, formant data were submitted to a smoothing spline analysis of variance (SS ANOVA; see Davidson, 2006; Derrick and Schultz, 2013).

2.3 Results

2.3.1 Startle indicators

Average EMG traces of every trial in each condition for the four muscles of interest (left SCM, right SCM, upper lip, and lower lip) are displayed in Figures 2.6. When EMG traces are temporally aligned to the stimulus onset (Figure 2.6), all of the muscles that are involved in the voluntary action of
saying /ba/ also show startle responses. These startle indicators are quite invariant and display typical reflexive patterns (as seen in Carlsen et al., 2009b; Maslovat et al., 2009, 2011), followed by more diffuse EMG activity associated with the subsequent voluntary action. Figure 2.6 also shows that the onset latencies of the startle EMG responses in the /ba/ trials were comparable with those in the baseline trials (Figure 2.4).

As seen in Figure 2.6, the more diffuse EMG activity for the voluntary movement was due to higher variability with regard to the onset of the voluntary action. Normalizing data to the lower lip opening onset allows us to better examine the muscle activity that is associated with the voluntary action and to clearly compare the difference between control and startle trials. When the EMG traces are normalized to the onset of the lower lip opening movement (i.e., temporally aligned to the lower lip opening onset, as marked by the vertical grey line in Figure 2.7), the EMG trace of voluntary action is now distinct (variability of time between onset of opening movement and EMG is low). As seen in Figure 2.7, in control trials both upper and lower lips were engaged in the lip compression prior to the opening, whereas only the lower lip showed EMG activity during the voluntary opening movement. Similar patterns were also observed in startle trials. It is noted that stronger EMG activity in response to the SAS was observed in the lip muscles as well as in the SCM muscles. The SCM muscles showed both startle responses (activity before zero, marked by the grey line in Figure 2.7) and EMG activity for the intended movement (activity after zero, marked by the vertical grey line in Figure 2.7).
Figure 2.6: Mean EMG activity and kinematic displacement of control (grey) and startle (black) trials in the vocalization condition with respect to the imperative stimulus. The top four channels represent the mean EMG across four muscles (from top to bottom): left SCM, right SCM, upper lip, and lower lip. The bottom two channels depict the mean upper lip and lower lip displacement trajectories, respectively.
Figure 2.7: Mean EMG activity and kinematic displacement of control (grey) and startle (black) trials in the *vocalization* condition with respect to the lower lip opening onset. The top four channels represent the mean EMG across four muscles (from top to bottom): left SCM, right SCM, upper lip, and lower lip. The bottom two channels depict the mean upper lip and lower lip displacement trajectories, respectively.
2.3.2 The StartReact effect

A summary of the results for all measures, including means, standard deviations, and p-values, are provided in Table 2.1.

<table>
<thead>
<tr>
<th>Experimental trials</th>
<th>Control</th>
<th>Startle</th>
<th>p-value</th>
<th>Significant</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RT of voluntary movement onset (ms)</strong></td>
<td>116.33 (32)</td>
<td>75.46 (30)</td>
<td>0.0036</td>
<td>*</td>
</tr>
<tr>
<td><strong>RT of acoustic burst release time (ms)</strong></td>
<td>268.33 (73)</td>
<td>203.64 (43)</td>
<td>0.018</td>
<td>*</td>
</tr>
<tr>
<td><strong>Lower lip peak-to-peak displacement (mm)</strong></td>
<td>18.5 (5.4)</td>
<td>21.17 (6.2)</td>
<td>0.007</td>
<td>*</td>
</tr>
<tr>
<td>( t_1/t_3 ) (transformed arcsine square root)</td>
<td>0.58 (0.08)</td>
<td>0.57 (0.06)</td>
<td>0.68</td>
<td>–</td>
</tr>
<tr>
<td>( t_2/t_3 ) (transformed arcsine square root)</td>
<td>0.99 (0.08)</td>
<td>1.00 (0.06)</td>
<td>0.68</td>
<td>–</td>
</tr>
<tr>
<td>( t_4/t_3 ) (transformed arcsine square root)</td>
<td>0.86 (0.08)</td>
<td>0.85 (0.06)</td>
<td>0.37</td>
<td>–</td>
</tr>
<tr>
<td>( t_5/t_3 ) (transformed arcsine square root)</td>
<td>0.55 (0.06)</td>
<td>0.55 (0.06)</td>
<td>0.89</td>
<td>–</td>
</tr>
<tr>
<td><strong>Syllable duration (ms)</strong></td>
<td>176.29 (62)</td>
<td>168.42 (51)</td>
<td>0.59</td>
<td>–</td>
</tr>
</tbody>
</table>

It was predicted that a SAS would induce faster onset of movements of a prepared syllable. Our results support this prediction. Shorter reaction times of voluntary movement onsets (time from stimulus presentation to
point A in Figure 2.5) were observed for startle trials (Mean = 75 ms, SD = 30 ms) than for control trials (Mean = 116 ms, SD = 32 ms; t(8) = 4.07, p < 0.01). The latency of the acoustic burst release (from stimulus presentation to point C in Figure 2.5) was also accelerated in startle trials (Mean = 204 ms, SD = 43 ms) compared to control trials (Mean = 268 ms, SD = 73 ms; t(8) = 2.97, p = 0.018). In addition to shorter latencies of the kinematic markers, the SAS triggered increased muscle activity (see Figure 2.6 and 2.7), yielding a significantly increased range of lower lip peak-to-peak displacement (Mean = 21.2 mm) than control trials (Mean = 18.5 mm), (t(8) = -3.83, p < 0.01).

However, the relative timing relationship of kinematic and acoustic events remained unaffected. The time between voluntary movement onset and lower lip opening onset (t1 in Figure 2.5) for control and startle responses were 30% and 29%, respectively, of latencies from voluntary movement onset to the time when the lower lip reached its lowest point (t3 in Figure 2.5). Secondly, latencies between lower lip opening onset and the time of the lower lip’s lowest point (t2 in Figure 2.5) for control and startle trials were 70% and 71%, respectively, of latencies from voluntary movement onset to the time of the lower lip’s lowest point (t3 in Figure 2.5). Thirdly, the latencies from the voluntary onset to the acoustic burst release (t4 in Figure 2.5) for control and startle trials were 58% and 56%, respectively, of t3. Lastly, the latencies between lower lip opening onset and acoustic burst release for control and startle trials were 27% and 26 %, respectively of t3. All these ratios were transformed using arcsine square root transformation. The results are summarized in Table 2.1. No statistical difference was found for
the time course of these kinematic events between control and startle responses, suggesting that the relative timing of the kinematic markers was not compressed during the startle trials.

The SS ANOVA results of formants are displayed in Figure 2.8, including the fit predicted by the SS ANOVA model and shading of 95% confidence intervals. For F1 and F2, frequency profiles between control and startle responses were within 95% confidence interval (as indicated by overlapping confidence interval bands in Figure 2.8). No difference in syllable duration was observed between control and startle trials ($t(8) = -0.56, p = 0.59$).

In addition to shorter latencies of the kinematic markers, the SAS triggered increased muscle activity, as indicated by EMG, yielding a significantly increased range of lower lip peak-to-peak displacement than control trials ($t(8) = -3.83, p < 0.01$).
Figure 2.8: SS ANOVA comparison of formant frequencies F1 and F2 over syllable /ba/ durations. The black line denotes the predicted fit for control responses and the white line denotes the startle responses. The dark grey and light grey bands surrounding the predicted fits represent a 95% confidence interval. Any white space between the transition lines for each condition represents a statistically significant difference between the given measures.
2.4 Discussion

Apart from other limb movements in response to auditory stimulus, the current experiment considers syllable production in speech as another movement task known to require cortical involvement. In syllable production, primary motor cortex and Broca’s area (i.e., left inferior frontal gyrus) are associated with linguistic processing including syllabification and phonetic encoding (Hickok and Poeppel, 2000, 2004, 2007; Indefrey and Levelt, 2004; Papoutsi et al., 2009; Tourville and Guenther, 2011). We predicted that prepared syllables would be subject to rapid release by a SAS. The current experiment found that both the lower lip voluntary movement and the acoustic burst were initiated earlier in startle trials while the acoustics of produced syllables remained intact. These results support our hypothesis that speech is subject to the StartReact effect despite its heavy cortical involvement.

It is conceivable that only the early part of the response (e.g., points A and B in Figure 2.5) would be included in the preplan such that a SAS would only trigger these early parts without accelerating the rest of the response (i.e., the mouth opening from point B to point C). However, we found that the response triggered by the imperative stimulus was executed as a coherent whole, as evidenced by a fixed timing relationship across all the kinematic markers. Likewise, the acoustic formant profiles of SAS-induced syllables (indicating the shape and position of the speech articulators), matched those of the voluntary responses (Figure 2.8). Thus, when a preplanned syllable is triggered by a SAS, both temporal and acoustic properties of the syllable
remain intact, indicating that the SAS effects rapid release of the entire prepared syllable.

It is noteworthy that the StartReact pathways largely overlap with the pathways for speech production. As summarized in Maslovat et al. (2011) and Carlsen et al. (2012), we propose that the StartReact response is mediated via an ascending thalamo-cortical pathway, generated by activation from reticular formation exerting on thalamus. Increased activation in thalamus provides inputs to primary motor cortex to initiate the cortically prepared movement via a descending corticospinal pathway. Similarly, speech production also relies on thalamo-cortical circuits and a descending corticospinal pathway. Receiving inputs from cerebellum, thalamus projects to primary motor cortex and Broca’s area. The commands are mediated via putamen and reticular formation and sent down to the phonatory motoneurones in the spine (Guenther et al., 2006; Iwata et al., 1996; Jürgens, 2002). The triggering by a SAS involves the initiation of the cortically prepared and stored syllable movements via the same StartReact pathways for upper limb movements. Shorter reaction times in StartReact responses are accounted for by increased neuron activation reaching faster above initiation threshold (see Carlsen et al., 2012, for details).

In SAS-induced responses, the voluntary movements for the prepared syllable were accelerated. In addition to the StartReact effect, we also found startle reflex activity in the perioral region. As both the startle reflex and the accelerated StartReact responses occur in the perioral region, it presents a potential confound to identify whether there are two events (i.e., a startle reflex followed by a voluntary movement) or there is only one responsive
activity in the perioral muscles. It is no surprise to observe a startle reflex in lip muscles as similar reflex activity in lips and orofacial muscles have been reported in other startle studies (Brown et al., 1991; Valls-Solé et al., 2008) and lip tapping studies (Bratzlavsky, 1979; McClean, 1991). As observed in Figure 2.6, the earliest EMG activity in SAS-induced responses is associated with the startle reflex, ranging from 40 ms to 90 ms. The startle reflex activity was present in all muscles measured, but more robust in the two SCM and lower lip muscles. Note that this reflex activity is distinct from the subsequent activity for the lower lip, which corresponds with the prepared but accelerated speech movement. Using TMS and SAS, Stevenson et al. (2014) showed that the startle reflex is dissociated from the StartReact responses. The startle reflex is modulated via a descending pathway from pontine reticular formation to nRPC (Carlsen et al., 2012; Yeomans and Frankland, 1995). When a SAS is presented, activations are projected to the cranial nerve VII (facial nerve), by which perioral muscles, including the orbicularis oris, levator and depressor labii muscles, are innervated. This innervation accounts for the observed startle reflex in the perioral muscles whereas accelerated voluntary responses are mediated by the dissociated pathways for StartReact responses. Therefore, SAS-induced EMG activity bears on two separate events associated with distinct neural pathways.

In SAS-elicited responses, the SCM muscles have been considered as a reliable startle indicator. The current experiment measured not only the SCM but also the perioral muscles in response to SAS. The results revealed that in addition to the SCM, the lower lip muscle also consistently exhibited
startle reflex activity prior to the activity for the voluntary movement (Figure 2.6). When data are normalized to the voluntary opening onset (Figure 2.7), two EMG events were observed in the lip muscles. The first one is associated with lip compression whereas the second one is directly related to the opening movement. While both lips are involved in lip compression, only the lower lip exhibited activity for the opening movement. Whether these observations are speech-specific or generic to any oral movements will require more experiments and empirical support.

In summary, this experiment is the first study applying the startle paradigm to prepared speech. In response to a SAS, the prepared syllable can be rapidly released with the temporal and acoustic properties of the response syllable preserved intact. In terms of associated neural pathways, the results suggest that the early startle reflex observed in the lips is mediated via reflex pathways dissociated from the StartReact pathways, and that upper limb movements and syllable production share similar neural pathways for StartReact responses. The rapid and accurate release of the prepared syllable by a SAS confirms that this startle paradigm can serve as a suitable design to further investigate other details of speech plans. A more significant implication of the startle paradigm design is that the StartReact effect is observed in tasks that are dependent on heavy cortical processing, such as syllable production.
Chapter 3

Startling spoken, mouthed, and non-speech movements

3.1 Introduction

Chapter 2 showed that a speech plan as long as a CV sequence (i.e., /ba/) can be prepared in advance and subject to rapid execution when elicited by a SAS. While only limited feedback may be available at such a short latency, the prepared syllable is performed as intended without any interruption to the synergistic system. While Chapter 2 has demonstrated that the StartReact effect is observed in prepared syllable production, it remains unclear whether these SAS-induced effects are speech-specific or generic to any lip movements. Minimally, if detailed information is specified for different speech tasks, these details ought to be preserved in SAS-induced responses. As such, we can apply the startle paradigm to different speech tasks in order to examine the pre-specified details and possible induced effects.

A number of studies have suggested that speech plans may encode multi-dimensional details, such as aerodynamics (e.g., Cho et al., 2002; Gick
et al., 2012; Murphy et al., 1997), muscular structure and coordination (e.g., Gracco and Löfqvist, 1994; Löfqvist and Gracco, 1997), and sensitivity to somatosensory feedback (e.g., Larson et al., 2008; Tremblay et al., 2003). It should be noted that different speech tasks may construct speech plans with different specifications across these dimensions. For example, Murphy et al. (1997) compared four different speech-related tasks: vocalized speech, mouthed speech (i.e., speech with articulatory movements, but no audible output), unarticulated speech (i.e., speech with vocalization, but no corresponding articulatory movements), and internal speech (i.e., silent speech without any articulation or vocalization). Their results show that the breathing patterns for vocalized speech and unarticulated speech are more similar, whereas the breathing pattern for mouthed speech resembles the pattern for internal speech. In addition to breathing patterns, spoken speech and mouthed speech also differ in EMG activity and correspondent kinematics. Compared with spoken speech, mouthed speech has shorter word duration, reduced EMG amplitude, and overall hypoarticulation in lip movements (Crevier-Buchman et al., 2011; Janke et al., 2010; Wand et al., 2009). While different multi-dimensional information is included in the pre-plans for these speech tasks, both spoken and mouthed speech tasks serve similar functions (i.e., both are speech-related) and both have a similar somatosensory basis during speech production. Tremblay et al. (2003) examine somatosensory feedback during speech production by applying mechanical perturbation to the jaw. In their study, the participant’s jaw was perturbed during the production of vocalized speech, mouthed speech, and non-speech movement. After training, adaptation to compensate for the perturbation
was observed in vocalized and mouthed speech. For both speech-related tasks, an after-effect was reported when perturbation was removed from trials. However, the training effect and after-effect were not found in the non-speech movement task. Their results suggest that somatosensory basis for speech-related tasks may be dissociated from the basis for generic non-speech movements.

In Chapter 2, participants were instructed to produce the sequence /ba/ from a mouth-closed condition. Intuitively, the initial component of the syllable production should be the release of the bilabial stop. However, as revealed by the results, the voluntary onset action was the two lips compressing against each other – in particular, the lower lip pushing upward against the upper lip. This lip compression occurred in 91% of spoken /ba/ responses. It is important to note here that this compression is not a side effect of StartReact itself; the same compression was also observed in 87% of control responses, suggesting that the observed lip compression is most likely a preparatory movement, part of the motor plan in preparation for the upcoming speech movements. We have proposed that the observed lip compression is associated with intraoral pressure required for the production of /b/; Chiu and Gick (2013) use a computational model to simulate the production of bilabial stops, reporting results supporting the view that anticipatory lip compression is required when intraoral pressure is implemented during production. Based on this previous work, we can predict that reduced lip compression should be observed in speech tasks where no or limited air pressure is required, as in mouthed instances of the syllable /ba/. Alternatively, there remains the possibility that this lip compression
could be simply a precursor to any opening movement (including a non-speech movement), in which case such compression should be observed to occur as frequently in other speech and non-speech tasks. A more probable prediction is that, since no intraoral pressure is required, much reduced lip compression would take place in a non-speech lip opening movement, compared with spoken and mouthed speech tasks.

In a simple RT task, all the details of the response have to be prepared in advance of stimulus presentation. The complexity of a prepared speech response does not affect reaction time, since the internal structure of the response (including duration) is specified before the presentation of the stimulus (cf. Klapp, 2003). Therefore, in terms of response latency, a reasonable prediction based on this would be that there should be no difference when triggered by an auditory stimulus regardless of the target response; that is, when elicited by an auditory stimulus (i.e., a control “go” signal), the latency of voluntary movement for spoken speech should be of no difference from that for mouthed speech or non-speech responses. Similarly, when a SAS is presented as an imperative GO signal, prepared mouthed and spoken speech should be released at a comparable latency.

This chapter examines how a SAS can elicit prepared mouthed-speech and non-speech movements. Two predictions can be made. First, reduced and less frequent lip compression would be expected for the mouthed speech and non-speech oral movement. Second, a SAS may induce accelerated release of prepared mouthed speech and non-speech movements, whereas no differences with regards to reaction time between conditions is expected. That is, the latency of the voluntary movement onset from control responses
for mouthed speech and non-speech movement should be comparable to that for Spoken speech; the latency of the voluntary movement onset from startle responses, though shortened, for all three conditions should have comparable latencies.

### 3.2 Methods

The experiment presented here was conducted to test the hypothesis that the StartReact effect and associated lip compression observed in spoken syllables can be observed in speech-like (i.e., Mouthed speech) and Non-speech (i.e., mouth opening movements) motor behaviours.

#### 3.2.1 Participants

Data were collected and analyzed from the same nine participants reported in Chapter 2 (3 male and 6 female; Mean = 23 years, SD = 4.2 years). Participants were all native speakers of North American English. Prior to the testing, participants signed an informed consent form and were naïve to the hypothesis under investigation. The experiment was conducted following the ethical guidelines established by the University of British Columbia.

#### 3.2.2 Apparatus, task, and procedures

The same apparatus from Chapter 2 were applied. Two tasks were designed in blocks. Throughout the testing session, participants were asked to start with their mouths closed in a relaxed posture without compressing their lips during the preparation. The first block (hereafter the Non-speech con-
dition) required participants to respond to an acoustic stimulus by opening their mouths. In the second block, participants were instructed to respond to the acoustic stimulus by mouthing a silent articulation of /ba/ (hereafter the Mouthed condition), in which no pulmonic air flow (and hence no phonation) was produced.

The testing procedures are identical to those reported in Chapter 2. Participants performed a single testing session of approximately 20 minutes. The testing block consisted of twenty control trials and five startle trials. The five startle trials were presented pseudo-randomly such that the first trial was never a startle trial, nor were there two consecutive startle trials.

3.2.3 Recording equipment

The same recording equipment and set-up from Chapter 2 were used.

3.2.4 Data reduction, dependent measures, and analyses

Two conditions and twenty-five trials in each condition for nine participants yielded 450 trials in total. A total of 40 of the 450 trials were excluded from analysis for the following reasons: anticipation (4 trials; 0.89%), data loss (1 trial; 0.2%), false startle (6 trials, 1.3%), hesitation (5 trials, 1.1%), startle indicator with late RT (>120 ms; 5 trials; 1.1%), and outlier filtering by subject (19 trials, 4.2%), where “outlier” was defined as a voluntary onset of 2 SD or more away from the mean. The remaining 410 trials were examined and further analyzed.

Following the analysis procedure utilized in Chapter 2, EMG activity collected from the SCM and lip muscles was aggregated and analyzed. Mean
EMG activity in relation to the imperative stimulus is shown in the results. Lip displacement trajectories in the Mouthed and Non-speech conditions are also included. It was expected that SCM and the lower lip muscles would show startle reflex activity in response to a SAS. Meanwhile, amplified EMG levels and accelerated lip displacements were also anticipated in SAS-induced responses.

The acceleration of lip movement was determined on the basis of three kinematic markers. As illustrated in Figure 3.1, the latencies of three kinematic markers were measured: (1) the voluntary movement onset (point A), defined as the initiation of lip compression, (2) the lower lip opening onset (point B), defined as the highest point of the lower lip, and (3) the lower lip displacement trough (point C), defined as the lower position of the lower lip. Measurement analyses were performed using a 3 Condition × 2 Stimulus repeated measures ANOVA. Greenhouse-Geisser correction was used for any violation of sphericity. Partial eta-squared ($\eta^2_p$) values are calculated as a measure of effect size.

Relative timing between the kinematic markers A, B, and C was calculated by measuring the time between these markers. The time between voluntary movement onset and lower lip opening onset is shown as $t_1$ (duration of lip compression), while $t_2$ marks the duration of the opening movement from the lower lip opening onset to the lower lip’s lowest position. The time frame for both of these events is $t_3$ (time between points A and C in Figure 3.1), and the relative timing of each kinematic event with respect to the entire time course was calculated. The ratios of $t_1/t_3$, $t_2/t_3$ for each trial
Figure 3.1: Lip displacement and response acoustic waveforms are also plotted with respect to the imperative stimulus. Point A marks the beginning of voluntary movement; point B marks the time to the lower lip opening onset; point C marks the lowest position of the lower lip. The intervals across these markers were labeled as $t_1$, $t_2$, and $t_3$. See texts for details.

(both startle and control) were calculated, with the ratio data transformed by trial using the arcsine square root transformation (this was used due to correlations between markers; McDonald, 2009). The transformed ratios were then aggregated by participant and analyzed using Student’s $t$-tests.

The trials were also analyzed for lip compression. In each experimental trial, lower lip movements were carefully examined in terms of their displacement trajectories. Following Chapter 2, the initiation of the lower lip movement was recorded as the voluntary movement onset (i.e., point A in
Figure 3.1). Since the lips began in a mouth-closed position, after the initiation of the voluntary movement, the upward movement observed in the lower lip is understood as lip compression. Conversely, direct downward movement of the lower lip at the onset of voluntary movement indicates an absence of lip compression. The number of trials with and without lip compression was calculated. Compression displacement (i.e., the vertical height between voluntary movement onset and lower lip opening onset) was also measured and analyzed.

3.3 Results

3.3.1 Lip compression

As discussed in Chapter 2, lip compression was observed at the initiation of voluntary movement in Spoken responses. This compression occurred in 91% of control responses and 87% of startle responses. The frequency of lip compression in the current design was also calculated, with the raw count of trials showing lip compression presented in Table 3.1. As can be seen, lip compression in Mouthed and Non-speech responses did not occur as frequently as in Spoken responses. The proportion of lip compression across control and startle was highest for Spoken, followed by Mouthed, and then Non-speech. For Spoken responses, 91% of trials showed lip compression, whereas for Mouthed responses, the occurrence of lip compression was 56%. Only 22% of Non-speech responses exhibited lip compression. Note that a decrease in the occurrence of lip compression from control to startle trials was also observed in Mouthed responses whereas an increase was observed...
in Non-speech responses.

One may wonder whether a training effect may be responsible for the reduced lip compression frequency in the Mouthed and Non-speech conditions. Additional analyses showed no effect of training. Please refer to Appendix A for further details.

Table 3.1: Numbers of trials with lip compression across conditions. Percentages were calculated from the trials included in analyses.

<table>
<thead>
<tr>
<th></th>
<th>Spoken</th>
<th>Mouthed</th>
<th>Non-speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>145 (91%)</td>
<td>104 (63%)</td>
<td>23 (14%)</td>
</tr>
<tr>
<td>Startle</td>
<td>26 (87%)</td>
<td>10 (24%)</td>
<td>22 (55%)</td>
</tr>
<tr>
<td>TOTAL</td>
<td>171 (90%)</td>
<td>114 (56%)</td>
<td>45 (22%)</td>
</tr>
</tbody>
</table>

Compression displacement (as measured by the vertical height between lower lip voluntary onset and lower lip opening onset) across conditions is summarized in Table 3.2. Only trials that exhibited lip compression were included for the analyses here. In order to perform a $2 \times 3$ ANOVA, missing values were filled in by a constant. Since there was no training effect, missing data was treated as random and substituted with the mean across participants in the same condition.

As the table shows, the largest compression displacement was observed in the Spoken condition, followed by the Mouthed condition, and then the Non-speech condition. Across all three conditions, more compression displacement was observed for startle responses than for control responses. Results revealed that a main effect for Condition ($F(2, 16) = 12.46, p < .01$,
\( \eta_p^2 = 0.61 \) was reported. No effect for Stimulus \((p = 0.06)\) or Condition \(\times\) Stimulus interaction \((p = 0.47)\) was noted. Post-hoc analyses found significant differences between Spoken and Non-speech \((p = 0.027)\) and between Spoken and Mouthed \((p = 0.006)\), but not between Mouthed and Non-speech \((p = 1)\).

### Table 3.2: Mean compression displacement across conditions (mm); standard deviations in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>Spoken</th>
<th>Mouthed</th>
<th>Non-speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>0.96 (0.73)</td>
<td>0.30 (0.41)</td>
<td>0.21 (0.12)</td>
</tr>
<tr>
<td>Startle</td>
<td>1.38 (0.95)</td>
<td>0.48 (0.42)</td>
<td>0.48 (0.50)</td>
</tr>
</tbody>
</table>

To further examine muscle activity and lip displacement during lip compression, EMG data and lip displacement trajectories were normalized to the lower lip voluntary movement onset. This marker was chosen because the lower lip demonstrated more robust displacement than the upper lip. In particular, the closed-mouth position of the lips at onset made it possible for the upper lip to show no obvious displacement at the instant of lip compression. Figure 3.2 illustrates upper lip and lower lip mean EMG activity and lip displacement trajectories in the Spoken condition. As the figure shows, in control responses (grey lines), both upper lip and lower lip produced EMG activity before the lower lip voluntary movement onset. The EMG activity onset preceded the voluntary movement onset due to the motor time required for the muscle to initiate the displacement. EMG ac-
tivity elicited under a SAS was of larger amplitude than that elicited under a control stimulus. Compared to the lower lip, the upper lip did not exhibit obvious displacement until after the lower lip had started its opening movement. That is, while the upper lip was actively engaged during lip compression (as evidenced by EMG activity), no displacement was observed for the upper lip. This absence might be due to either any downward movements being canceled out by the upward movement from the lower lip, or lip compression induced deformation of the lip.

Figure 3.3 depicts lip EMG activity and displacement trajectories in the Mouthed condition. The averaged EMG activity for both lips across control responses in the Mouthed condition was of smaller amplitude than the same measure averaged across Spoken control responses. This reduced EMG activity may be due to fewer trials exhibiting lip compression in the Mouthed condition. As observed with Spoken responses, the upper lip did not generate obvious displacement at the instant of lower lip voluntary movement onset. Displacement was only observed when the lower lip started the opening movement and dragged the upper lip with it.

Upper lip and lower lip mean EMG activity and displacement trajectories from the Non-speech Control condition are shown in Figure 3.4. In Non-speech control responses, the lower lip displayed a clear EMG event, whereas the upper lip did not show much activity during the response. Figure 3.4 shows mean EMG activity of control and startle responses in the Non-speech condition. As revealed in the control responses (Figure 3.4 left), only limited upper-lip EMG activity was observed, appearing to occur later than the lower lip voluntary movement onset (i.e., right to the grey vertical line).
It is likely that the observed upper lip EMG was involved in the lower lip opening movement, rather than voluntary upward compression. SAS-elicited startle reflex activity was more robust in the lower lip than the upper lip. As observed with both Spoken and Mouthed responses, no obvious displacement from the upper lip took place until the lower lip opening onset.
Figure 3.3: Mean upper lip (left) and lower lip (right) EMG activity and displacement trajectories in the Mouthed condition across participants. Data in all channels were normalized to the lower lip opening onset (vertical line). The distance of lip displacement between control and startle responses does not reflect the absolute distance between control and startle responses.

As revealed by Figures 3.2 and 3.3, for Spoken and Mouthed responses, the voluntary movement involved both lips compressing against each other. Non-speech responses, on the other hand, did not involve upper lip and lower lip time-locked activation (Figure 3.4). For Non-speech control responses, mean latency to the EMG onset of the upper lip was 180.3 ms (SD = 65
ms) and 127.8 ms for the lower lip (SD = 45 ms). While the lower lip was active for the voluntary (upward compressing) movement, the upper lip did not show activity until the lower lip started to move downward for the opening movement (at 157 ms from the stimulus onset, see Table 3.3 in the next section). This EMG pattern suggests a lack of voluntary compression
between the upper and lower lips; in this case, displacement from the lower lip is more likely to be anticipatory to the opening movement. It is possible that this anticipatory movement may serve a different function from voluntary compression in the speech-related responses (i.e., Spoken and Mouthed responses).

3.3.2 Startle indicator

In Chapter 2, we observed distinct startle reflex activity in the two SCM and lower lip muscles when a speech response was elicited by a SAS. Figure 3.5 aligns mean EMG activity and lip kinematic trajectories from the Spoken (repeated from Figure 2.6), Mouthed, and Non-speech responses in relation to the stimulus onset. As the figure shows, larger EMG amplitude from the SCM and lips were found in SAS-elicited responses than in control responses. More importantly, for both Mouthed and Non-speech responses, the startle reflex activity was observed in the two SCM and lower lip muscles, but not in the upper lip muscle. A window ranging from 30 ms to 120 ms after the onset of the imperative stimulus was labeled for all three conditions. This window marks the range for startle reflex activity. As revealed, the observed reflex activity is distinct from a second EMG event that is associated with the voluntary mouth opening. This echoes with the findings in Chapter 2.
Figure 3.5: Mean EMG activity and kinematic displacement of control (grey) and startle (black) trials in the Spoken (repeated from Chapter 2), Mouthed, and Non-speech condition. All data were normalized to the imperative stimulus. The top four channels represent the mean EMG across four muscles (from top to bottom): left SCM, right SCM, upper lip, and lower lip. Each channel is of the same scale but vertically arranged for visualization. The bottom two channels depict the mean upper lip and lower lip displacement trajectories, respectively. Red boxes mark the window of startle reflex activity, ranging from 30 ms to 120 ms after the onset of the imperative stimulus.
3.3.3 Kinematic markers, reaction times, and timing

A summary of the results for all measures, including means and standard deviations, are presented in Table 3.3. Note that the lower lip opening onset and lower lip displacement trough were not observed in Chapter 2 for the Spoken response. These measures of Spoken responses are restated and analyzed here for comparison.

Chapter 2 reported that RTs for voluntary movement onset were shorter in SAS-induced responses (Mean= 75 ms) than those in control responses (Mean = 116 ms). In the present experiment the table illustrates slower RT was observed in Mouthed and Non-speech responses when elicited by a SAS. Main effects for Condition (F(2, 16) = 4.44, p = 0.049, \( \eta^2_p = 0.04 \)) and Stimulus (F(1, 8) = 17.87, \( p < .01, \eta^2_p = 0.32 \)) were found. A Condition \( \times \) Stimulus interaction was also reported (F(2, 16) = 5.21, \( p = .02, \eta^2_p = 0.03 \)). Post-hoc analyses using Tukey’s HSD found that for control responses, the reaction time from each task was not significantly different from the others (all \( p > .05 \)). Similarly, no differences with regards to reaction time for startle responses were reported between tasks (all \( p > .05 \)). Across conditions, pairwise comparisons (using a Holm-Bonferroni correction; significance level at 0.01667) confirmed that there were no differences between Spoken and Non-speech (\( p = 0.06 \)) and between Mouthed and Non-speech (\( p = 0.52 \)). However, significant difference was noted between Spoken and Mouthed (\( p = 0.001 \)). Between control and startle responses, pairwise comparisons (using a Holm-Bonferroni correction; significance level at 0.01667) confirmed that latencies of startle responses are significantly shorter than latencies of
control responses in the Spoken ($p = 0.004$) and Non-speech ($p = 0.002$) conditions. However, latencies between control and startle responses were not significantly different in the Mouthed condition ($p = 0.02$).

Considering the lower lip opening onset, shorter latencies were observed in SAS-induced responses from all three conditions. Main effects for Condition ($F(2, 16) = 22.99$, $p <.01$, $\eta^2_p = 0.2$) and Stimulus ($F(1, 8) = 14.78$, $p <.01$, $\eta^2_p = 0.26$) were found, but not in the Condition × Stimulus interaction ($p = 0.27$). Post-hoc analyses using Tukey’s HSD found that for control responses, the reaction time from each task was not significantly different from the others (all $p >.05$). Similarly, no differences with regards to reaction time for startle responses were reported between tasks (all $p >.05$). Comparing across conditions, pairwise comparisons (using a Holm-Bonferroni correction; significance level at 0.01667) confirmed that latencies was longer for Spoken than Silent and than Non-speech (both $p <.001$). No difference was reported between Silent and Non-speech responses ($p = 0.07$).

Similarly, shorter latencies measured from the lower lip displacement trough were observed in SAS-induced responses. Main effects for Condition ($F(2, 16) = 9.76$, $p <.01$, $\eta^2_p = 0.06$) and Stimulus ($F(1, 8) = 11.68$, $p <.01$, $\eta^2_p = 0.29$) were found, but not in the Condition × Stimulus interaction ($p = 0.38$). Post-hoc analyses using Tukey’s HSD found that for control responses, the reaction time from each task was not significantly different from the others (all $p >.05$). Similarly, no differences with regards to reaction time for startle responses were reported between tasks (all $p >.05$). Comparing across conditions, pairwise comparisons (using a Holm-Bonferroni correction; significance level at 0.01667) confirmed that latencies was longer
for Spoken than Silent ($p < .001$) and than Non-speech ($p = 0.002$). No difference was reported between Silent and Non-speech responses ($p = 0.04$).
Table 3.3: Mean reaction times (in ms) to dependent markers across conditions; standard deviations in parentheses. The voluntary movement onsets of Spoken responses from Chapter 2 are included for comparison.

<table>
<thead>
<tr>
<th></th>
<th>Spoken</th>
<th>Mouthed</th>
<th>Non-speech</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>control</td>
<td>startle</td>
<td>control</td>
</tr>
<tr>
<td>Voluntary movement onset</td>
<td>116 (32)</td>
<td>75 (30)</td>
<td>128 (40)</td>
</tr>
<tr>
<td>Lower lip opening onset</td>
<td>198 (60)</td>
<td>142 (44)</td>
<td>157 (51)</td>
</tr>
<tr>
<td>Lower lip displacement</td>
<td>377 (89)</td>
<td>302 (42)</td>
<td>339 (72)</td>
</tr>
</tbody>
</table>
In Chapter 2, it was reported that while shorter latencies at different kinematic markers were observed in startle responses, the relative timing between these kinematic markers remained unaffected. In the current experiment the results also show no significant difference between control and startle values in Mouthed responses for the arcsine square-root-transformed relative time data between voluntary movement onset and lower lip opening onset (i.e., $t_1/t_3$; control Mean = 0.44, startle Mean = 0.47, $p = 0.38$), or between lower lip opening onset and the lower lip displacement trough (i.e., $t_2/t_3$; control Mean = 0.96, startle Mean = 0.94, $p = 0.48$). Timing across markers was also unaffected in the Non-speech condition. No significant difference was found between control and startle values for the arcsine square-root-transformed relative time data between voluntary movement onset and lower lip opening onset (i.e., $t_1/t_3$; control Mean = 0.42, startle Mean = 0.45, $p = 0.58$), or between lower lip opening onset and the lower lip displacement trough (i.e., $t_2/t_3$; control Mean = 0.99, startle Mean = 0.96, $p = 0.55$).

The results suggest that the relative timing across these kinematic markers is not altered by SAS. These relative time frames for control and startle trials are shown in Figure 3.6 with the relative time intervals representing untransformed data.
Figure 3.6: Relative timing ratio (non-transformed) of various kinematic time markers for startle and control trials. The whiskers for each bar are standard error bars. The time frame for these events is the time between lower lip voluntary movement onset and lower lip displacement trough \((t_3)\). Time intervals include the time between voluntary movement onset and lower lip opening onset \((t_1)\) and between lower lip opening onset and lowest lower displacement trough \((t_2)\).
3.4 Discussion

The experiment presented in this chapter was designed to test whether multi-dimensional (in this case aerodynamic) details may be revealed by SAS and whether the startle paradigm has a different impact on Mouthed speech versus Non-speech movements. The results supported our two predictions: (1) Differentiated lip compression frequency was observed for different speech and non-speech tasks, and (2) SAS was not observed to impact the latency of voluntary movements across different tasks. Much as we saw with the Spoken responses in Chapter 2, the results of the current study confirm that the timing profile of lip kinematics for Non-speech and Mouthed responses remains unaffected when elicited by a SAS. These findings provide support for the StartReact effect (i.e., prepared responses are accelerated but unaffected in other measures) and demonstrate that both prepared speech and non-speech oral responses are subject to this effect.

In parallel to the analysis of Spoken responses in Chapter 2, EMG activity from SCM and lip muscles, lip kinematics, and reaction times at different kinematic markers of the Mouthed and Non-speech responses were examined in this study. As shown in Figure 3.5, for both Mouthed and Non-speech responses, SCM and lower lip EMG activity exhibited startle reflex behaviour when triggered by a SAS. This result echoes the findings in Chapter 2. In addition to initiating startle reflex activity, the SAS also produced an accelerated initiation of the prepared response for both Mouthed and Non-speech movements. Although Spoken responses exhibit more frequent, longer, and larger lip compression than Mouthed and Non-speech responses, all these
prepared responses are susceptible to early release by SAS.

The prediction regarding differentiated lip compression for Spoken, Mouthed, and Non-speech movements was supported. While 91% of Spoken responses showed lip compression, only 56% of Mouthed responses and 22% of Non-speech responses demonstrated compression when elicited by a normal control stimulus. For Spoken responses, a SAS did not appear to have an impact on the occurrence frequency of lip compression. Frequent lip compression observed in Spoken responses supports the view that such lip movements are part of the plan for a Spoken task. While non-startled Mouthed responses revealed less lip compression than Spoken responses did overall, the percentage of Mouthed responses with lip compression decreased even further in SAS-elicited trials. As summarized in Table 3.1, lip compression was observed least frequently in the Non-speech condition. Compared with Mouthed responses, a much lower lip compression frequency in Non-speech responses suggests that non-speech lip movement does not require lip compression as part of the preparatory movement. Lip compression may not be directly connected with mouth opening, but may be performed in anticipation of speech behaviours, possibly in anticipation of the need to contain an increase in intraoral pressure associated with an oral stop. It should be noted that EMG activity from the upper lip and lower lip was temporally aligned in Mouthed and Spoken responses, whereas in Non-speech responses, EMG activity from the upper lip was not observed to be temporally aligned with lower lip EMG activity, suggesting a lack of active involvement of the upper lip in Non-speech responses. Meanwhile, lip compression in Non-speech and Mouthed responses is unlikely to be the result of a training effect, since the
initiation of lip compression did not increase in frequency towards the end of the block. As such, lip compression appears to be a speech movement that is speech-specific, rather than a preparatory kinematic movement for the upcoming opening movement. A more general implication is that details about speech motor control for speech and non-speech movements may be revealed by SAS.

In addition to the lower frequency of lip compression, significantly shorter latencies to the lower lip opening onset and lower lip displacement trough for Mouthed and Non-speech responses also characterized a differentiated lip compression from Spoken responses. As a SAS reliably elicits prepared movements, these results suggest that lip compression may not be pre-tuned for Mouthed speech tasks. If lip compression were part of the speech-related motor plan, we should have observed a comparable amount and frequency of compression in both Mouthed and Spoken responses. The contrast in the frequency of lip compression between these two speech modes can be best accounted for by the hypothesis that different muscular structures are involved in the implementation and anticipation of intraoral pressure for Silent and Speech tasks. The implementation of intraoral pressure as part of a speech motor plan requires further investigation.

Compared with Spoken responses, Mouthed responses exhibited less frequent lip compression (Table 3.1). Meanwhile, it is noted that longer latencies to the voluntary movement onset were observed for Mouthed responses when elicited by a SAS (96 ms vs. 75 ms for Spoken and 78 ms for Non-speech; see Table 3.3). Recall that voluntary movement onset correlates with the initiation of lip compression; this being the case, the long latency of
Mouthed responses to the voluntary movement onset may be best accounted for by assuming optionality of lip compression in the prepared program. That is, the choice to incorporate orofacial muscle tensing in a Mouthed response may result in extra online workload and consequently increase the reaction time. Compared with Spoken responses, the performance latency for the Mouthed responses was also longer. As Spoken responses are more “common” than Mouthed responses (speakers perform speech tasks every day and are very experienced with them, whereas mouthing a silent sound is not a commonly performed task), extra inhibitory effects, such as a concerted effort to hold back air flow, may be required for Mouthed responses. This additional inhibition may also lead to increased processing load and longer reaction time for execution. If these inhibitory efforts are included in the speech plan, their timing in relation to other movements may be pre-specified, in which case no delay in reaction time should be observed. The current results reveal the opposite to this prediction, suggesting that a potential inhibitory effort may be implemented after the initiation of the response, consequently resulting in a longer reaction time latency.

As mentioned in the Introduction (Section 3.1), a finding of different breathing patterns would indicate different aerodynamic structures encoded in speech tasks. While Spoken and Mouthed speech tasks may involve different aerodynamics, this kind of differentiation can also be found in different spoken tasks. For example, Gick et al. (2012) show that /p/ and /m/ require different degrees of muscle tension to resist intraoral pressure, even though both sounds are bilabial stops. In the production of /p/, intraoral pressure is built up during lip closure and the perioral muscles are activated in order
to keep the mouth closed and maintain the air pressure within. Compared to /p/, the production of the bilabial nasal /m/, which requires very limited or no aspiration or air pressure, does not require as intensive muscle activity in the perioral areas. Comparing /pa/ with /ma/ would be useful to further test the possibility of implementation of intraoral pressure in the speech plan, though there could be other unknown kinematic differences between theses sounds that would confound such a comparison. The current experiment design did not directly or independently test the implementation of intraoral pressure, or the controlled kinematics of lip compression by comparing spoken /ba/ and mouthed /ba/. Applying the startle paradigm to /ba/ vs. /ma/ would likely improve our understanding of the role of intraoral pressure in the speech plan. This design, which requires more careful analysis of lip kinematics, will be exploited in future studies.

As all of the trials in the current experiment involved prepared sequences, we can conclude that the observed latency differences do not reflect differences in planning, but in execution. In Klapp’s (2003) model, the complexity of a prepared response does not affect reaction time, but the sequence of the elements does. Recently, Maslovat et al. (2014) showed that effects on reaction time may be associated with an inability to prepare timing of the elements of a prepared response. The current results suggest that there may be some other potential mechanisms, such as inhibition of air flow and vocalization, that impact on the reaction time of Mouthed startle responses. Further investigation into the timing of inhibitory efforts will call for future research.
Chapter 4

Startling syllables with pre-speech movements

4.1 Introduction

Chapters 2 and 3 examined the elicitation of prepared Spoken syllables, Mouthed syllables, and Non-speech lip movements. When a SAS is used as an imperative stimulus, shorter latencies of voluntary movement onsets are induced. We also observed lip compression in elicited speech responses. The observed lip compression appeared to be speech-specific as it occurred with significantly greater frequency in Spoken vs. Mouthed or Non-speech responses, suggesting that lip compression may be associated with the need to contain air flow required for Spoken speech and thus to be included in the speech plan. As proposed by McClean and Clay (1995), in connected speech, lip compression is necessary in order to guarantee a full closure for the bilabial sound and to retain possible oral pressure within. This chapter tests whether, in a simple RT task, lip compression that is independent of aerodynamic factors is observed in all speech-related tasks and is elicited at a short latency by SAS. This will be tested by observing the production of
both Spoken and Mouthed syllables as well as Non-speech mouth opening
movements starting from a mouth-open posture, thus effectively adding a
preceeding mouth-closing movement to prepared sequences.

As suggested by the results in Chapter 3, lip compression appears to be
speech-specific and is neither guaranteed nor fully realized for other oromo-
tor tasks. It should be noted that lip compression may also be achieved by
other forces. Löfqvist and Gracco (1997) considered lip compression (i.e.,
negative lip aperture in their terms) as a virtual target for bilabial stops.
Boucher (2008, p. 297) further proposes that such compression implies an
overshoot of the vocal-tract space. As the major contrast distinguishing
Spoken from Mouthed speech is the presence vs. absence of air flow, not
articulatory movements or lip kinematics, it is thus predicted that when
starting with a mouth-open condition, Spoken and Mouthed speech shar-
ing comparable articulatory movements and lip kinematics should generate
comparable lip compression.

Temporally speaking, from a simplistic point of view, one can consider
a mouth-open response as the linear combination of a mouth-closing move-
ment plus a mouth-closed response (cf. Perrier et al., 1996; Sanguineti et al.,
1998). That is, producing a /ba/ from a mouth-open position would be con-
sidered as the production of a mouth-closing movement followed by a pro-
duction of /ba/ from a mouth-closed condition. In a simple RT task, adding
this additional pre-speech closure movement should not affect reaction time.
It is predicted that prepared Spoken /ba/, Mouthed /ba/, and Non-speech
oral movements are subject to early release by SAS and no differences with
regards to reaction time is expected.
This chapter will report on an experiment on SAS-induced responses from a mouth-open position and present data to test the above predictions.

4.2 Methods

4.2.1 Participants

Data were collected and analyzed from the same nine participants reported in Chapter 2 (3 male and 6 female; Mean = 23 years, SD = 4.2 years). Participants were all native speakers of North American English. Prior to the testing, participants signed an informed consent form and were naïve to the questions under investigation. The experiment was conducted following the ethical guidelines established by the University of British Columbia.

4.2.2 Apparatus, task, and procedures

The same apparatus and procedures from Chapter 3 were applied, except that participants were asked to begin each task with their mouths open. The first experiment block required participants to respond to an acoustic stimulus by closing their mouths (hereafter the Non-speech condition). In the second block, participants were instructed to respond to the acoustic stimulus by mouthing a silent articulation of /ba/ (hereafter the Mouthed condition). In the final block, participants responded to the acoustic stimulus by producing a vocalized /ba/ (hereafter the Spoken condition).

The testing procedures used for this experiment were identical to those reported in Chapters 2 and 3. Participants performed a single testing session of approximately 20 minutes. The testing block consisted of twenty control
trials and five startle trials. The five startle trials were presented pseudo-randomly such that the first trial was never a startle trial, nor were there two consecutive startle trials.

4.2.3 Data reduction, dependent measures, and analyses

Nine participants were exposed to three conditions, each containing twenty-five trials for a total of 675 trials. A total of 43 trials were excluded from the analyses for the following reasons: incorrect movement (i.e., starting with mouth closed, 5 trials; 0.74%), hesitation (2 trials; 0.3%), anticipation (5 trials; 0.74%), bad audio recording (1 trial; 0.15%), and data loss (2 trials; 0.3%), and outlying data points (defined as those voluntary onsets being at least 2 SD away from the mean) (28 trials; 4.1%). The remaining 632 trials were examined and further analyzed.

EMG activity measured from the SCM and lip muscles were aggregated and analyzed. Individual trials were first aggregated by participants and conditions. Mean EMG activity was taken across aggregated by-participant means. Lip displacement trajectories from all the conditions are also included. It was expected that SCM and the lower lip muscles would show startle reflex activity in response to the SAS; amplified EMG levels and accelerated lip displacements were also anticipated.

Latencies to three kinematic markers (denoting the lower lip voluntary onset, lower lip opening onset, and lower lip displacement trough). In the current experiment design, participants started with their mouths open. Within this context, the lower lip voluntary onset was defined as the instant when the lower lip started to move upward for the closure (point A in Figure
Figure 4.1: Schematic Spoken response with lip displacement and response acoustic waveforms plotted with respect to the imperative stimulus. Point A marks the beginning of voluntary movement; point B marks the time to the lower lip opening onset; point C marks the acoustic onset; point D marks the lowest position of the lower lip. The intervals across these markers were labeled as $t_1$, $t_2$, $t_3$, $t_4$, and $t_5$. See text for details.

4.1); the lower lip opening onset was defined as the instant when the lower lip started to descend for the opening burst of the syllable /ba/ (point B in Figure 4.1); the lower lip displacement trough was defined as the lowest point reached by the lower lip during the opening movement (point D in Figure 4.1).

To measure the timing of the initial acoustic burst release, the acoustic signals were first filtered (at a level of 10Hz) using PRAAT (Boersma and Weenink, 2009). A narrow-band spectrogram (bandwidth of 43 Hz;
wavelength of 30 ms) was displayed in PRAAT to mark the acoustic burst release (point C in Figure 4.1). Reaction times to points A, B, C and D were measured and analyzed. Measurement analyses were performed using a 3 Condition × 2 Stimulus repeated measures ANOVA. Greenhouse-Geisser correction was used for any violation of sphericity. Partial eta-squared ($\eta_p^2$) values are calculated as a measure of effect size.

Relative time between the kinematic markers A, B, C, and D was measured (Figure 4.1), with the time differences labelled as follows: $t_1$ measures the latency between voluntary movement onset and lower lip opening onset; $t_2$ measures the latency between opening movement and acoustic onset; $t_3$ measures the latency between lower lip voluntary onset and acoustic onset; $t_4$ measures the latency between lower lip opening onset and lower lip displacement trough. The time frame for all of these events is $t_5$ (the time between points A and C in Figure 1); the latency of each kinematic event with respect to $t_5$ was also calculated. For each trial (startle and control), the ratios $t_1/t_5$, $t_2/t_5$, $t_3/t_5$, and $t_4/t_5$ were calculated and transformed using the arcsine square root transformation (this was used due to correlations between markers; McDonald, 2009). The transformed ratios were then aggregated by participant and analyzed via Student’s $t$-tests.

In Chapter 3, we observed lip compression during the production of elicited bilabial stops. The current design asked participants to begin each task with their mouths in an open position. As predicted by Löfqvist and Gracco (1997), a negative lip aperture is anticipated in the production of the bilabial /b/. From a mouth-open position, the two lips start to move toward each other. After the contact, the lower lip, which usually induces greater
force, pushes against the upper lip and reverses its direction of movement (from downward to upward); this is shown as Point M in Figure 4.2.

![Figure 4.2: Lip compression markers. Points M and N mark the onset and offset of lip compression, respectively (see text for details). Lip compression is here defined as the latency between Points M and N ($t_c$). Lip compression displacement measures the change in lower lip displacement within $t_c$.](image)

The opening burst is mainly determined by the lower lip (and the jaw). As the two lips come into contact, the initiation of opening from the lower lip affects upper lip movement, pulling the upper lip downward when the opposite forces between the two lips are not strong enough to separate them. Once the opposing force generated by the two lips becomes greater than the stickiness between them, the two lips move apart. This series of steps produces a w-shaped movement trajectory in the upper lip (the upper channel in Figure 4.2).

To measure lip compression during the mouth-open experiment, two tem-
Figure 4.3: Schematic lip displacement trajectories (top) from the upper and lower lips and correspondent lip aperture profile (bottom). Data was taken from one trial from one of the participants. The two lips reach an equilibrium point and lip aperture stabilizes. Data was taken from one example trial from a participant. The vertical grey line marks the onset of lip compression. See text for details.

poral points were marked. The onset of lip compression was defined as the point at which the movement trajectory of the upper lip changes (Point M
in Figure 4.2). Arrival at this point reliably indicates that the two lips have come into contact and that the closure force from the lower lip is affecting the upper lip movement trajectory. The offset of lip compression was defined by the lower lip (Point N in Figure 4.2), which is a reliable agonist for the opening movement. The duration of lip compression \( t_c \) was defined as the latency between Points M and N in Figure 4.2. Lip compression duration was calculated as the interval between the onset and offset of lip compression \( t_c \). Lip compression displacement was calculated as the change in lower lip vertical displacement between the onset and offset of lip compression \( t_c \).

Lip compression measurements were then analyzed by a 2 Stimulus (control vs. startle) × 2 Condition (Mouthed vs. Spoken) repeated-measures ANOVA.

Figure 4.3 shows schematic lip displacement trajectories and the correspondent lip apertures. The upper and lower lip data from one trial from one participant was plotted. The lip displacement trajectories in relation to the imperative stimulus (the grey line in Figure 4.3) were illustrated in the top figure of Figure 4.3. To obtain lip aperture data, the absolute vertical difference between the upper and lower lips were calculated. The vertical height of the lower lip was subtracted from the vertical height of the upper lip (Figure 4.3 bottom). When participants initiate the prepared response, the two lips move closer to each other and lip aperture starts to decrease. When the two lips make contact, lip aperture stabilizes (bottom of Figure 4.3). However, while the lower lip is moving upward and pushing against the upper lip, the upper lip may continue moving downward for the lip closure. Lip aperture does not stay constant; an inflection is observed in the
lip aperture profile in Figure 4.4.

Figure 4.4: Schematic lip displacement trajectories (top) from the upper and lower lips and correspondent lip aperture profile (bottom). Data was taken from one trial from one of the participants. This figure shows that lip aperture continues to decrease after the two lips make contact. The vertical grey line marks the onset of lip compression. An inflection of lip aperture profile is observed. See text for details.
4.3 Results

4.3.1 Startle indicator

For all three conditions, mean EMG activity and lip displacement with respect to the imperative stimulus were plotted in Figures 4.5. Across all three conditions, startle reflex activity was observed in both SCM muscles and the lower lip in SAS-induced responses, whereas only limited activity was observed in the upper lip muscles in the same context. In each case, the onsets of startle reflex activity in the SCM muscles and lower lip occurred within 100 ms of the imperative stimulus.

As observed in mouth-closed responses (Chapters 2 and 3), the observed startle reflex activity in the current mouth-open condition was consistently present in SCM and lower lip muscles, distinct from EMG activity for voluntary movements, and not task-dependent. In the current mouth-open design, voluntary movement started at the instant when both lips came towards each other for the closure movement. Unlike in the mouth-closed experiments, the upper lip was engaged in the voluntary closure movement and showed EMG activity from the initiation of the closure. For all three conditions, a larger lip displacement range was observed in SAS-elicited responses.
Figure 4.5: Mean EMG activity and kinematic displacement of control (grey) and startle (black) trials in the Spoken, Mouthed, and Non-speech condition. All data were normalized to the imperative stimulus. The top four channels represent the mean EMG across four muscles (from top to bottom): left SCM, right SCM, upper lip, and lower lip. Each channel is of the same scale but vertically arranged for visualization. The bottom two channels depict the mean upper lip and lower lip displacement trajectories, respectively. Red boxes mark the window of startle reflex activity, ranging from 30 ms to 120 ms after the onset of the imperative stimulus.
4.3.2 Kinematic markers, reaction times, and timing

A summary of reaction times derived from kinematic markers is presented in Table 4.1. Concerning lower lip voluntary movement onset, longer latencies were observed for control responses than for SAS-elicited responses across all three conditions. Two-way ANOVA results reported a main effect for Stimulus ($F(1, 8) = 24.78, p < .01, \eta^2_p = 0.29$). No effect was found for Condition ($p = 0.054$) or Condition $\times$ Stimulus ($p = 0.7$).

Table 4.1: Mean reaction time (in ms) at dependent markers across conditions; standard deviations are in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>Spoken</th>
<th>Mouthed</th>
<th>Non-speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voluntary movement onset</td>
<td>Control</td>
<td>119 (26)</td>
<td>136 (31)</td>
</tr>
<tr>
<td></td>
<td>Startle</td>
<td>91 (22)</td>
<td>101 (27)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>99 (22)</td>
<td></td>
</tr>
<tr>
<td>Lower lip opening onset</td>
<td>Control</td>
<td>265 (69)</td>
<td>282 (84)</td>
</tr>
<tr>
<td></td>
<td>Startle</td>
<td>210 (28)</td>
<td>230 (37)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Lower lip displacement trough</td>
<td>Control</td>
<td>423 (94)</td>
<td>461 (125)</td>
</tr>
<tr>
<td></td>
<td>Startle</td>
<td>352 (46)</td>
<td>382 (68)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>–</td>
<td></td>
</tr>
</tbody>
</table>

Latency to the lower lip opening onset was only measured and analyzed for Mouthed and Spoken responses. As reported above for the lower lip voluntary movement onset, shorter latencies were observed in SAS-elicited responses than control responses. Results showed main effects for Condition ($F(1, 8) = 5.96, p = 0.04, \eta^2_p = 0.03$) and for Stimulus ($F(1, 8) = 7.39, p = 0.03, \eta^2_p = 0.19$), but no Condition $\times$ Stimulus interaction ($p = 0.68$). Compared to the control responses, startle responses showed shorter latencies before the lower lip displacement trough was reached. A main effect
for Stimulus was found ($F(1, 8) = 10.6, \ p = .01, \ \eta^2_p = 0.17$). No other effect or interaction was reported. Note that standard deviations for both Mouthed and Spoken responses gradually increased from voluntary movement onset, to lower lip opening onset, and then lower lip displacement trough. This increase of standard deviation was driven by one participant (P19) exhibiting much longer lip compression during the closure. For this participant, longer latencies to the lower lip opening onset and displacement trough were observed. These longer latencies resulted in the observed large standard deviations across participants.

While the lower lip initiated the voluntary movement at around the same time for all three conditions, the lower lip varied in the time taken to reach its peak position in the closure (i.e., the opening onset) and its lowest position (i.e., displacement trough). As summarized in Table 4.1, shorter latencies to these markers were observed for Spoken responses than for Mouthed responses. These shorter latencies may be due to the presentation order of the blocks. In the current design, the presentation order of the blocks was fixed: participants performed all tasks in the order Non-speech > Mouthed > Spoken. As they were asked to begin each response from a mouth-open condition, the degrees of opening decreased as the tasks proceeded. The largest opening (measured by the distance between the two lips at the lower lip voluntary onset) was observed for Non-speech (Mean = 23.5 mm, SD = 8.6 mm), followed by Mouthed (Mean = 18.09 mm, SD = 8.3 mm), and then Spoken responses (Mean = 16.79 mm, SD = 8.1 mm). As such, the shorter latencies to lower lip peak position observed in the Spoken condition may be accounted for by the differences in the degree of opening at the starting
While shorter latencies at different kinematic markers were observed in startle responses, the relative timing between these kinematic markers remained unaffected. For Non-speech responses, only the latency between lower lip voluntary movement onset and lower lip opening onset (i.e., $t_1$ in Figure 4.1) was measured. The t-test result showed no significant difference between control (Mean = 204 ms, SD = 37 ms) and startle responses (Mean = 206 ms, SD = 45 ms; $p = 0.93$). For Mouthed responses, the lower lip kinematic makers at the lower lip’s voluntary movement onset, opening onset, and displacement trough were measured. The $t$-test results showed no statistical difference between control and startle values for the arcsine square root transformed relative time data between voluntary movement onset and lower lip opening onset (i.e., $t_1/t_5$; control Mean = 0.68, startle Mean = 0.69, $p = 0.48$) or between lower lip opening onset and lower lip displacement trough (i.e., $t_4/t_5$; control Mean = 0.77, startle Mean = 0.75, $p = 0.52$).

For Spoken responses, no significant difference was reported between control and startle values for the arcsine square root transformed relative time data between voluntary movement onset and lower lip opening onset (i.e., $t_1/t_5$; control Mean = 0.70, startle Mean = 0.69, $p = 0.54$), between lower lip opening onset and voice onset (i.e.,$t_2/t_5$; control Mean = 0.63, startle Mean = 0.52, $p = 0.14$), between voluntary movement onset and voice onset (i.e., $t_3/t_5$; control Mean = 0.85, startle Mean = 0.81, $p = 0.21$), or between lower lip opening onset and displacement trough (i.e., $t_4/t_5$; control Mean = 0.74, startle Mean = 0.76, $p = 0.53$). As in the mouth-closed
conditions (Chapters 2 and 3), the relative timing across these kinematic markers in the mouth-open condition was not altered by SAS-elicitation. The relative time frames for control and startle trials are illustrated in Figure 4.6, with the relative time intervals representing untransformed data.

Figure 4.6: Relative timing ratio (non-transformed) of various kinematic and acoustic time markers for startle and control trials. The time frame for these events is the time between lower lip voluntary movement onset and lower lip displacement trough ($t_5$). Time intervals include the time between: voluntary movement onset and lower lip opening onset ($t_1$), lower lip opening onset and voice onset ($t_2$), voluntary movement onset and voice onset ($t_3$), and lower lip opening onset and lower lip displacement trough ($t_4$).
4.3.3 Lip kinematics

Table 4.2 summarizes lip compression onset latency and duration in Mouthed and Spoken responses. Observed lip compression onset latencies (illustrated by the onset latency for Point M in Figure 4.2) were longer for Mouthed responses than for Spoken responses, and longer for control responses than for startle responses. Main effects for Condition ($F(1, 8) = 17.61, p < .01$) and Stimulus ($F(1, 8) = 25.07, p < .01$) were observed. No Condition $\times$ Stimulus interaction was reported ($p = 0.76$). Mouthed and Spoken responses showed no difference for lip compression duration ($p = 0.91$). While shorter mean lip compression durations were observed in startle responses than in control responses, the difference was not significant ($p = 0.5$). No interaction of Condition $\times$ Stimulus was reported ($p = 0.86$).

Table 4.2: Mean lip compression onset latency (in ms), lip compression duration (in ms) and lip compression displacement (in mm); standard deviations are in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>Mouthed</th>
<th>Spoken</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lip compression onset latency (ms)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>233 (54)</td>
<td>216 (47)</td>
</tr>
<tr>
<td>Startle</td>
<td>190 (37)</td>
<td>169 (29)</td>
</tr>
<tr>
<td>Lip compression duration (ms)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>59.9 (66)</td>
<td>57.5 (38)</td>
</tr>
<tr>
<td>Startle</td>
<td>50.5 (28)</td>
<td>49.7 (19)</td>
</tr>
<tr>
<td>Lip compression displacement (mm)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>3.70 (4.3)</td>
<td>4.09 (4.5)</td>
</tr>
<tr>
<td>Startle</td>
<td>5.93 (5.6)</td>
<td>5.05 (5.1)</td>
</tr>
</tbody>
</table>

No main effect for Condition ($p = 0.61$) or Stimulus ($p = 0.11$) was reported with regards to lower lip compression displacement, whereas a
Stimulus × Condition interaction was found (F(1, 8) = 6.18, \( p = 0.04 \), \( \eta_p^2 = 0.004 \)).

Compared with the upper lip, the lower lip showed more robust EMG activity and displacement. To observe EMG activity and lip kinematics of voluntary responses, both upper lip and lower lip data were normalized to the lower lip voluntary onset (point A in Figure 4.1). Figures 4.9, 4.8, and 4.7 present mean EMG activity and lip displacement trajectories for the Spoken, Mouthed, and Non-speech conditions, respectively. Across all three

Figure 4.7: Upper lip (left) and lower lip (right) EMG activity and displacement trajectories in the Spoken condition. Data in all channels were normalized to the lower lip opening onset (vertical line).
conditions, for both lips, larger EMG activity and displacement ranges were observed in startle responses than in control responses. Compared with the upper lip, the lower lip generated more EMG activity and exhibited larger displacement for both control and startle responses. It is noted that,

Figure 4.8: Upper lip (left) and lower lip (right) EMG activity and displacement trajectories in the Mouthed condition. Data in all channels were normalized to the lower lip opening onset (vertical line).

for Mouthed and Spoken responses, the lower lip demonstrated two EMG events, whereas the upper lip only showed one instance of EMG activity. The first EMG event measured in the lower lip was temporally bound to the event observed in the upper lip. This EMG event is considered to be
related to the closing movement. Comparable EMG activity for the closure movement was also seen in Non-speech responses (Figure 4.9). The second EMG activity observed in the lower lip, on the other hand, was responsible for the opening of the bilabial burst.

Figure 4.9: Upper lip (left) and lower lip (right) EMG activity and displacement trajectories in the Non-speech condition. Data in all channels were normalized to the lower lip opening onset (vertical line).
4.4 Discussion

The results of this chapter show that the StartReact effect is observed for utterances produced from a mouth-open starting position for all three conditions in this study (Spoken, Mouthed, and Non-speech), as evidenced by the presence of startle reflex activity in the SCM and the lower lip and by accelerated voluntary movement onset for the prepared response. All the responses were elicited at shorter latencies by a SAS; stronger EMG activity and larger displacement amplitudes were observed for SAS-induced responses than for control responses. As predicted, adding an additional pre-speech closure movement does not affect reaction time. Earlier onset of lip compression for Spoken responses than Mouthed responses were consistent with the smaller degree of mouth opening observed for Spoken responses. Meanwhile, relative timing across different kinematic markers was unaffected by SAS (Figure 4.6).

Comparing Tables 3.3 with 4.1, for both Spoken and Mouthed syllables, voluntary movement onsets for mouth-closed and mouth-open responses were comparable. It is therefore suggested that in a simple RT task, adding an additional preparatory movement does not appear to yield any additional delay of execution. The present study took the position that the mouth-open condition is a temporal combination of a mouth-closing movement and a mouth-closed condition. While an overshoot from the mouth-closing movement is expected, such an overshoot does not contribute more lip compression to the existing mouth-closed speech plan. Lip compression appears to fall within a narrow range, regardless of the starting position.
of the mouth. Further research would be needed to determine whether the compression from the mouth-closed speech plan has been overwritten or otherwise modified. Future research into lip compression would be needed.

Different degrees of mouth opening may affect the timing, strength, and duration of lip compression. In the current design, the degree of mouth opening in each participant’s initial posture was not controlled for. Given different beginning positions, different lip forces and lip movement velocities may be generated, which in turn may affect lip compression kinematics. The observed lip deformation may thus be the result of pure biomechanical interaction between the two lips. In instances where the upper lip is affected by the lower lip, it is not always clear that the upper lip has been pushed upward and forced to change its trajectory. In fact, after contact between the two lips is initiated, the influence coming from the lower lip may not be great enough to change the moving trajectory of the upper lip. The force generated by the upper lip may be strong enough to resist the impact from the (upward-moving) lower lip. As a result, the upper lip may only be affected in terms of its velocity, while maintaining its downward movement. In turn, measured lip compression may not necessarily reflect differentiated lip compression for Spoken and Mouthed responses.

To summarize, the results show that preparatory oral movements can be included in a speech plan and may be subject to rapid release by a SAS. While lip compression is speech-specific in the mouth-closed condition, it is not unique to Spoken speech for the mouth-open condition.
Chapter 5

Pitch planning in English and Taiwanese Mandarin

5.1 Introduction

Previous chapters have demonstrated that the SAS experimental paradigm is capable of uncovering the content of speech plans. Using this method, a prepared CV syllable can be elicited at shorter latencies, and prepared CV syllables can be performed rapidly and accurately, though some aspects of coordination may be disturbed. The present chapter tests whether phonemic and non-phonemic aspects of pitch are included in a speech plan.

As reported in Chapter 2, prepared English CV syllables elicited by SAS are initiated at shortened latencies, with lower-lip movements for /ba/ initiated earlier in SAS-elicited responses (Mean = 75 ms) than in control responses (Mean = 116 ms). In English SAS-triggered syllable production, throughout the syllable, lip kinematics and vowel F1 and F2 (Figure 5.1) are unaffected by the presentation of a SAS, indicating that these parameters are encoded in the speech plan and are not dependent on afferent feedback. The lack of affect on F1 and F2 indicates that SAS-induced CV syllables
share comparable vowel quality with control responses. These responses are therefore considered “accurate.” However, in addition to formants, other measurements, such as pitch (i.e., fundamental frequency $f_0$) and speaking rate, may also indicate different properties of the syllable. Results for parameters relating to pitch control, such as pitch height and contour profile, are not discussed in Chapter 2. This leaves an important gap, as it remains unclear whether or not SAS elicitation impacts pitch. If a SAS imposes any effects on pitch, at least three questions ought to be answered. First, how is pitch affected by SAS? Second, are SAS-induced responses with affected pitch still considered “accurate?” Third, if these SAS-induced responses are not accurate, is there any correction or adjustment based on feedback? To answer these questions, the current chapter investigates the pitch height and contour of SAS-induced responses.

The SAS paradigm is ideally suited for investigating pitch control, both because a SAS can elicit rapid response of prepared movement sequences, allowing less time for effects of sensory feedback, and because the associated startle reflex provides a natural, physiological pitch perturbation. Following presentation of a SAS, a brief physiological startle reflex response occurs (muscle pre-motor reaction time around 40 ms), followed by accelerated release of the prepared movement sequence (see Stevenson et al., 2014, for details); by the time of the voice onset (Mean = 204 ms), this initial startle reflex has long since been resolved. In an experiment conducted by Baer (1979), startle responses were observed when loud clapping was introduced near the participants’ ears. The participants’ continuous phonation was perturbed by this startling acoustic stimulus, with a consequent increase
Figure 5.1: SS ANOVA comparison of formant frequencies F1 and F2 over syllable /ba/ durations. The black line denotes the predicted fit for control responses and the white line denotes the startle responses. The dark grey and light grey bands surrounding the predicted fits represent a 95% confidence interval. Any white space between the transition lines for each condition represents a statistically significant difference between the given measures. NB: The figure is a revised version of Figure 2.8.

in the fundamental frequency of the phonation at a latency of 50 ms (see Baer, 1979, for details). The perturbation lasted for approximately 100 ms, after which the overall frequency of phonation was higher than before the perturbation. Baer argues that tension in the larynx increases after the perturbation due to a protective closure reflex that occurs in response to the unexpected loud sound. When an unexpected startling auditory stimulus was presented during constant phonation, the startle reflex caused an increase in laryngeal tension resulting in a momentary elevation in pitch height, followed by a correction (i.e., dropoff) back to baseline level.
Extrapolating from Baer (’s 1979) findings for constant phonation to the production of spoken syllables, let us consider the sequence of pitch-related events beginning with presentation of the SAS. First, a speaker who is planning to produce a spoken syllable hears a SAS; after about 40 ms, the first premotor reaction associated with the startle reflex can be observed, initiating an increase in laryngeal tension. As soon as 20 ms after the startle reflex has begun to perturb laryngeal tension, any initial somatosensory (e.g., proprioceptive) feedback-based correction may begin to take place (cf. Larson et al., 2008). Over 100 ms later, the physiological reflex and the initial response having resolved, voicing onset occurs (∼ 204 ms); vocal fold vibration offers a second opportunity for somatosensory (e.g., vibrotactile, aerotactile) feedback to take effect at a further latency of ∼ 20 ms after the pitch onset (cf. Larson et al., 2008). If auditory feedback plays a role in correction, compensation would occur at an additional latency from as early as 100 ∼ 150 ms (Hain et al., 2000) up to 210 ms (Jones and Munhall, 2002) following voicing onset. Thus, if initial pitch height is pre-specified in the speech plan (i.e., if there exists a specified baseline to which the system might be corrected), proprioceptive feedback streams should have had sufficient time to correct for SAS perturbation to laryngeal configuration prior to vowel onset. However, if initial pitch height is not part of the speech plan, an uncorrected increase in onset pitch height may be expected to persist in SAS-induced responses. Taking the above observations into account, we expect vowel-initial pitch height to be elevated by SAS perturbation. Figure 5.2 illustrates windows for potential somatosensory and auditory feedback for a SAS-induced response.
While elevated pitch height is anticipated in response to a SAS, it is also necessary to consider whether perturbed pitch control affects response accuracy. In the experiment reported in Chapter 2, the target response was a nonsense syllable /ba/. Since this is not a word in English, perturbed pitch control cannot produce a meaning change or shift. Therefore, response accuracy cannot be determined in this design. In addition, even for a grammatical English word, such as /ti/ (‘tea’), pitch height does not create a meaning contrast. For example, /ti/ produced in a higher pitch range, from a 3-year-old toddler, conveys the same meaning /ti/ produced in a much lower range by a 65-year-old man. Thus, examining SAS-induced English pitch alone cannot sufficiently address the question of elicited response ac-

Figure 5.2: Schematic timeline of potential feedback for SAS-elicited speech
accuracy.

Mandarin Chinese, on the other hand, serves as a qualified candidate to investigate this issue. As a tonal language, Mandarin Chinese uses differentiated pitch level and contours to distinguish word meanings. The syllable /ba/ can be combined with four lexical tones to produce four different lexical content words (cf. Table 5.1). Studying SAS-induced Mandarin pitch can thus reveal whether or not pitch contour profiles are affected by SAS, resulting in a change in word meaning. If pitch contour profiles are prespecified in the speech plan, they should not be perturbed by a SAS. In this case, the induced responses can be considered accurate responses.

<table>
<thead>
<tr>
<th>Matched syllable</th>
<th>Tone</th>
<th>Pitch control</th>
<th>Word meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ba/</td>
<td>Tone 1</td>
<td>High-level</td>
<td>‘eight’</td>
</tr>
<tr>
<td></td>
<td>Tone 2</td>
<td>Rising</td>
<td>‘pull’</td>
</tr>
<tr>
<td></td>
<td>Tone 3</td>
<td>(Mid-) Falling (-rising)</td>
<td>‘target’</td>
</tr>
<tr>
<td></td>
<td>Tone 4</td>
<td>Falling</td>
<td>‘father’</td>
</tr>
</tbody>
</table>

Compared with pitch height, pitch contour in Mandarin appears to be fairly stable, possibly because of its more crucial role in maintaining lexical contrasts. Liu et al. (2007), for example, showed that while Mandarin-speaking mothers use higher pitch level to address infants than to address adults, the contours associated with phonemic contrasts are maintained and even exaggerated between adult-directed and infant-directed speech. Thus, lexical contrasts can be maintained by preserving the contour profiles that
are used to differentiate lexical tones, despite changes in absolute pitch height. If absolute pitch height is part of the speech plan, a correction to baseline level in SAS-induced responses is anticipated at some point during the vowel (i.e., auditory feedback-based correction), compromising pitch contour. However, if pitch contour is pre-specified but absolute height is not (as we might expect from previous literature we described above), the contour profile should be preserved when elicited by a SAS, even if at a higher pitch range. Thus, while we predict an elevated pitch height at vowel onset, we do not expect to see a rapid dropoff in pitch signifying feedback-based correction following the physiological response. Rather, we expect to see pitch contours maintained, though at an elevated level.

The current experiment was designed to examine how pitch in tonal and atonal languages is produced in response to a SAS. English, a non-tonal language, is compared with a Mandarin dialect spoken in Taiwan (hereafter Taiwanese Mandarin). The issue of how pitch control affects response accuracy under SAS elicitation is discussed.

5.2 Methods

5.2.1 Participants

Fifteen Taiwanese Mandarin native speakers were recruited for the study, of which seven (2 male and 5 female) exhibited consistent startle reflex in response to SAS (e.g., shoulder shrugging or facial grimace); all the data collected from these seven participants were analyzed. Prior to testing, participants signed an informed consent release and were left unin-
formed about the research question under investigation. The experiment was conducted following the ethical guidelines established by the University of British Columbia (UBC). For the Taiwanese Mandarin experiment, all the testing was conducted in a sound-attenuated room at National Chiao Tung University, Hsinchu, Taiwan. Data from Stevenson et al. (2014) were reanalyzed to identify SAS-induced responses in English for comparison. Apparatus and procedures were as described in Stevenson et al. (2014), except for minor changes as described here.

5.2.2 Apparatus, task, and procedures

Participants sat in an upright chair facing a computer monitor (Viewsonic, VE175, 17”, 75 Hz refresh rate) at a distance of approximately 0.5 meters and were instructed to look straight ahead at the monitor and respond to an acoustic stimulus by vocalizing the target syllable as quickly as possible. A visual display of the syllable /ba/ was presented on the monitor concurrently with an acoustic stimulus of either 80 dB (control) or 124 dB (startle). Participants were instructed to respond to the auditory stimulus by vocalizing the target syllable /ba/ while acoustics and video were collected.

In the Taiwanese Mandarin experiment, the syllable /ba/ was matched with each of the four lexical tones and presented in blocks, so that all syllables in each block contained the same tone (in the order of Tone 1, 2, 3, and 4). Target syllables were visually displayed using Taiwanese phonetic orthography (Zhuyin bopomofo), with tonal information specified.

All testing trials began with a warning tone (500 ms, 440 Hz, 80 dB).
The acoustic imperative stimulus and the visual display of the target syllable followed the warning tone by a random time delay of between 1500 and 2500 ms. This auditory signal was either a control stimulus (80 ± 2 dB, 100 ms, 1,000 Hz) or a startling stimulus (124 ± 2 dB, 40 ms, 1,000 Hz, <1ms rise time). The acoustic stimuli were amplified (R Long Audio amplifier AK302) and then presented via a loudspeaker placed directly behind the head of the participant. The acoustic stimulus intensities were measured using a sound-level meter (TES sound level meter model: 1350A, “A” weighted scale, impulse response mode) at a distance of 30 cm from the loudspeaker (approximately the distance to the ears of the participant). Prior to the testing session, a baseline startle trial (124 ± 2 dB, 40ms, 1,000 Hz, <1ms rise time) was introduced.

Participants performed a single testing session of approximately 15 minutes. The testing block consisted of twenty control trials and five startle trials. The five startle trials were presented pseudo-randomly such that the first trial was never a startle trial, nor were there two consecutive startle trials.

5.2.3 Recording equipment

Acoustic production was recorded by the computer program PsycoPy (www.psychopy.org). Acoustic data were collected by a microphone directly connected to the computer. Data were sampled at 44.1 K Hz; collection began 160 ms before the presentation of the stimulus and was terminated 2000 ms later. Video was recorded using Lumix digital camera (Panasonic Model DMC-LX3).
5.2.4 Data preparation and statistical analyses

A total of 34 of the 700 Taiwanese Mandarin trials (4.8%) were excluded from analysis, due either to slow voice onset (defined as 2 SD from the mean) or to participants’ unfamiliarity with the trial event at the beginning of the testing. For both English and Taiwanese Mandarin, acoustic signals were processed using Praat (Boersma and Weenink, 2009, http://www.fon.hum.uva.nl/praat/). Syllable boundaries were visually examined by the experimenter and manually marked. Before marking syllable boundaries, the acoustic waveforms were first filtered via de-emphasis at 50 Hz.

The voice onset was marked based on the appearance of substantial acoustic wave bursts; the end of the syllable was marked as the end of periodic waveforms. Acoustic burst release times and syllable durations were submitted as dependent measures and analyzed using paired Student’s t-tests.

The formants were extracted using the LPC formant tracker, with a window from 0 to 2000 Hz, and a window length of 0.0125 s. Data were then analyzed across normalized durations. Voicing onset was identified and marked based on the appearance of substantial acoustic wave bursts; the end of the syllable was identified and marked as the end of the periodic waveforms. Acoustic burst release times and syllable durations were used to measure reaction time and submitted as dependent variables. Measurement analyses were performed using a 2 Stimulus × 4 Tone repeated measures ANOVA. Pitch and formant transitions from both English and Taiwanese Mandarin were analyzed using a smoothing spline analysis of variance (SS
ANOVA) to examine the fit of each variant (cf. Davidson, 2006; Derrick and Schultz, 2013).

5.3 Results

As with the English results reported in Stevenson et al. (2014), in Taiwanese Mandarin, significantly shorter latencies for acoustic burst release time were reported for SAS-induced responses than control responses (Table 5.2). Main effects were observed for Stimulus \((F(1, 6) = 10.87, p = 0.02)\), Tone \((F(3, 18) = 8.94, p < .01)\), and the Stimulus \(\times\) Tone interaction \((F(3, 18) = 4.9, p = 0.01)\). Post-hoc analyses revealed that the reaction time latency for Tone 4 was significantly shorter than that for Tone 1 and Tone 2. The main effects for Tone and Stimulus \(\times\) Tone interaction may be interpreted as a practice effect, as reaction time latencies gradually decreased from Tone 1 to Tone 4. Syllable durations are summarized in Table 5.2. No effect for Stimulus or Tone was reported. No statistically significant difference was found for either F1 or F2 (as indicated by the overlaps between the control and startle fits in Figure 5.3), suggesting that formant profiles, acting in a phonemic role, are preserved in SAS-induced responses.
Table 5.2: Average acoustic burst release times and syllable durations (in ms) across 4 Taiwanese Mandarin tones. Standard deviations are shown in parentheses.

<table>
<thead>
<tr>
<th>Tone</th>
<th>Burst release time</th>
<th>Syllable duration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>control</td>
<td>startle</td>
</tr>
<tr>
<td>Tone 1</td>
<td>297.07 (70)</td>
<td>225.49 (49)</td>
</tr>
<tr>
<td>Tone 2</td>
<td>272.37 (67)</td>
<td>233.01 (48)</td>
</tr>
<tr>
<td>Tone 3</td>
<td>244.24 (74)</td>
<td>207.96 (55)</td>
</tr>
<tr>
<td>Tone 4</td>
<td>209.40 (68)</td>
<td>180.06 (43)</td>
</tr>
</tbody>
</table>

Figure 5.3: Taiwanese Mandarin F1 and F2 range for control and startle responses. SS ANOVA results of F1 and F2 frequency range across 4 tones. Each shaded band represents a 95% confidence interval. The white space between the transition lines for each condition represents a statistically significant difference between the given measures.

For English, SAS-induced pitch levels were significantly higher than those of control responses, as indicated by the distinct SS ANOVA curves shown in Figure 5.4. The onset frequency in SAS-induced responses was
elevated by 40 ~ 50 Hz, compared to control responses. It is also noted that pitch contour, although not phonemic in English, was unaffected by the SAS.

![Figure 5.4: SS ANOVA comparison of pitch over English /ba/ durations.](image)

Each shaded band represents a 95% confidence interval. The white space between the transition lines for each condition represents a statistically significant difference between the given measures.

For Taiwanese Mandarin, as with English, SAS-elicited pitch levels were elevated and distinguished from control responses throughout most of their duration (Figure 5.5). At vowel onset, substantial pitch elevation was observed for all four tones, with an increase of 39 Hz (≈ 222 cents) for Tone 1, 27 Hz (≈ 189 cents) for Tone 2, 23 Hz (≈ 166 cents) for Tone 3, and 30 Hz (≈ 167 cents) for Tone 4. Pitch contour for each tone, on the other hand, remained largely unchanged by the SAS. Also visible in Figure 5.5, tone-specific variance is maintained across the different tones (i.e., Tones 3
and 4 exhibit more variance than Tones 1 and 2 in control responses, and this variance is preserved in startle conditions).

Figure 5.5: SS ANOVA results of pitch profiles across 4 Taiwanese Mandarin tones. Other legends follow Figure 5.3.

The prediction that fundamental frequency will increase in response to a SAS is supported by the SS ANOVA results. Elevated pitch contour was observed in startle responses for both English and Taiwanese Mandarin (Figures 5.4 and 5.5). The SS ANOVA results for pitch show a substantial difference (~ 40 Hz) between control and startle responses.
5.4 Discussion

In this chapter, a simple RT task using SAS was conducted to compare Taiwanese Mandarin, a language in which different pitch contour profiles produce contrasts in meaning, with English, in which pitch contour is not used to distinguish word meanings. The purpose of the experiment was to investigate how formant and pitch control are prepared in the speech plan and how these details are revealed in SAS-induced responses. Essentially, the experiment revealed that SAS-induced responses are both rapid and accurate.

As indicated by the results, the introduction of a SAS triggers a rapid release of prepared syllables. This result echoes the findings for English reported in previous chapters of this dissertation. SS ANOVA results for F1 and F2 in both English and Taiwanese Mandarin also confirm that formants remain largely unaffected by a SAS. As formants play a phonemic role in both English and Taiwanese Mandarin, the lack of effect on formant profiles in response to SAS elicitation suggest that a prepared target syllable ought minimally to include phonemic information (i.e., information necessary for lexical contrast). Note that there is a tendency for English F1 to increase in the middle of the response during startle trials. Increased vocal effort is usually accompanied by higher pitch in fundamental frequency and a lower jaw, which in turn elevates F1 values. Therefore, the observed elevation of F1 values in the middle of production may be attributed to the physiological jaw-lowering response to the SAS.

For both English and Taiwanese Mandarin, the pitch height of SAS-
induced responses is significantly elevated throughout the syllable’s duration. As reported in Hain et al. (2000), pitch compensation can occur as early as 100 ~ 150 ms after the onset of auditory feedback. Thus, for SAS-induced responses, if any feedback correction in pitch were to occur, it could start as early as 100 ms after the voice onset. The current results revealed that no evidence of feedback-based correction of pitch level to a pre-specified baseline was observed at any stage in the English production. Thus, it appears that absolute pitch height is not pre-specified in speech plans for English. The elevation can be explained as relating to a physiological startle reflex, with an increase in laryngeal tension presumably acting as a protective maneuver in response to the SAS (Baer, 1979).

In a tonal language such as Taiwanese Mandarin, pitch plays a phonemic role in speech production; on the assumption that contrastive phonemic information should be pre-specified in the speech plan to guarantee response accuracy, one might predict that, for Taiwanese Mandarin, such a plan would include both absolute pitch height and contour profile, while this may not be the case for a non-tonal language such as English. The results from the current study, however, show that both languages respond similarly to the SAS, with initial pitch heights being compromised (elevated) in both languages, while contour profiles are preserved throughout the syllable for all four Taiwanese Mandarin tones. This difference between pitch height and contour profiles may relate to the observation that speakers are capable of producing reliable pitch contours despite a wide range of baseline vocal pitch levels in the population. Pitch control realized within a range of levels, as opposed to an absolute height, has been associated with speakers’ long-
term exposure to the sound inventory of the language, to other speakers of the same language, and to musical training from different ages (Deutsch et al., 2006, 2009). As summarized in Table 5.2, syllable durations across the four Taiwanese Mandarin tones ranged from 210 ms to 313 ms. The effects of afferent (sensory) feedback should thus have been expected to begin as early as $30\% \sim 50\%$ of the way through the syllable’s production. Our results show that for all SAS-induced tonal contours, no attempt to correct the pitch to an absolute or baseline level was observed (Figure 5.5).

In Figure 5.5, the falling contour in SAS-induced responses starts to level out slightly around $50\%$ of the way through the syllable’s duration. One potential explanation could be feedback in action. Previous studies have suggested that pitch regulation is sensitive to auditory feedback (e.g., Houde and Jordan, 2002; Jones and Munhall, 2002; Xu et al., 2004); in continuous vowel phonation, speakers compensate for perturbed auditory feedback at a latency around 210 ms (Jones and Munhall, 2002). Similarly, during the production of bi-tonal sequences, when speakers’ feedback is shifted, the production is compensated at a latency around 164 ms (Xu et al., 2004). In the current study, for Tone 4 responses, the average syllable duration was 205 ms for control responses and 210 ms for startle responses. If the leveling of pitch contour were the result of feedback correction, the compensation occurring around 100 ms ($50\%$ of the way through the syllable’s duration) is earlier than possible feedback compensation as otherwise reported. If the significant pitch drop is the result of feedback correction, the correction is anticipated to occur much later in time. Moreover, if afferent feedback loops back during the production, the pitch contour and level are anticipated to be
in a similar profile and range as those in control responses (i.e., the dashed line in Figure 5.5). Instead of becoming more similar to control responses, the leveling of Tone 4 startle responses is more likely to be accounted for by the creakiness in the second half of the syllable.

Together with the findings reported in Chapter 2, the observations we have made here reveal a good deal of similarity between plans for speech and other SAS-induced (e.g., upper limb) movements. In SAS-induced speech, overall trajectories for a prepared response are generally maintained (for instance, in the context of vowel formants, pitch, and lip movements). Likewise, just as a SAS causes increased EMG muscle activity compared with control responses for upper limb movements (e.g., Carlsen et al., 2012), similar effects are also observed in SAS-induced speech movements (cf. Chapter 2). In the present study, amplified activity in laryngeal muscles elicited by a SAS may also account for increased laryngeal tension and concomitant elevation in pitch level.

To conclude, SAS has not been found to introduce any perturbation to pre-specified formant and pitch contour profiles. Response accuracy thus remains unaffected. However, higher-pitched syllables are observed in response to a SAS; this suggests a limited capacity in the preplanning of speech tasks. While pitch contour is likely to be included in speech plans under feedforward control, absolute pitch height may be implemented at a later stage in production. This suggestion, that pitch contour and pitch height are introduced at different stages of production, calls for further research.
Chapter 6

General discussion

6.1 Summaries

This dissertation employs the design of startling auditory stimulus to trigger prepared speech and investigates different aspects of the planning of spoken syllables. The results from each chapter are summarized as follows.

Chapter 2 demonstrates that the StartReact effect is also observed in prepared syllable production. The implication of this study is that this methodology is also applicable to motor behaviours with heavy cortical involvement. In that chapter, it has been shown that a syllable such as a CV sequence can be prepared in advance and is subject to early release by SAS. In addition to the early release, additional lip compression was observed at the beginning of the voluntary spoken /ba/. This compression was seen in both control and startle responses. Chapter 3 followed up on the observations of both latency and lip compression and further examined whether these effects are speech-specific or generic to other speech-like and non-speech movements. As revealed by the results, lip compression occurred most frequently in Spoken syllables, followed by Mouthed and then Non-speech movements. Compared with Spoken responses, Mouthed and Non-speech responses exhibited less compression displacement and shorter
compression duration. Differentiated frequencies and kinematic profiles of lip compression across Spoken, Mouthed, and Non-speech responses suggest that such lip compression is part of the speech plan for Spoken speech, and possibly more associated with air flow required for Spoken speech. The latencies to the voluntary movement onsets for all three conditions were comparable. However, slightly longer reaction time and more reduced lip compression were observed in the Mouthed speech, compared with the other two. It is conjectured that an inhibitory effect may be induced for Mouthed speech since it is not as well-rehearsed as Spoken speech. Therefore, the potential inhibitory effect may be responsible for this observation.

To compare with the speech-specific lip compression observed in chapter 3, Chapter 4 further tested whether lip compression resulting from movement overshoot is present in all responses and subject to early release by SAS. An additional pre-speech movement (i.e., a mouth closing movement) was added to the beginning of the prepared /ba/. The results showed that all three response types (Spoken, Mouthed, and Non-speech) respond similarly in terms of both latency and lip compression.

Chapter 5 investigates how suprasegmental information are included in speech plans. By comparing English and Taiwanese Mandarin, the results show that pitch levels are subject to be perturbed and elevated by SAS while pitch contour and formants remain unaffected, suggesting that phonemic gestures may be included in speech plans and are more resistant to external perturbation.
6.2 Theoretical implications

The experimental results of this chapter not only address the immediate research questions but also lead to further discussion and implication about previous theoretical work, as I will discuss in the following sub-sections.

6.2.1 Startle paradigm as a window into speech motor planning

It is commonly acknowledged that a speech event not only specifies the intended linguistic function but also coordinates across multiple muscular structures. As defined by Maas et al. (2008, p. 107), a speech motor program is “a set of processes responsible for transforming an abstract linguistic (phonological) code into spatially and temporally coordinated patterns of muscle contractions that produce speech movements.” For example, to successfully produce a CV syllable /ba/, the lips are bound to make a closure followed by a burst into the vowel /a/. Note that the production of /b/ is not as simple as making a bilabial closure. It also involves the distinctive properties that differentiate /b/ from other phonemes, such as /p/, in which aspiration on the burst is required for English. In addition to respiratory and laryngeal mechanisms for maintaining and modulating air flow and phonation, the planning of a syllable as simple as a CV /ba/ would minimally require coordinated lip and tongue kinematics as well as perioral muscular activities anticipating the aerodynamics within.

To understand these coordinated patterns of muscle activity for speech production, McClean and Tasko (2002) examined the relationship between
supralaryngeal kinematics and laryngeal control, and between supralaryngeal and respiratory configurations. Participants were directed to produce a designated phrase (“a bad daba”) with different intensity levels and speaking rates. Their results show a strong and positive correlation between lip kinematics and laryngeal postures for different fundamental frequencies and intensity levels, although individual variations were observed across participants. A more robust correlation was observed between jaw kinematics and laryngeal configuration. McClean and Tasko (2003) further demonstrated that orofacial EMG activity is positively correlated with orofacial kinematics (i.e., distance and speed) across variations in speech rate and intensity. The authors’ results also showed that lip kinematics and EMG are highly correlated with speech rate and intensity, suggesting a tight coupling between orofacial movements and laryngeal and respiratory movements. Similar results are reported in Wohlert and Hammen (2000).

Similarly, other studies also find a tight coupling between lower lip, jaw, and laryngeal movements. For example, Gracco and Løfqvist (1994) demonstrated that lip displacement and glottal configurations are highly correlated in terms of timing during consonant production, as are jaw lowering and glottal closure velocity during vowel production. Based on these observed tight couplings, Gracco and Løfqvist (1994) proposed that speech motor behaviours are organized on a functional basis to successfully produce phonemic segments, and that the motor actions required for a phonemic segment may be stored together as a motor program. While Gracco and Løfqvist (1994) proposed that phonemic segments are the fundamental units of speech production, they also note that these units may be combined into sequences
through the use of blending motions (i.e., trajectories). Speech units, on this view, are sequential compositions of different motor programs moderated by coarticulation mechanics. Syllable construction is encompassed within the programming of units through context-dependent adjustment.

In addition to a high correlation between lip kinematics and laryngeal configurations, research has also found a direct effect of suprasegmental adjustments on lip kinematics. Kelso et al. (1985) showed that different speaking rates and stress patterns may induce different lip kinematics. Larger displacements are observed in normal speech than in fast speech; greater displacements and longer durations are observed under stress. Lip closing and opening movements are also affected by speaking rate to a variable extent (Adams et al., 1993). These observed tightly coupled relationship between orofacial kinematics and laryngeal configurations suggest that speech motor control involves motor commands across muscular and kinematic organizations, including both segmental and suprasegmental components.

If speech plans indeed pre-specify coordination at a muscular level with correspondent kinematics, a SAS-elicited response should be able to reveal the details of such a speech plan. With only limited or no feedback correction, a SAS can reliably elicit prepared speech motor responses. The results from Chapters 2 - 4 showed that a prepared speech event as long as a CV sequence is subject to rapid release by SAS. Uncompromised formants in SAS-induced responses indicate that muscular coordination across the oral, lingual (tongue), and laryngeal control were performed largely unperturbed by SAS. A more general view from this is that an entire CV syllable may be considered as a coherent event. In Chapter 5, prepared syllables were
performed with phonemic contrasts, but phonemic parameters were not independently performed or triggered, making these results inconclusive with regard to planning at a phonemic level. Moreover, Chapter 5 only examined pitch contour as one of the phonemic parameters. An alternative argument for sub-phonemic information to be part of a pre-plan is implicated in studies on inner speech, for example. In inner speech (un-articulated speech), corollary discharge providing a sensory prediction of the motor command would affect listeners’ perception (Scott et al., 2013). Mouthing and imagining speech tasks affect listeners’ perception, based on the sensory prediction of their motor commands for the tasks. When mouthing [afa], for example, speakers are more likely to categorize a sound token from the [ava]-[aba] continuum as [ava]; in other words, they choose the token alternative in which the labiodental feature is shared between their sensory prediction (from [afa]) and their auditory perception (from [ava]). Their results suggest that speech planning may include information detailed to the sub-phonemic level. With our current design, it would be too soon to conclude if phonemic parameters can act as discrete units for speech production, and more specifically be subject to rapid release by SAS. A more reasonable design to examine independently prepared and triggered phonemic parameters would be applying a choice RT task or a Go vs. No Go task using SAS (e.g., Carlsen et al., 2008; Kumru et al., 2006), in which phonemic parameters may be controlled as factors and further be examined if they can be independently elicited.
6.2.2 Forward control

For decades in the speech motor control literature, speech production was considered to be dependent on inter-articulator coordination between planning and feedback regulation. For example, perturbation studies investigating coordination across articulators and muscles have emphasized the interdependence between forward and feedback control (e.g., Abbs and Gracco, 1984; Kelso et al., 1984). In particular, several of these studies observed compensatory responses from the upper lip when external force perturbed the closure movements of the lower lip, suggesting a predictive, open-loop process (Abbs and Gracco, 1984; Kelso et al., 1984; Shaiman and Gracco, 2002). These findings reveal that speech motor behaviour relies on forward control and suggest that some aspects of articulation involve “preplanning” while others are automatic responses. Speech plans may be revealed when appropriate perturbation or stimulus is delivered. Note, however, that the forward control revealed by these studies is conditional; forward control movements only arise when there is perturbation and feedback information available. In other words, speech production cannot normally be observed under forward control alone.

Inner speech and efference copy also reflect the content of planning. For example, as reported by Ylinen et al. (2014), when the auditorily presented vowel is concordant with the rehearsed item, a suppression effect was observed in the auditory cortex; when the auditorily presented vowel is different from the rehearsal, the auditory cortex would show enhanced activity. Their findings revealed that the planning may be as detailed as the level of
phonemes or syllables (not distinguished by Ylinen et al.). Efference copy studies similarly demonstrate that phonemic vowels are part of speech plans (Niziolek et al., 2013).

Along with the above studies, the results from this dissertation also reveal that prepared speech may be performed under largely forward control, i.e., when little or no feedback correction is introduced. Speech and non-speech movements are performed as intended but with shorter latencies. While there appear to be some consequences of the elicitation by SAS, such as augmented EMG, larger acoustic amplitude, elevated pitch level, and greater magnitude of lip displacement, these affected aspects of performance instead suggest that they are not central to the task as specified in the speech plan, and may thus be more susceptible to the SAS perturbation. This is also in line with the Minimal Intervention Principle, which argues that variabilities in task-irrelevant dimensions are allowed and feedback information is used to correct only those that interfere with the intended goal of the task (Todorov and Jordan, 2002, 2003). As such, un-perturbed lip movement trajectories (e.g., lip compression and bilabial burst) and phonemic details are more associated with forward control and are resistant to other corrections or variations.

6.3 Conclusion and future work

Results in this dissertation lead to several broader implications. First, as the StartReact effect is also observed in heavy cortically determined processes like speech, speech motor control may share more in common than
previously believed with body motor control in terms of neurological pathways and execution of motor commands. Second, the results not only uncover a number of aspects of oral motor control that can be pre-specified in a speech plan, but also demonstrate how these aspects may be planned and executed in forward control when only limited feedback correction is available. Third, speech plans may be detailed to the phonemic level whereas suprasegmental control is more susceptible to SAS perturbation.

These and other questions arising from this work can be further tested and confirmed through experimental and simulation approaches. These will call for future research.
Bibliography


Carlsen, A. N., Maslovat, D., Lam, M. Y., Chua, R., and Franks, I. M.


Crevier-Buchman, L., Gendrot, C., Denby, B., Pillot-Loiseau, C., Roussel,


tions between auditory and somatosensory feedback for voice \( F_0 \) control. 


McClean, M. D. and Tasko, S. M. (2002). Association of orofacial with la-


Appendix A

Chapter 3 additional figures

To further examine the phenomenon of lip compression in Mouthed and Non-speech responses, the frequency of lip compression was analyzed. Figures A.1 and A.2 depict the numbers of trials with lip compression occurring in the first and second halves of the block. Overall, more lip compression was observed in Mouthed responses (Figure A.1) than in Non-speech responses (Figure A.2), although some individuals showed the opposite pattern (e.g., subject 8). As shown by the figures, the frequency of trials with lip compression is comparable in the first and second halves. Training effects from the first half of the block did not influence lip compression frequency in the second half.
Figure A.1: Number of trials with lip compression in Mouthed condition. The bars in black represent trials occurring in the first half of the block; the grey bars represent trials occurring in the second half of the block.

Figure A.2: Number of trials with lip compression in Non-speech condition. The bars in black represent trials occurring in the first half of the block; the grey bars represent trials occurring in the second half of the block.