The involvement of the speech production system in prediction during comprehension: An articulatory imaging investigation

Eleanor Drake

A thesis submitted in fulfilment of requirements for the degree of
Doctor of Philosophy

School of Philosophy, Psychology and Language Sciences
University of Edinburgh

2016

Declaration

I hereby declare that this thesis is of my own composition, and that it contains no material previously submitted for the award of any other degree. The work reported in this thesis has been executed by myself, except where due acknowledgement is made in the text.

Eleanor Drake

Abstract and lay summary

This thesis investigates the effects in speech production of prediction during speech comprehension. The topic is raised by recent theoretical models of speech comprehension, which suggest a more integrated role for speech production and comprehension mechanisms than has previously been posited. The thesis is specifically concerned with the suggestion that during speech comprehension upcoming input is simulated with reference to the listener's own speech production system by way of efference copy.

Throughout this thesis the approach taken is to investigate whether representations elicited during comprehension impact speech production. The representations of interest are those generated endogenously by the listener during prediction of upcoming input. We investigate whether predictions are represented at a form level within the listener's speech production system. We first present an overview of the relevant literature. We then present details of a picture word interference study undertaken to confirm that the item set employed elicits typical phonological effects within a conventional paradigm in which the competing representation is perceptually available. The main body of the thesis presents evidence concerning the nature of representations arising during prediction, specifically their effect on speech output. We first present evidence from picture naming vocal response latencies. We then complement and extend this with evidence from articulatory imaging, allowing an examination of pre-acoustic aspects of speech production.

To investigate effects on speech production as a dynamic motor-activity we employ the Delta method, developed to quantify articulatory variability from EPG and ultrasound recordings. We apply this technique to ultrasound data acquired during mid-sagittal imaging of the tongue and extend the approach to allow us to explore the time-course of articulation during the acoustic response latency period. We investigate whether prediction of another's speech evokes articulatorily specified activation within the listener's speech production system

The findings presented in this thesis suggest that representations evoked as predictions during speech comprehension do affect speech motor output. However, we found no evidence to suggest that predictions are represented in an articulatorily specified manner. We discuss this conclusion with reference to models of speech production-perception that implicate efference copies in the generation of predictions during speech comprehension.

Acknowledgements

Foremost, I would like to thank Martin Corley. Without his encouragement I would not have started this program of study. Without his patience, insight, knowledge and support I would never have completed it. I take this opportunity to recognise the huge contribution that Martin made in developing and running the Delta ultrasound processing that converted digital images to a data format that could be analysed for the purposes of this thesis.

I would also like to thank Sonja Schaeffler, who was always ready to discuss ideas and who made sure that I took the opportunity to share those ideas with others. Without her encouragement and enthusiasm I would not have encountered many of the concepts and people central to the work reported here.

I thank: Jim Scobbie for allowing me access to the Queen Margaret University ultrasound recording studio and for including me in the community there; Alan Wrench and Steve Cowan for their assistance with ultrasound recording; Conny Heyde for being a great friend and ceaseless source of information on things ultrasound-related; Robin Lickley for inspiring and encouraging my interest in speech processing. I have shared an office with a number of fellow PhD students over the years. They have all contributed to my thinking, particularly Sam Miller, Ian Finlayson and Paul Brocklehurst.

Thank you to my family, Shay Drake, Sarah Drake, Isabel Sebastian, Gabriel Drake, for their support and love. Most of all, thank you to my son, Daniel Drake-Wallace, for helping me keep everything in perspective.

Contents

Abstract and lay summary ii

Acknowledgements iii

Chapter 1: Introduction 1

1.1. Speech motor activation during comprehension 1

1.2. Prediction during comprehension 3

1.2.1. Pre-activation during comprehension 3

1.2.2. Form preactivation 7

1.2.3. Pre-activation in Comprehension-Production 10

1.3. Speech motor activation in prediction in comprehension? 12

1.3.1. Joint Action and Prediction-by-simulation 13

1.3.2. Forward modelling as an action prediction mechanism 14

1.4. Analysing speech as action 17

1.5. The current study 18

Chapter 2: Item-Set Suitability 22

2.1. Introduction 22

2.2. Method 25

2.2.1. Participants 25

2.2.2. Materials 25

2.2.3. Procedure 26

2.3. Results 26

2.3.1. Error data 27

2.3.2. Response latency data 27

2.4. Discussion 29

Chapter 3: Prediction elicitation 31

3.1. Introduction 31
3.2. Method 36
3.2.1. Participants 36
3.2.2. Stimuli 36
3.2.3. Procedure 37
3.3. Results 38
3.3.1. A general note on data analysis 38
3.3.2. Experiment 2a Findings 39
3.3.3. Experiment 2b Findings 41
3.4. Discussion 43
Chapter 4: Are listener-generated predictions specified at a speech-sound level? 46
4.1. Introduction 46
4.1.1. Abstract to paper 46
4.1.2. Main Introduction to paper 47
4.2. Method 50
4.2.1. Participants 50
4.2.2. Stimuli 50
4.2.3. Procedure 51
4.3. Results 52
4.3.1. Experiment 3a Findings 53
4.3.2. Experiment 3b Findings 54
4.3.3. Experiment 3c Findings 55
4.3.4. Lexical (homophone) Analyses 58
4.3.5. Filler v Experimental Item Analyses 64
4.4. Discussion 64
4.4.1. General implications 66
Chapter 5: Exploring speech as action: An ultrasound imaging approach 68

5.1. Introduction 68

5.2. Motor activity during response making 68

5.3. Speech imaging approaches and findings 71

5.4. Speech imaging to study prediction-as-simulation 73

5.5. The ultrasound imaging approach 75

Chapter 6: Articulatory Imaging Implicates Prediction During Spoken Language Comprehension 77

6.1. Introduction 77

6.1.1. Paper abstract 77

6.1.2. Prediction within the production system 78

6.2. Method 80

6.2.1. Participants 80

6.2.2. Materials 80

6.2.3. Procedure 80

6.2.4. Data treatment and analysis approach 81

6.2.5. Recording Quality 82

6.3. Results 84

6.3.1. Response Latencies 85

6.3.2. Acoustic durations 85

6.3.3. Ultrasound Analysis 85

6.4. Discussion 89

Chapter 7: Are Articulatory Effects of Prediction During Spoken Language Comprehension Speech Sound Specific? 92

7.1. Introduction 92

7.1.1. Somatotopic activation during listening 93

7.1.2. Somatotopic activation as an aspect of prediction during listening 93

7.1.3. Predictive emulation 94

7.1.4. Current Experiment 96

7.2. Method 96
7.2.1. Participants 96
7.2.2. Materials 96
7.2.3. Procedure 97
7.3. Results 97
7.3.1. Response latencies 98
7.3.2. Ultrasound Analysis 98
7.3.3. Lexical analysis 101
7.4. Discussion 104
Chapter 8: Conclusion 106
8.1. Concluding remarks 113
Appendix A: PWI studies summary 115
Appendix B: Context studies 118
Appendix C: Experimental Item Details 119
References 124

List of tables

Table 2.3.1: Experiment 1, response type (correct vs. error) by condition 27

Table 2.3.2: Experiment 1 model coefficients (in ms) for naming latencies 28

Table 2.3.3: Experiment 1 (subset) model coefficients (in ms) for naming latencies, no obs = 776 29

Table 3.3.1: Experiment 2a summary of best-fit model for response latency data (effect stated in ms). 40

Table 3.3.2: Experiment 2b summary of best-fit model for response latency data acquired in Experiment 2b (effects are stated in msecs) 42

Table 4.3.1: Recorded errors in Experiments 3a-c: Figures refer to total errors/numbers of errors in which distractor was produced in error (percentages in brackets). 53

Table 4.3.2: Fast and slow responses excluded from further analyses in Experiments 3: Figures indicate raw number of responses that were below 100ms ('fast') or above 2499ms ('slow'), and (in brackets) percentages of non-error data points in each condition. 54

Table 4.3.3: Experiments 3a-c model coefficients (in logits) for the likelihood of producing an error 56

Table 4.3.4: Experiments 3a-cModel coefficients (in ms) for naming latencies 57

Table 4.3.5: Experiments 3a-c model coefficients (in ms) for naming latencies (sentence contexts only) 62

Table 4.3.6: Experiments 3a-c model coefficients (in ms) for naming latencies (sentence contexts only, data sub-setted by category type; full match and homophone) 63

Table 6.2.1: Discrimination scores calculated for the participants shown in Figure 6.2.1 83

Table 6.3.1: Model coefficients (in Delta units) for differences from Control condition: Details of Context by Onset model. 87

List of figures

Figure 2.3.1: Experiment 1 mean response latencies by condition for correct responses 28

Figure 4.3.1: Experiments 3a-c response latencies to pictures, reported by category (category mismatch = homophone in full-overlap condition, category match = full match in full-overlap condition)..... 58

Figure 6.2.1: Multidimensional scaling of Delta differences between all articulations produced by each of two participants, measured from 0.1s before consonant onset to vowel offset..... 84

Figure 6.3.1: Articulatory movement over time in producing the onsets of picture names which match or do not match predictions from the sentence stem. 88

Figure 7.3.1: Frame-to-frame change in ultrasound tongue image during pre-acoustic articulation.. 100

Figure 7.3.2: Frame-to-frame change in ultrasound tongue image during pre-acoustic articulation in full-overlap condition (homophone analysis) 103

Figure 7.3.3: Frame-to-frame change in ultrasound tongue image during pre-acoustic articulation rime-overlap condition (homophone analysis)..... 103

Figure 7.3.4: Frame-to-frame change in ultrasound tongue image during pre-acoustic articulation onset-overlap condition (homophone analysis)..... 104

Chapter 1: Introduction

Comprehension is, in part, a proactive process. In addition to analysing incoming perceptual information, comprehenders synthesize anticipated input. There is considerable evidence to indicate that anticipated input is specified at semantic and syntactic levels. In this thesis we focus on synthesis at a phonological-phonetic level, and investigate the proposal that prediction at this level invokes the speech-motor system through a process of "prediction-by-simulation" (Pickering & Garrod, 2013). This proposal has generated widespread interest, particularly with respect to the potential nature of representations generated during prediction, and the mechanism by which these might be associated with neural motor activation (e.g., Alario & Hamamé, 2013; de Ruiter & Cummins, 2013; Hartsuiker, 2013; Hickok, 2013; Mylopoulos & Pereplyotchik, 2013; Oppenheim, 2013). We investigate the question of whether speech sound representations activated predictively during comprehension engage the speech production system in a manner comparable to those activated during planning of one's own speech. Specifically, we investigate whether competition between predicted and to-be-produced item is identifiable at a speech-output level.

In this chapter we introduce the theoretical framework under which prediction during comprehension is treated as a speech production phenomenon. We focus on Pickering and Garrod's suggestion that prediction can involve simulation within the listener's speech production system (referred to as "prediction-by-simulation"; Pickering & Garrod, 2013). Although itself a framework rather than an evidence-tested model, Pickering and Garrod's suggestion is inspired by evidence that: (i) hearing speech activates the neural speech motor system; (ii) comprehension involves the prediction of upcoming input; (iii) joint action involves predictive simulation of others' actions. We briefly review evidence supportive of these three claims. We then summarise Pickering and Garrod's unifying argument; that conversation is a form of joint action, that prediction during comprehension can therefore be understood as action prediction, and that action prediction involves simulation within the observer's own neural motor planning system. We review response latency and articulatory imaging evidence to indicate that phonological competition during planning of one's own speech produces observable traces in speech output. We conclude by presenting the approach employed in the current study, in which we investigate whether "planning" during prediction of another's speech evokes similar traces in speech output.

1.1. Speech motor activation during comprehension

Pickering and Garrod argue that comprehension and production must be treated as interconnected processes. They argue that, in line with current thinking concerning the integrated nature of action and perception (e.g., see Hommel, 2004; Gentsch, Weber, Synofzik, Vosgerau, & Schütz-Bosbach, 2016), speaking (i.e., action) and comprehension (i.e., perception) processes are best understood and

modelled as interwoven. The proposal that prediction-by-simulation involves activation within the listener's (neural) speech motor system is underpinned by evidence that listening to speech evokes speech-motor activity (Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Pulvermüller, Huss, Kherif, del Prado Martin, Hauk, & Shtyrov, 2006; Watkins & Paus, 2004; Wilson, Saygin, Sereno, & Iacoboni, 2004; for reviews see Gambi & Pickering, 2013; Scott, McGettigan, & Eisner, 2009). Motor activation has been associated with two types of comprehension-production resonance: referential resonance and communicative resonance (Fischer & Zwaan, 2008; Willems & Hagoort, 2007). Referential resonance describes activation elicited by the linguistic content of the listened-to-material, and involves the representation or simulation of motor acts referred to by the speaker (e.g., hearing "kick" activates leg areas; Hauk, Johnsrude, & Pulvermüller, 2004; Tettamanti et al., 2005). Communicative resonance describes activation related to the phonetic content, or surface form of speech, and involves representation or simulation of the motor activity involved in speech production itself (e.g., hearing /khIk/ activates areas involved in the articulation of that sound stream). It is communicative resonance with which the studies reported in this thesis are concerned.

Communicative resonance was initially observed in response to meaningless speech sounds presented acontextually: Listening to speech sounds has repeatedly been demonstrated to evoke somatotopically specified activation within the primary motor cortex (for a review, see Schomers, Kirilina, Weigand, Bajbouj, & Pulvermüller, 2015). For example, presentation of labially articulated syllables (e.g., /ba/) evokes greater activation within the primary motor cortex sub-region associated with lip movement execution, whereas presentation of lingually articulated syllables (e.g., /ka/) evokes greater activation in the region associated with tongue movement (Pulvermüller et al., 2006). Such effects can be seen not only in the speech motor cortex, but in the articulatory effectors (e.g., tongue and lips) themselves: Electromyographic activity within a given articulator is raised when the participant is auditorily presented with a speech sound involving that articulator (Fadiga et al., 2002; Watkins, Strafella, & Paus, 2003). These effects have been observed in response to acoustic presentation of meaningless syllables in a comprehension-free context, and are therefore neither evidence of predictive activation nor evidence that motor-activation is a pre-requisite for speech processing during comprehension.

More recent research findings suggest that communicative resonance may play a causal role in speech comprehension. Transcranial magnetic stimulation (TMS) of somatotopically relevant areas of the speech motor and premotor cortex has been demonstrated to affect speech comprehension in an articulatorily specified manner (Schomers et al., 2015): People are quicker to match heard words to pictures when the place of articulation of the word onset is consistent with the neural speech-motor region activated via TMS than when the area activated relates to a competing articulator. For example, people are quicker to select the relevant picture for labial-onset words such as "pool" when TMS pulses are applied to the primary motor cortex area associated with lip movement than when pulses are applied to the area associated with tongue movement. The reverse is true for lingual onset words such as "tool". The fact that comprehension task performance (word-picture matching) is affected by externally-induced somatotopically-specific simulation of the primary motor cortex

implies that activation within the speech motor system impacts semantic processing. That is to say, neural speech motor activation contributes to comprehension. This evidence supports an action-perception model of language processing (e.g., Fadiga & Pulvermüller, 2010). Whilst this is consistent with Pickering and Garrod's argument that production and comprehension are interwoven processes, the study described above did not address whether the putative relationship between speech motor activation and comprehension involves a predictive element.

There remains considerable controversy as to whether, how, and why speech production processes might be recruited speech during comprehension (see Galantucci, Fowler, & Turvey, 2006 and Scott, 2009 for reviews). As outlined above, Pickering and Garrod (2007; 2013) propose that communicative resonance arises not only in response to externally generated perceptual input, but that it is also evoked during preactivation of speech motor representations during comprehension: They suggest that listener predictions of upcoming input are implemented in the (neural) speech production system, and that these predictions ultimately result in activations within the speech-motor system.

1.2. Prediction during comprehension

Phenomenological evidence suggests that, when language is used in an everyday setting (i.e. dialogue), interlocutors make use of an ability to predict each other's material: We finish one another's sentences (Clark & Wilkes-Gibbs, 1986), and inter-speaker turn intervals can be shorter than intra-speaker intervals (de Ruiter, Mitterer, & Enfield, 2006; Stivers et al., 2009; Heldner & Edlund, 2010; Sacks, Schegloff, & Jefferson, 1974; Wilson & Wilson, 2005). Strongly predictive accounts of language comprehension hold that multi-level constraints within the input lead listeners to pre-activate upcoming items or features in an online fashion (see Kutas, DeLong, & Smith, 2011, for details and discussion; see also Martin, 2016, for discussion of cues as constraints). This pre-activation "most likely does not reach the level of consciousness", and therefore tends to be investigated within experimental paradigms that allow the use of online measures (Kutas et al., 2011).

The nature of predictively elicited representations can be investigated by the manipulating the degree of overlap between a predicted item and a presented item at the level of interest (e.g., conceptual, semantic, syntactic, phonological). This approach has been employed in visual world, event related potential (ERP) and picture naming studies, and has provided significant insight into the role and nature of top-down processing during comprehension. We briefly summarise these approaches and relevant findings below.

1.2.1. Pre-activation during comprehension

Prediction during comprehension cannot be directly measured. It is instead inferred on the basis of behavioural or neurophysiological effects. If an effect is to be interpreted as evidence of predictive activation it should be observable prior to, or in the absence of, perceptual presentation of the

predicted item (e.g., Pickering & Garrod, 2013). Eye-tracking within a visual world paradigm allows the capture of relevant behavioural data. Participant eye movements are recorded online, allowing fixations and saccades to be time-locked against linguistic information (see Kamide, 2008). It is understood that linguistic material guides the listener's attention allocation, which in turn determines gaze direction (e.g., Huettig, Rommers, & Meyer, 2011). Within the visual world paradigm, linguistic material is acoustically presented whilst the participant views a "visual world" on screen. The visual world contains potential referents and/or distractors alongside irrelevant fillers. Typically, constraints provided by the linguistic and visual information combined lead some images to be preferentially attended (i.e., attract a higher number of fixations than others). When an image is preferentially attended prior to, or in the absence of, its name being presented, it is understood that participants have anticipated that the item will be referred to within upcoming spoken input (see Huettig, 2015).

Electroencephalographic (EEG) recording provides another means to collect data pertinent to the question of whether participants predict upcoming input during comprehension. EEG captures temporally high resolution data concerning electrical activity within the brain. Electrodes placed on the participant's scalp allow changes in the electrical activity of the brain to be recorded over time. Much of the activity recorded does not relate specifically to the event of interest. However, the impact of this unrelated activity can be minimised by averaging over many trials. This allows researchers to isolate components related to the event of interest. Signature peaks and troughs in electrical activity (referred to as "event related potentials"; ERPs) provide a means to infer online neural processing.

Although ERPs have a somewhat unknown temporal delay relative to neural activity, they allow for estimation of the time point during stimulus presentation at which specific effects developed. As with eye-tracking, the potential this offers to determine whether effects arose before perceptual presentation of a specific stimulus has been crucial in determining whether an effect can be said to arise of prediction (see Osterhout, Kim, & Kuperberg, 2007, for a discussion). Whereas eye-tracking data concerning prediction has commonly been obtained in paradigms where the linguistic input is acoustic, ERP data has tended to be collected in paradigms where the linguistic input is orthographic. Orthographic material in such studies is typically presented via rapid serial visual presentation (RSVP), in which the words of a sentence fragment are presented one-by-one, in order, in written form on a computer screen. This allows maximal experimental control over when participants process specific lexical items but limits the extent to which findings might inform about the processing of spoken, as opposed to written, language.

Eye-tracking data suggest that listeners make predictions that incorporate semantic information. A key finding in this respect is that verb meaning induces preferential looking to semantically appropriate objects (Altmann & Kamide, 1999; Mani & Huettig, 2012). For example, upon hearing a sentence featuring the verb "eats" listeners preferentially fixate edible items (e.g., CAKE) over non-edible items (e.g., BALL). This is not the case when the verb does not impose a selectional constraint

within the visual context. The fact that preferential fixation to edible items occurs prior to acoustic presentation of the edible item suggests that the (semantic) selectional constraint imposed by the verb causes participants to shift visual attention to items that satisfy the constraint. That is, participants predict upcoming input at a semantic level. Predictive shifts in visual attention have also been observed where the verb constrains upcoming information through their argument structure: Dative verbs elicit higher numbers of fixations to potential recipient images than do monotransitive verbs (Boland, 2005).

ERP data indicate that semantic predictions are also elicited during reading (Federmeier & Kutas, 1999; see also Federmeier & Kutas, 2001 for a version in which the critical item is presented in picture rather than orthographic form). This evidence rests on the understanding that N400 response indexes semantic unification, i.e., fit between elements of a sentence or utterance, with N400 amplitude being inversely related to the degree of semantic fit between the critical item and its context (see Osterhout et al., 2007). When context is provided by a high-cloze sentence fragment, items that provide equally unlikely completions may differ in the degree to which they overlap with the high-cloze target at a semantic feature level. For example, the sentence fragment "They wanted to make the hotel look more like a tropical resort so along the driveway they planted" is a high-cloze sentence fragment that typically elicits the completion "palms". The words "pines" and "tulips" are both plausible yet unlikely completions to the sentence. However, "pines" overlaps with "palms" at a semantic feature level to a greater extent than does "tulips" (for example, both pines and palms have trunks, whereas tulips do not). The amplitude of the N400 response to the presented cloze item (i.e., "palms", "pines", "tulips") was found to be inversely related to the degree of semantic overlap between the presented item and the expected high-cloze item. This suggests that the typical completion ("palms"), although not presented in the stimulus, is activated[1] to at least a semantic feature level. This is therefore interpreted as evidence that readers pre-activate semantic features of high-cloze items.

The N400 data described above are not consistent with an integration-only account (i.e., an account that attributes effects to difficulty integrating the presented cloze item with the linguistic context in which it is presented). This is because items that are equally unlikely in a given context will be equally difficult to integrate in that context. Therefore, if the N400 effect reflected integration costs alone, items that are equally unlikely would elicit N400 components of equal amplitude (for further discussion see Federmeier, 2007). In the studies described above this was found not to be the case, although unlikely items presented in the study above did not differ in their cloze-probability. Instead,

[1] Throughout this thesis, the term "activation" is used when referring to neuroimaging data to signal "significant difference from resting state on the measure employed in that study", rather than an actual increase in activity: our perspective is agnostic with respect to whether such activity reflects activation or inhibition at a neurobiochemical level, and we, personally, are, at present, unqualified to comment further on this topic: nor is it of immediate pertinence to the current study. When the term "activation" is used in a more general sense it is as the more commonly employed equivalent to "invocation".

N400 amplitude varied as a function of the degree to which an unlikely item shared semantic features with the (absent) high-cloze item. The fact that N400 amplitude was conditioned on aspects of this relationship to the high-cloze item suggests that the high-cloze item was subject to representation at a semantic feature level, despite being absent from the stimulus. Given that the high-cloze item was not presented, evidence to suggest that its semantic features were active is interpreted as evidence that the item was subject to representation at this level through a top-down process of prediction.

An ERP approach has also been employed to investigate predictive activation of syntactic features during comprehension. Modulations in ERP amplitude can be associated with the presentation of gender- incongruous adjectives or determiners. This has been interpreted as evidence that listeners and readers preactivate syntactic features of upcoming lexical items (Wicha, Bates, Moreno, & Kutas, 2003; Wicha, Moreno, & Kutas, 2003; Wicha, Moreno, & Kutas, 2004; Otten & van Berkum, 2008; Otten, Nieuwland, & Van Berkum, 2007; Foucart, Ruiz-Tada, & Costa, 2016). For example, Van Berkum, Brown, Zwitserlood, Kooijman, and Hagoort (2005) presented participants with "predictive two-sentence mini-stories", read aloud in Dutch (mean cloze probability = .86, e.g., "The burglar had no trouble locating the secret family safe. Of course, it was situated behind a big but unobtrusive painting/bookcase"). Dutch nouns have syntactic gender (e.g., painting = schilderij$_{neu}$ / bookcase = boekenkast$_{com}$), requiring inflection on the prenominal adjective (e.g., big = groot$_{neu}$/grote$_{com}$). Although the form of the adjective is determined by the noun, listeners encounter the adjective first. If listeners anticipate the syntactic gender of the upcoming noun, this will inform their expectations concerning the inflection on the adjective. The adjective could therefore violate gender agreement with a predicted noun even when there is no violation with respect to the noun that is actually subsequently heard (e.g., grote$_{com}$ violates gender agreement for predicted schilderij$_{neu}$ although not for the actually heard boekenkast$_{com}$). Van Berkum and colleagues found this to be the case: Deflections in ERP amplitude time-locked to the point at which the adjectival inflection was presented were larger when the inflection was not gender congruent with the high-cloze (i.e., predicted) noun than when it was. This is interpreted as evidence that at the point of adjective presentation participants have already developed expectancies conditioned on syntactic features of the noun; that is, syntactic features of the noun are preactivated through prediction.

The nature of ERP modulations associated with syntactic gender mismatch with likely upcoming words varies considerably between studies. The "canonical" N400 component associated with a meaning-related violation is observed in data from electrodes located in centro-parietal regions. The negative-going deflection starts between 200 and 300 milliseconds after orthographic or acoustic presentation of the word, and peaks around 400 milliseconds post-presentation. Modulations associated with predicted gender mismatch have been observed more frontally (e.g., van Berkum et al., 2005) and later (e.g., Wicha et al., 2003). Such variations can lead to questions concerning the validity of interpreting these findings as evidence of the prediction of syntactically-specified lexical items. Discussion of matters concerning the localization of the N400 component source is beyond the scope of this thesis (see Lau, Phillips, & Poeppel, 2008 for a review and meta-analysis). However,

we note that the claim that syntactically-specified representations are preactivated through prediction during comprehension is consistent with the suggestion that prediction during comprehension involves the retrieval of specific lexical items. This suggestion is supported by evidence from whole-brain frequency-specific EEG oscillatory data, in which it has been observed that high-cloze contexts are associated with theta band power increases not see in lower cloze contexts (Wang et al., 2012).

The suggestion that specific lexical items are pre-activated during comprehension (e.g., Lau et al., 2008; Kutas et al., 2011; Wang et al., 2012) raises the possibility that predicted information might be subject to form-level representation. It is this possibility which underlies Pickering and Garrod's proposal that neural speech motor activation observed during listening may in part reflect predictive processing, specifically prediction-via-simulation. Below we review evidence of phonological form prediction during comprehension.

1.2.2. Form preactivation

The ERP approach described above with respect to the investigation of semantic and syntactic preactivation has also been used to investigate whether phonological features of (anticipated) upcoming words are preactivated**.** In a study exploiting the rule that, in English, the phonological form of the indefinite article varies according to whether the word it precedes begins in a vowel or a consonant, DeLong and colleagues (2005) found N400 amplitude effects associated with phonological form appropriacy. Participants read high-cloze sentences word-by-word (as presented by RSVP; e.g., "The day was breezy so the boy went outside to fly a/an…"). High-cloze target nouns began with either a consonant (e.g., "kite") or a vowel (e.g., "airplane"), requiring the preceding indefinite article to take either the form "a" or "an" respectively. When the indefinite article was inappropriate to the phonological form of the high-cloze probability noun, N400 amplitude was higher than when the article form was appropriate. N400 amplitude in response to non-match indefinite articles varied systematically in accordance with the cloze-probability of the anticipated noun. When a sentence context less strongly biased toward a specific cloze item, the N400 amplitude in response to an incongruous indefinite article was reduced. As the two forms of the indefinite article differ only at a phonological form level (i.e. do not differ at a syntactic or semantic level), it was concluded that the N400 effect on the article arose because the surface form of the anticipated cloze item was preactivated via prediction, thus influencing the anticipated form of the preceding indefinite article (DeLong et al., 2005). The surface form prediction effect first reported by DeLong and colleagues has been replicated and found to be present for L1 speakers but not L2 users of a language (Martin, Thierry, Kuipers, Boutonnet, Foucart, Costa, 2013).

N400 amplitude in response to the cloze noun itself, as opposed to the preceding article, appears to be associated with the presence of phonological overlap between the item presented and the high cloze probability predicted item (Ito, Corley, Pickering, Martin, & Nieuwland, 2016): When very high-cloze sentence contexts (mean cloze probability = .93) were presented via serial visual presentation at

a slow rate (500ms on, 200ms off per word), the N400 response to unexpected cloze items was attenuated when the word overlapped at coda position with very high-cloze probability nouns. For example, given the sentence context "The juice isn't cold enough, so Alice is adding some … now", N400 was attenuated in response to "dice", which overlaps with the predicted word "ice", as opposed to "wine", which does not overlap with the high-cloze target in onset or coda position. This effect was not apparent at a faster presentation rate (300ms on, 200ms off) nor in medium cloze (M = .65) contexts.

As yet, ERP modulations associated with predicted surface form have been found only when the context is presented word-by-word in written form. Current ERP findings therefore cannot distinguish between whether the surface form effect is phonological or orthographic in nature (i.e., whether participants pre-activate sound or letter form features of the anticipated word, or both). Effects appear to emerge earlier than a classical N400 (i.e., at around 250ms); this was noted particularly by Martin and colleagues, who use the term "N400a" (Martin et al., 2013). N400(a) amplitude has previously been shown to be modulated by the degree to which a presented orthographic word form matches an anticipated word form, suggesting that orthographic representations are pre-activated: Orthographic neighbours of high-cloze items elicit smaller N400 amplitudes than do words that do not overlap with the predicted word at an orthographic level (Laszlo & Federmeier, 2009; see also Kim & Lai, 2012). It has been argued that early effects are consistent with contextual support for a high-cloze word leading to top down activation of orthographic form features of the word (Kim & Lai, 2012), as the N250 has been found to be sensitive to the degree of orthographic overlap between a prime and target (e.g., Grainger & Holcomb, 2009; Holcomb & Grainger, 2006; Kiyonaga, Midgley, Holcomb, & Grainger, 2007). However, it should be noted that the greater support for an orthographic interpretation is largely a function of the fact that historically researchers tended to use orthographic presentation, for both pragmatic and theoretical reasons. Where phonological overlap has been investigated within an acoustic priming ERP paradigm, the N400 effect is similar to that found in orthographic priming studies, and does not appear to be dependent on conscious processes (e.g., Praamstra & Stegeman, 1993; Radeau, Besson, Fontenau, & Caastro, 1998; for an alternative view see Perrin & Garcia-Larrea, 2003).

Whether form representations are elicited during comprehension can also be investigated within the visual world paradigm (introduced above). This is achieved by including in the "visual world" an item that shares form features with an item that is predicted by the linguistic input. As detailed above, eye gaze is understood to reflect attentional focus, and therefore to provide inform as to the content of representations active at the time. It has previously been demonstrated that when participants hear a word they look to phonologically related items more often than to those that are unrelated (Allopenna, Magnuson, & Tanenhaus, 1998). If listener predictions are similarly specified at the form level, items that are phonologically related to the predicted word should be preferentially attended over non-related items. In the case of speech sound predictions, this would involve a higher number of fixations to items that overlap at a phonemic level with the high-cloze item than to those that do not.

Evidence is emerging to suggest that this may be the case (Ito, Corley, & Pickering, submitted; see also Dahan & Tanenhaus, 2004, for potentially relevant evidence from constraining sentences and heard words, and Chabal & Marian, 2015 for a related visual search paradigm).

In the study reported by Ito and colleagues (submitted), highly-predictive sentences were presented acoustically in English to participants (e.g., "The tourists expected rain when the sun went behind the cloud"). Sentences were presented in 4 conditions; in each condition the on-screen image contained 3 unrelated distractor items and 1 critical item. The four conditions were formed by manipulating the properties of the critical item such that its name either; (i) was the predicted word (e.g., cloud), (ii) partially shared the phonological form of the predicted word (e.g., clown), (iii) was unrelated to the predicted word, (iv) in Japanese partially shared phonological form with the predicted item. Critical items were fixated most often in the condition where the high-cloze item was present on screen, but were fixated more often in the partial phonological overlap condition (English) than in the unrelated and Japanese conditions. This effect was observable approximately 2 words prior to acoustic presentation of the highly predictable word, and was interpreted as suggesting that the phonological form of upcoming words is activated by that point in L1 speakers.

The finding described above may be experiment-specific. As the authors themselves highlight, the fact that participants were initially required to name the relevant pictures in isolation may have "boosted word form activation". Pictures, even when to-be-ignored, can elicit phonological word form activation (e.g., Meyer & Damian, 2007). In the study reported by Ito and colleagues, pictures were displayed on-screen from 1000ms prior to onset of the critical word. It is possible that picture presentation elicited automatic word form activation related to the images displayed. Such activation could, theoretically, interact with lexical level preactivation of anticipated items to guide attention to the phonologically-related image, as opposed to attention being guided by independent word form activation of predicted items. This interpretation would require to be ruled out in further investigations. This is particularly the case because the time interval between picture presentation and presentation of a highly predictable word was consistent: In some cases the picture represented the highly predictable word. The consistent time interval may therefore have led participants to develop an association between the two events, in which picture presentation acted as a trigger to focus on specific predictions.

A replication of the finding reported by Ito et al. would be desirable, as the form effects were small and short-lived (in the view of Ito and colleagues; see manuscript p. 28), and statistical modelling was generally anti-conservative. However, the finding itself is highly relevant: Unlike in the ERP studies of form preactivation described above, the evidence indicates that phonological form can be preactivated to an extent that items that overlap only partially at this level are preferenced over those that do not overlap. This finding is critical, in that it provides much stronger evidence of predictive activation of phoneme-level representations than can evidence that electrophysiological activity differs when an item fulfils phonological expectations than when it does not. Further, although Ito

and colleagues suggest that their study cannot distinguish between orthographic and phonological preactivation, the fact that experimental stimuli were delivered acoustically and pictorially favours a phonological account to a much greater degree than studies in which stimuli were presented orthographically (e.g., DeLong et al., 2005; Ito et al., 2016; Martin et al., 2013).

In summary, both eye-tracking and neural signal recording approaches have provided evidence that readers and listeners pre-activate features of anticipated input. Semantic and syntactic features appear to be preactivated in both written and spoken contexts. There is some evidence to suggest that phonological features are also preactivated, under certain circumstances. When stimulus material is presented in written form, readers appear to anticipate surface form (e.g., readers anticipate the indefinite article form "a" before a consonant onset and "an" before a vowel onset). This has been interpreted as indicating that comprehenders generate phonological representations of anticipated items, although the extent to which this interpretation can be generalised to listening as opposed to reading during RSVP remains open to question: ERP evidence from written word studies does not allow a distinction to be made between phonological and orthographic preactivation. However, the finding that listeners are more inclined to look to items when they are phonologically related to predicted input suggests that listeners do, at times, preactivate anticipated input at a speech-sound level. This allows the possibility that communicative resonance might, at times, reflect processes involved in preactivation at a speech-sound level during linguistic processing. It does not, however, directly relate to (neural) speech-motor activation during comprehension.

1.2.3. Pre-activation in Comprehension-Production

In addition to being used in studies of online language comprehension, cloze-probability and contextual constraint manipulation have been used to study language production (and, often incidentally, relationships between comprehension and production). Production studies have, to a greater extent than comprehension studies, tended to be situated within an integration framework rather than a prediction framework. That is, they have considered what features ease integration rather than what or how features might be represented during prediction of upcoming material. However, the fact that the sentence-cloze approach has been used across comprehension and mixed modality (comprehension-production) studies allows findings to be interpreted within a prediction framework.

Production activity in relevant comprehension-production designs is typically elicited by asking participants to complete sentences (e.g., Bock & Miller, 1991) or to name pictures or words presented within a sentence context (Bentrovato, Devescovi, D'Amico, & Bates, 1999; Bentrovato, Devescovi, D'Amico, Wicha, & Bates, 2003; Forster, 1981; Gollan, Slattery, Goldenberg, van Assche, Duyck, & Rayner, 2011; Griffin & Bock, 1998; Jacobsen, 1999; Manenti, Repetto, Bentrovato, Marcone, Bates, & Cappa, 2004; Piai, Roelofs, & Maris, 2014; Roe, Jahn-Samilo, Juarez, Mickel, Royer, & Bates, 2000; Stanovich & West, 1979; Wicha, Orozco-Figueroa, Reyes, Hernandez, Gavaldón de Barreto, &

Bates, 2005; see Appendix B for study-by-study summaries). For example, Roe and colleagues investigate sentential context effects via auditory presentation of a sentence-stem followed by presentation of an image-to-be-named (Roe, Jahn-Samilo, Juarez, Mickel, Royer, & Bates, 2000). Participants were required to name pictures in three sentential contexts; semantically congruent, semantically incongruent, and semantically neutral (e.g., The dog was chasing the CAT/ Peter eats his soup with a CAT/ Now please say CAT). Compared to a neutral context, congruous sentential context facilitated picture naming (as measured by voice-key response latency). This effect remained steady across the experiment. Incongruous sentential context sometimes slowed picture naming response times, but this effect was only observable in by-participant analyses of trials presented early in the experiment. This led the authors to suggest that sentential interference is, "[…] vulnerable to a build-up of strategies and expectations across the course of the experiment.", in contrast to the facilitative component of "sentential priming" (i.e. sentence context), which they suggest is invulnerable to repetition effects due to being a, "relatively automatic process" (p.763). The suggestion that naming facilitation reflects automatic processing elicited via sentence comprehension was subsequently supported by the findings of a lexical decision task (Aydelott & Bates, 2004; see chapter 3 for further details).

The way in which semantic and syntactic expectancies generated by sentential context interact to influence production has primarily been studied in languages that have grammatical gender (e.g., Italian, Spanish). When pictures are presented for naming following acoustic presentation of a sentence context, latencies in partially congruent contexts (i.e. those that are congruent at either a semantic or a syntactic level) are no slower than those obtained in neutral sentence contexts. Latencies in a fully incongruent context (i.e., when both the noun's gender and its meaning are incongruent given the sentential context) are slower than those in a neutral context (Bentrovato et al., 1999; see also Wicha et al., 2005; for word-naming versions, in which partial congruence is facilitatory, see Bentrovato et al., 2003; Manenti et al., 2004).

Whether sentential context generates expectancies that influence production at a phonological form level has, to the author's knowledge, not been studied directly prior to the work presented in this thesis. However, indirect evidence pertinent to the question can be found in studies that investigate whether sentential context modulates lexical frequency effects. The term "lexical frequency effect" here refers to the finding that, during simple picture naming, response latencies for high-frequency words are shorter than those for low-frequency words. In both comprehension and production, lexical frequency effects are generally assumed to arise primarily during lexical access, specifically at the point of form retrieval (for production see Jescheniak & Levelt, 1994; Levelt, Roelofs, & Meyer, 1999; for comprehension see; Forster & Chambers, 1973; Murray & Forster, 2004; Rayner, 1998; see Kitteridge, Dell, Verkuilen, & Schwartz, 2008 and Knobel, Finkbeiner, & Caramazza, 2008 for alternative viewpoints based on case studies of patients with aphasia; see Gollan et al., 2011 for a review).

It appears that the lexical frequency effect in picture naming is modulated by sentential context; frequency effects are observable in low sentential constraint contexts (i.e. when no lexical item has high cloze probability given the sentence context), but abolished or reduced in high constraint contexts (i.e., when the sentence context has an associated high-cloze lexical item; Griffin & Bock, 1998; Gollan et al., 2011; Piai et al., 2014). Under a prediction framework account, the interaction is observed because high constraint contexts guide form retrieval prior to presentation of the picture-to-be-named. If frequency effects arise during form retrieval, and high-cloze contexts elicit form retrieval prior to picture presentation, frequency effects will then not be observable at picture presentation as they will have occurred at a point prior to picture presentation. This interpretation is supported by the finding that neural activity associated with lexical access is raised prior to picture presentation in high-cloze contexts compared to low-cloze contexts (Piai et al., 2014, p. 154).

In summary, effects of sentential constraint on picture naming suggest that auditorily and orthographically presented high-cloze contexts orthographically affect speech production processes. To some degree, at least, the effects seem to arise of automatic processing. High-cloze contexts appear to impact processing during speech production at semantic and syntactic levels. Cloze-probability and picture name frequency interact, suggesting that these variables share a processing locus. If, as has been suggested, frequency effects arise during word form retrieval, this may be indirect evidence that sentential context affects word form activation within the speech production system through predictive preactivation.

1.3. Speech motor activation in prediction in comprehension?

As it has become largely accepted that sentential context can cause listeners and readers to preactivate aspects of upcoming material, researchers have begun to focus on the mechanisms by which such prediction is achieved. One suggestion which has gained considerable traction, and which we investigate in this thesis, is the suggestion that; "predictions are generated by the language production system" through motor simulation (Pickering & Garrod, 2007, p.105; see also Schiller, Horemans, Ganushchak, & Koester, 2009). The suggestion that motor simulation is one route to prediction during comprehension is situated within a framework that treats linguistic interchange as a special form of joint action. Pickering and Garrod, in line with others (e.g., Clark, 1996), suggest that conversation can be understood as a case of joint action in which the action is speech. Drawing an analogy from the joint action prediction literature, Pickering and Garrod suggest that when people listen during conversation their predictions of upcoming input are action predictions, involving forward modelling within the listener's own motor system. Below we briefly introduce the concept of joint action, with specific reference to the proposed mechanism of prediction; forward modelling within the observer's neural motor system.

1.3.1. Joint Action and Prediction-by-simulation

Joint action has been defined as "any form of social interaction whereby two or more individuals coordinate their actions in space and time to bring about a change in the environment" (Sebanz, Bekkering, & Knoblich, 2006, p.70). Joint action requires prediction of a co-actor's actions, in that it requires that individual bodies (and minds) are co-ordinated (e.g., Allport, 1924, cited in Sebanz, Bekkering & Knoblich, 2006). It has been proposed that such action prediction may be achieved through a process of action emulation (that is to say, motor simulation; Wilson & Knoblich, 2005), and that this engages processes typically recruited during the planning and execution of a person's own actions (Knoblich, Butterfill, & Sebanz, 2011).

Task sharing paradigms have provided a means to investigate action planning within a joint action environment (e.g., Sebanz, Knoblich, Prinz, & Wascher, 2006; Tsai, Kuo, Hung, & Tzeng, 2008; Tsai, Kuo, Jing, Hung, & Tzeng, 2006; Knoblich et al., 2011; see Obhi & Sebanz, 2011; Wenke, Atmaca, Holländer, Liepelt, Baess, & Prinz, 2011 for reviews). The pattern of performance observed indicates that two participants performing individual tasks within a joint environment act in a manner similar to that observed when one individual performs both tasks (rather than when two participants perform their own individual tasks in isolation; e.g., Sebanz, Knoblich, & Prinz, 2003; Welsh, 2009). This supports the notion that we plan (predict) other's actions in a manner functionally similar to the way that we plan our own actions (Welsh, 2009), an argument that underlies Pickering and Garrod's proposal that the listener "plans" the speaker's upcoming input using their own speech production system.

The effects observed in task sharing paradigms may be explained by either an actor co-representation account (e.g., Dolk, Hommel, Colzato, Schütz-Bosbach, Prinz, & Liepelt, 2014; Philipp & Prinz, 2010; Vlainic, Liepelt, Colzato, Prinz, & Hommel, 2010; Wenke et al., 2011) or a task co-representation account (e.g., Atmaca, Sebanz, & Knoblich, 2011; Sebanz, Knoblich, & Prinz, 2005). Under the actor co-representation account, representations of the other partner's activity specify only when the other partner should act. By analogy, if the joint action task is conversation, then the listener's predictive representation of the task will detail only when they should speak. Under the task co-representation account, partners form representations of both when the other partner should perform and *what* they will perform. Again by analogy, the listener would predict the content of what the speaker would say (as Pickering and Garrod propose). ERP findings provide some evidence to support a task co-representation account interpretation of joint action: When participants performed a key push task in which either they, their partner, or neither partner should respond, with intended responder identity being indicated by a colour cue, P300 amplitudes were greater in "partner response" than "neither respond" trials (Tsai et al., 2006). The P300 component is widely taken to index inhibitory processes. This finding was therefore interpreted as evidence that participants formed representations within their own neural motor systems of their partner's (unobserved) actions, which

then had to be inhibited in order to avoid their own responding (see Baus, Sebanz, de la Fuente, Branzi, Martin, & Costa, 2014, p. 396). It should be noted, however, that this approach cannot be adapted to investigate whether listeners represent at a speech-motor level what speakers will say. This is because the P300 indexes inhibition at a general level; its amplitude could therefore be used to infer whether listeners predicted that the speaker would speak but not what sounds they would speak (i.e., the content of what they would say).

There is currently, to the author's knowledge, no direct evidence concerning whether participants in spoken dual-task paradigms generate form-level representations of each other's turn. However, as with sentential context effects (see section 1.2.3), indirect evidence is provided by an investigation of lexical frequency effects: Baus and colleagues (2014) employ an electrophysiological approach to investigate lexicalization processes during a task sharing paradigm (as indexed by P200 modulations associated with lexical frequency/ activation; see, for example, Strijkers, Holcomb, & Costa, 2011), in order to address "whether lexical processes in speech production are involved during the anticipation of a task-partner's utterance" (p. 396). Participants performed a go no-go picture naming task in two conditions; (i) alone; (ii) alongside a partner. In both conditions, pictures were presented in 3 colours (blue, black, red) and participants were to name only pictures presented in red (self-go trials). In the joint condition the partner (a confederate) named only pictures presented in blue, and neither participant named pictures in black. In this way, both blue and black pictures represented no-go trials for the experimental participant in both lone and joint conditions. In blue picture trials, however, there was potential for the participant to form a representation of their partner's action (other-go), whereas in black picture trials neither participant would be expected to act. Participants' neural activity in the two no-go conditions was compared. An interaction between word frequency and trial-type (other-go v. none-go) indicated that the magnitude of the word frequency effect was influenced by trial-type. The authors interpret this interaction as indicating that, "when participants did not have to name a picture, traces of lexical access were present as long as another person had the intention to name it" (p.403).

1.3.2. Forward modelling as an action prediction mechanism

As noted above, Pickering and Garrod propose that prediction during comprehension can be achieved via forward modelling within the production system. This suggestion is made in the context of models and evidence from the wider motor control literature, particularly in relation to joint action (see Wolpert, Doya, & Kawato, 2003; Brown & Brüne, 2012, for a review). Under this suggestion, predictive emulation of others' actions during joint action is thought to engage forward modelling processes typically associated with planning and executions of one's own actions (see Sebanz & Knoblich, 2009; Wilson & Knoblich, 2005, for reviews; see also Knoblich & Flach, 2001; Avenanti, Candidi, & Urgesi, 2013).

Many models and theories of action planning and/or execution incorporate the generation of "forward models" (i.e., predictions of the state of the motor apparatus or sensory input at future point in time; see Grush, 1997; 2004; Hickok, 2012; Miall & Wolpert, 1996; HMOSAIC; Haruno, Wolpert, & Kawato, 2003). The internal forward model is considered to be internal to the central nervous system. It represents movement that will be executed within the peripheral nervous system, and "captures the forward or causal relationship between actions and their consequences" (Wolpert & Flanagan, 2001, p.R729). The mechanism by which forward models are thought to be generated is efference copy (also referred to as "corollary discharge"; for classical evidence from the visual domain see Sperry, 1950; von Holst & Mittelstadt, 1950). The efference copy is a copy of a motor command, thought to be emitted along the route from motor-command generation to motor execution. In this strict sense, the term "efference copy" refers to a specific neurobiological event that occurs as an outgoing (efferent) signal from the motor cortex to the peripheral nervous system is copied to the sensory cortex. This input activates within the central nervous system a representation of the sensory input associated with execution of the copied motor command. The sensory representation is referred to as the forward model. The forward model can be compared to input from the perceptual system in order to determine whether the action command goal was achieved in terms of its sensory consequences.

Indirect evidence from behavioural and neuroimaging studies supports the suggestion that forward models of the actions of others are generated within the neural regions associated with motor action (e.g., Aglioti, Cesari, Romani, & Urgesi, 2008; Kourtis, Sebanz, & Knoblich, 2013; Ramnani & Miall, 2004; van Schie, Mars, Coles, & Bekkering, 2004). Activation of motor-associated neural regions is observed prior to visual presentation of (expected) predictable movements (Kilner, Vargas, Duval, Blakemore, & Sirigu, 2004; see also Haueisen & Knösche, 2001). Increases in premotor activation can be observed when participants can predict the "future course of partly invisible action sequences" (Stadler, Schubotz, von Cramon, Springer, Graf, & Prinz, 2011). Beta oscillation changes in the neural motor system appear to be involved in predictive timing of sensory input as well as being related to anticipatory motor planning and response selection (e.g., Arnal & Giraud, 2012; Cheyne, 2013). Behavioural evidence that motor simulation involving forward models may be involved in action prediction is provided by the finding that action production interferes with action prediction. The degree of interference observed is modulated by the degree of overlap between the executed and predicted actions (Springer, Brandstädter, Liepelt, Birngruber, Giese, Mechsner, & Prinz, 2011; see also Mulligan, Lohse, & Hodges, 2015 for evidence that a secondary motor action task impacts expert observers' accuracy in predicting the action trajectories).

The importance of a forward model of one's own speech output during articulation is well recognized (Heinks-Maldonado et al., 2006; Christoffels, Formisano, & Schiller, 2007; Hawco, 2009). Indirect evidence of efferent copy can be observed in, for example, auditory cortex suppression during vocalization (see Flinker, Chang, Kirsch, Barbaro, Crone, & Knight, 2010, for evidence and a review; see also Blakemore, Frith, & Wolpert, 1999 for evidence from the somatosensory domain). Of particular interest to the question of speech predictions during comprehension is the fact that auditory

cortex suppression occurs whether speech is fully implemented or imagined (Tian & Poeppel, 2010). This finding led Tian and Poeppel to suggest that "speakers construct a forward model incorporating phonological information under conditions when they do not speak (i.e., do not use the production implementer)" (p. 340). The suggestion is in keeping with a much earlier proposal that, "Although typically associated with sensorimotor systems, corollary discharge might also apply to inner speech or thought, which can be regarded as our most complex motor act"(Ford, Mathalon, Heinks, Kalba, Faustman, & Roth, 2001, p. 2069, citing Jackson, 1958).

Within the Pickering and Garrod framework, the terms "efference copy" and "forward model" appear to be used in an metaphorical rather than a literal sense (for comments see Alario & Hamamé, 2013; Howes, Healey, Eshghi, & Hough, 2013; Jaeger & Ferreira, 2013). The abstract levels of representation at which forward models are suggested to operate (phonological, syntactic, semantic) are considerably removed from the point at which true efferent copies of motor command output are generated (see above). Other speech processing frameworks that incorporate the notion of forward models have been explicit in stating that their forward models operate at a conceptual level and do not involve the production of true efference copies (e.g., Tian & Poeppel, 2010; Hickok, 2012). We return to this matter in further chapters, whilst noting at this point that this matter is not central to the investigation carried out in this thesis: The picture word interference and phonological competition effects to which inform our experimental design and interpretations are understood to arise at a phonological (i.e., abstract) level but have been shown to have effects that are observable at an articulatory level when participants plan their own speech. This is consistent with understanding detailed in the above paragraph that, with respect to speech, forward models may involve representation at a phonological level within the speech production system.

In summary, action emulation provides a parsimonious explanation of the mechanism underlying prediction at a phonological-phonetic level during comprehension. To date, however, evidence cited in support of the view that prediction through the production system is specified at a phonetic-phonological level (and therefore likely to result in speech-motor activation) has been largely circumstantial. Researchers have observed that speech-motor activation occurs during spoken language comprehension and have suggested that this activation may reflect the simulation of upcoming speech sound input through emulations (Watkins & Paus, 2004; Pickering & Garrod, 2004; Schiller et al. 2009). Evidence that action prediction interferes with action production (and vice versa) is cited to support this proposal, under the understanding that dialogue is a form of joint action and that findings from the wider joint action literature are transferrable to spoken language comprehension. The work reported in this thesis investigates specifically whether speech production shows evidence of interference from speech action prediction (i.e., prediction-by-simulation during comprehension).

1.4. Analysing speech as action

Speech differs considerably from other forms of action, not least in that the action is primarily heard rather than seen (see Stevens, 2005, for evidence that visual and motor imagery make differing contributions to action prediction). This difference means that it is particularly important that there be speech-specific evidence to supplement argument based on analogy from other motor action domains. It also means that it is particularly difficult to capture speech action data: Limb movement and movement-readiness data can be non-invasively captured via magnetic or infra-red tracking (e.g., Kilner, Hamilton, & Blakemore, 2007; Stanley, Gowen, & Miall, 2007), or motor evoked potentials (e.g., Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995). Ultrasound imaging currently offers the only risk-free, non-invasive means to image articulatory movement during speech production at a temporal resolution suitable for studies of a psycholinguistic nature.

As noted above, speech activity is typically inferred by reference to its acoustic sequelae, for example voice onset latencies or speech error rates. Such measures do not capture articulatory variability, but historically articulatory variability has been understood to arise of "motor noise" and therefore to be constant across conditions in psycholinguistic experiments. The acoustic approach has been enormously informative, grounding the development of highly influential models of speech production (e.g., Levelt, 1993; Dell, 1988; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997). However, the "motor noise" assumption, whilst pragmatic, is not fully tenable (see Smolensky, 1988 for a philosophical treatment of this issue): Vocal response latencies encompass both cognitive and motor components (Riès, Legou, Burle, Alario, & Malfait, 2012; 2015), and articulation contains traces of cascade from higher-processing levels (e.g., Frisch & Wright, 2002; Goldrick & Blumstein, 2006; McMillan & Corley, 2010; Pouplier, 2007; see Pouplier & Goldstein, 2010 for comment and Chapter 5 of this thesis for details). This may not be particularly problematic within the context of speech production models that intentionally do not extend to incorporate articulation (e.g., Levelt, 1993; Levelt, Roelofs, & Meyer, 1999). However, a framework that treats speech (and potentially comprehension) as a motor activity (e.g., Pickering & Garrod, 2013) requires speech action data. Therefore in this thesis we analyse both response latency and articulatory imaging data in order to investigate whether speech motor behaviour shows traces of activation of the (neural) speech motor system during comprehension related prediction-by-simulation.

The primary active articulator involved in speech production is the tongue (see Keating, 1994). As an internal muscular hydrostat, its movements are particularly difficult to capture and measure. Potential data capture techniques include electromagnetic articulography (EMA; e.g., Narayanan et al., 2014), magnetic resonance imaging (MRI; see Narayanan, Nayak, Lee, Sethy, & Byrd, 2004) and ultrasound tongue imaging (Stone, 2005; for an introduction to speech imaging techniques see Ball, 2016). EMA involves attaching coils to the surface of the tongue, and is therefore somewhat invasive. MRI is non-invasive but currently does not have a temporal resolution appropriate to capture dynamic speech production. These techniques have not yet been widely employed to address psycholinguistically

motivated research questions. This is likely in part due to the fact that articulation has historically been considered to be outside the scope of psycholinguistic models (see above), and in part due to the practical difficulties of capturing and analysing data. However, the Delta method of ultrasound tongue image analysis has been demonstrated to be an appropriate and feasible approach for addressing psycholinguistic questions within an experimental environment (e.g., McMillan, 2008; McMillan & Corley, 2010). In this thesis we develop and extend the technique to investigate pre-acoustic speech action within a prediction-eliciting environment.

1.5. The current study

In the studies reported in this thesis, we elicit prediction through the acoustic presentation of high-cloze sentence fragments, and determine its effect on production activity elicited via picture-naming. We are interested in whether prediction during comprehension elicits speech sound representations within the speech motor system. We address this question by manipulating the degree to which high-cloze targets overlap picture name targets at a phonological level, and examining whether the effects of this manipulation are observable in speech motor output. Production level effects of phonological overlap between the predicted and to-be-produced word would provide evidence to suggest that lexical predictions are indeed represented in a form that contacts the listener's speech motor system, in keeping with Pickering and Garrod's prediction-by-simulation proposal.

As noted previously, vocal response latency studies consistently report an effect of phonological similarity when a distractor is presented perceptually (see above for references to the literature; see also Chapter 2 and Appendix A of this thesis for evidence and discussion): When participants are required to name pictures in the presence of distractor pictures or words, naming response latencies are longer than when the picture is named in isolation. Where there is phonemic overlap between the target and distractor name, this interference effect is reduced. For example, naming latencies to the image TAP are shorter when it is accompanied by auditory presentation of the word TAN than when it is accompanied by an entirely non-overlapping word form such as CONE. This effect indicates that perceptually available distractors are subject to representation at a speech-sound level in a form that contacts the speech production system. If comprehension-related predictions are represented at a speech-sound level within the speech production system, we would expect to see a comparable effect, with response latencies being shorter when the high-cloze target overlaps the picture name in onset- or rime- position than when it does not. This possibility is investigated in the first part of this thesis: In Chapter 2 we present a study in which we confirmed that the perceptual phonological similarity effect described above was observable for items that would be used in our subsequent prediction studies. In chapters 3 and 4 we present prediction studies in which we tested whether phonological overlap between a predicted item and a picture-to-be-named affects picture naming latencies. We observed a phonological overlap effect when items were presented perceptually but not when items were elicited through prediction.

In the second part of the thesis, we investigate whether articulation shows evidence of interference from speech motor system activation elicited during comprehension-related prediction. Although articulatory imaging is a nascent method of psycholinguistic investigation, a number of studies have used this approach to investigate the effect of phonological competition arising of one's own speech planning: These studies have elicited phonological competition either by requiring participants to perform tongue-twisters (i.e., to repeat alternating phonologically similar sequences; McMillan & Corley, 2010; Pouplier & Goldstein, 2010) or by generating a word onset order expectancy and then violating it (e.g., McMillan, Corley, Lickley, 2009; Pouplier, 2007). When speech motor activity is recorded via articulatory imaging (electropalatography, electromagnetic articulography, or ultrasound tongue imaging), it reveals traces of phonological level competition during speech planning (see Chapter 5 for details and discussion of techniques and findings). Where items compete at word onset, speech motor activity is more variable than when items do not compete at word onset. If comprehension-related predictions are represented at a speech-sound level within the speech production system we would expect to see a comparable effect of predictions, with speech motor activity revealing traces of competition from prediction-related activation. In Chapter 5 we introduce the technique used to investigate this possibility. In Chapters 6 and 7 we present studies in which we tested whether comprehension-related predictions induce interference at an articulatory level and whether this interference is speech-sound specific.

The question of whether interference is speech-sound specific is crucial in considering whether speech motor effects reflect prediction-by-simulation as opposed to a more general articulatory response to expectancy violation. As reviewed above (see Section 1.1), a number of studies have found that speech motor activation in response to perceptual presentation of speech sounds is speech-sound specific at a neural and effector level: Electromyelographic activity within a given articulator is raised when the participant is auditorily presented with a speech sound involving that articulator. Listening to speech sounds evokes somatotopically-specified activation within the primary motor cortex. Comprehension task performance (word-picture matching) is affected by externally-induced somatotopically-specific simulation of the primary motor cortex. Such somatotopically specific responses to hearing speech sounds have been investigated under paradigms which involve the presentation of single words or syllables in isolation. When sounds are presented in isolation each item can be associated with one primary articulator. This allows the researcher to investigate whether sounds associated with involvement of a given articulator during production evoke somatotopically specific activation within the listener's primary motor cortex. It also allows the articulatory contrast between labial and lingual onsets to be exploited in order to investigate whether somatotopic activation of speech motor regions plays a causal role in speech comprehension.

Unfortunately, the question of whether comprehension-related predictions are associated with somatotopic activation of the neural speech-motor system does not lend itself to the approaches employed to investigate the effects of perceptual presentation. When investigating reactive speech-motor activation associated with sensory presentation of a linguistic item, only the critical item need

be presented in each trial. In order to study proactive speech-motor activation associated with prediction-by-simulation, it is necessary to present linguistic content sufficient to constrain potential upcoming input to an extent that the listener predicts the critical item (at some level). The presentation of such linguistic content will in itself induce reactive speech-motor activation. The techniques that are employed to study reactive somatotopic activation during perception would currently not allow differentiation between prospective activation associated with prediction and retrospective activation associated with perception[2]. A neural imaging approach is therefore not currently amenable to addressing the questions central to this thesis. However, given that activation at a phonological level has been found to leave traces in speech motor output (see above), direct observation of speech motor behaviour allows a potential window on speech sound representation within the speech production system during prediction of upcoming linguistic input.

Chapters 6 and 7 of this thesis report studies in which we employed an ultrasound tongue imaging approach in order to allow direct observation of speech motor behaviour within a predictive context. We manipulated the degree and nature of form overlap between predicted and to-be-named items. This allowed us to investigate whether predictions are specified in an articulatorily-specified manner, as would be anticipated under Pickering and Garrod's prediction-by-simulation proposal. To pre-empt our findings, it appears that predictions elicited during comprehension do make contact with the listener's speech motor system. However, we did not find evidence that predictions are represented at a speech segment level in a production-accessible manner; rather, it appears, the predictions are represented at a syllable (or phonological word) level.

This finding is compatible with the general proposal that the speech production system is active during the generation of lexical predictions during comprehension. It does not support the proposal that the production of individual speech segments is simulated within the listener's speech-motor production system during prediction of upcoming spoken input. The findings are consistent with a speech processing architecture proposed under the Hierarchical State Feedback Control model (HSFC; Hickok, 2012), under which phonology is represented at lower and higher levels. The higher level involves syllable level units and the lower level involves motor-encoded segment level units, with both levels including production and comprehension processing streams. The HSFC incorporates forward models at pre-motor levels, and allows for a mechanism that inhibits the cascade of activation

[2] Disruptive TMS has been employed within a visual world paradigm in order to investigate the potential involvement of the cerebellum in prediction-during-comprehension (Lesage, Morgan, Olson, Meyer & Miall, 2012). This study provided evidence to suggest that processing within the cerebellum supports "prediction during motor control and language processing". In this study predictions were associated with semantic-syntactic constraints provided by the verb. Slowed predictive processing in the presence of disruptive TMS is therefore interpreted at reflecting cerebellar involvement at that level rather than a speech-motor level. It has separately been proposed that the internal (i.e., forward) models operating at an articulatory level are cerebellar (e.g., Hickok, 2012). Argryopoulos (2015) provides a review relating to the fact that "whether cerebellar internal models in language comprehension recruit language generation mechanisms remains an outstanding question".

to motor-encoded segment representations during listening for comprehension. The HSFC model was not developed to account for prediction during spoken language comprehension. However, our data suggest that it provides a suitable model through which to improve understanding of the interactions between spoken language comprehension and production which, as Pickering and Garrod have highlighted, appear to underlie the predictions we make during comprehension.

Chapter 2: Item-Set Suitability

2.1. Introduction

As detailed in Chapter 1, the main body of this thesis investigates whether prediction during comprehension evokes representations within the speech motor system of the listener. Predictions are elicited via presentation of high cloze sentence fragments. Production activity is observed during picture naming. Response latencies and articulatory activity are analysed to determine whether these are impacted by the presence and nature of predictions. Interference at an articulatory level would indicate that predictions are represented at a speech-motor level. If interference is reduced when there is phonological-phonetic overlap between the predicted and produced words, this would suggest that speech-motor representations are articulatorily specified.

The rationale for the above approach rests, to some extent, on evidence from picture word interference (PWI) studies: When a picture-to-be-named is presented in the presence of a perceptually available distractor, phonological-phonetic overlap results in reduced interference effects (see Chapter 1 and below for details). However, previous evidence of interference at an articulatory level comes not from PWI studies but from error-elicitation studies involving tongue twisters or SLIP tasks (e.g., McMillan, 2008; Pouplier, 2007; see Chapter 5 for more details). For theoretical and practical reasons, the items used in error-elicitation studies are typically monosyllabic, with "distractors" (or competing items) often differing from the target by only one phonological feature. For example, the tongue twister words ("cop" – "top") employed by McMillan and Corley (2010) differ only by place of articulation of the onset consonant. This degree of similarity between target and distractor is not typically seen in phonological PWI studies, in which competing items usually differ by multiple phonemes and features (e.g., swear – BEAR; Damian & Bowers, 2009).

In addition to differing in the degree of featural overlap between competing items, the design of PWI and articulatory interference (error elicitation) studies differs at a syllabic level. Articulatory interference studies almost exclusively investigate the effects of competition at word onset position (i.e., competing items share a syllable rime; e.g., c*op* - t*op*), whereas PWI effects are most stable and most frequently investigated in onset overlap conditions (e.g., *co*p – *co*t); see Appendix A details). PWI and error elicitation studies further differ in that PWI studies typically group monosyllabic and multisyllabic words together, with the first syllable of competing multisyllabic words overlapping fully (see Meyer & Schriefers, 1991, for an exception), whereas studies in which articulatory interference has been observed have exclusively employed monosyllabic items.

In this chapter we present a pilot study, conducted in order to confirm that PWI facilitation effects are observable for items of the type that have previously revealed articulatory interference in error elicitation paradigms (i.e. exclusively monosyllabic words, with an onset competition condition included). The PWI study we report includes the items that would be included in subsequent studies reported in this thesis. This allows us to confirm that when the distractor items we employ in further studies reported in this thesis are perceptually available (as opposed to being represented solely via top-down prediction processes as they are in subsequent studies in this thesis), they elicit the effects typically observed within a PWI paradigm. This allays the concern that, were phonologically mediated effects not to be observable within the prediction paradigm, this could be due to peculiarities of the item-set rather than differences in the nature of representations elicited. Below we review the PWI paradigm and relevant findings prior to presenting details of the current study.

Picture word interference (PWI) and picture picture interference (PPI) paradigms have been used to investigate the scope and nature of phonological encoding during word production (Appendix A provides an overview of relevant studies). Participants are required to name a sequence of pictures while ignoring distractor words. The phonological relationship between the distractor words and pictures to be named is systematically varied. Distractor words may be presented in written form, superimposed on the picture-to-be named (e.g., Damian & Dumay, 2007; de Zubicaray, McMahon, Eastburn, & Wilson, 2002; Glaser & Düngelhoff, 1984; Lupker, 1982), auditorily (e.g., Damian & Martin, 1999; Meyer & Schriefers, 1991), or pictorially (e.g., Meyer & Damian, 2007; Morsella & Miozzo, 2002; Navarette & Costa, 2005). Regardless of distractor modality, participants are usually slower to name pictures in the presence of a distractor than when pictures are presented in isolation. Participants are quicker to name pictures when the distractor word shares phonological segments with the picture-to-be-named than when it does not (e.g., Damian & Dumay, 2007; de Zubicaray & McMahon, 2009; Lupker, 1982; Meyer & Schriefers, 1991). This effect is commonly referred to as a "phonological facilitation effect", although it has also been interpreted as a phonetic effect (see Lupker, 1982).

The phonological facilitation effect is assumed to arise either at a word-form retrieval or at an encoding level where phonological segments of both the distractor word and the target picture name are activated (e.g., Roelofs, 1997; Schriefers, Meyer, & Levelt, 1990; Starreveld & La Heij, 1995; Damian & Martin, 1999). It is argued that, when the distractor word and target picture name share phonological segments, either competition is reduced (see Meyer & van der Meulen, 2000, p., 315), or segments are pre-activated by the distractor and are then more readily activated in response to the target (see Meyer, Schriefers, Levelt, 1990, p. 99).

When distractors are presented orthographically, the size of the phonological facilitation effect is generally not affected by the position of the phonological overlap when the target and distractor are presented simultaneously (begin v. end related; Meyer, 1996; Damian & Dumay, 2007). Similarly, when distractors are presented pictorially (i.e., in picture-picture interference paradigms) a

phonological facilitation effect is observed in both word-onset and word-rime overlap positions (e.g., Meyer & Damian, 2007; Morsella & Miozzo, 2002; Navarrete & Costa, 2005). However, when distractor words are presented auditorily, effects tend to be larger and occur earlier if overlap occurs at the beginning of a syllable/word than at the end (e.g., Meyer & Schriefers, 1991). The mechanism underlying this finding continues to be debated. Phonological facilitation associated with auditorily presented rime-overlap distractors appears to decrease over the course of childhood (ages 5 to 11 years; Brooks & MacWhinney, 2000). It has been suggested that the decrease over the course of development reflects a move from holistic to incremental encoding. An alternative interpretation of auditory PWI findings is that onset-overlap facilitates lexical selection via a cohort effect (for argumentation and rebuttal see Wilshire, Singh, & Tattersall, 2015).

As noted above, effects of phonological overlap within PWI are assumed to arise at phonological level (see Roelofs, 1997). In this respect, understanding of the effect is subject to more general questions concerning the nature of phonological encoding: Accounts of the form-encoding process differ as to whether abstract phonemes of a morpheme are activated in sequence (e.g., Dell, Juliano, & Govindjee, 1993; Hartley & Houghton, 1996; Houghton, 1990; Sevald & Dell, 1994; Vousden, Brown, & Hartley, 2000) or are activated simultaneously and then selected in sequence (e.g., Levelt et al., 1999; Roelofs, 1997; 2004; 2015). Minimalist accounts assume that articulation can start as soon as the initial segment of an utterance becomes available (e.g., Dell, Juliano, & Govindjee, 1993; Jordan, 1990; MacKay, 1987; 2012). Non-minimalist accounts assume that articulation cannot start until the first syllable or phonological word has been planned: Segments are sequentially assigned to a frame during the syllabification and prosodification processes necessary to allow the generation of appropriate motor commands via access to the mental syllabary (see Wilshire et al., 2015 for a review; see also Roelofs, 2015). Syllabification of a word can commence once the first few segments have become available, but the output will be buffered until all segments are available. Buffered forms can only be extended from left to right (i.e., sequentially from word onset to word end), as indicated by the finding that anticipatory knowledge of phonological segments of words to be said only speeds response times when the segments occur at the beginning of the word (see Roelofs, 2015, p.35 for a summary).

Whilst differences between minimalist and non-minimalist accounts are pertinent to investigations described later in this thesis, the study described in this chapter seeks simply to replicate previously observed effects; as such it does not contribute to this debate. Phonological effects have previously been observed in studies that employ primarily monosyllabic items (e.g., de Zubicaray, McMahon, Eastburn, & Wilson, 2002; mean syllables per item = 1.25, mean phonemes per item = 4.6). However, subsequent studies in this thesis employ exclusively monosyllabic words, with the same items acting as both targets and distractors (for reasons of articulatory comparability and experimental control; see Chapters 3, 4, and 5): Given the existence of non-minimalist accounts of the relationship between form-encoding and articulatory execution, we were particularly concerned to confirm that the

item set to be used throughout this thesis was capable of eliciting phonological facilitation effects typically observed in PWI studies.

In the current study we chose to present distractor items orthographically and at a stimulus onset asynchrony (SOA) of 0ms. That is, we presented pictures-to-be-named simultaneously with a distractor word. This ensured that participants had access to the full lexical form of the distractor whilst formulating the picture name. Findings of semantic interference and phonological facilitation have consistently been found at 0ms SOA (see Appendix A), whereas semantic interference is sometimes not observed at positive SOAs and phonological facilitation is sometimes not observed at negative SOAs. We therefore employed a 0ms in order that the conditions during perceptual presentation should be as comparable as possible to those included in later experiments in which the distractor representation was elicited via prediction (following Pickering & Garrod, 2013, we assume that predictions are subject to representation at multiple linguistic levels).

2.2. Method

2.2.1. Participants

Fourteen participants (7 male) between the ages of 18 and 20 years of age (mean age = 19.2 years) took part in the study. All participants were monolingual speakers of British English, had normal or corrected-to-normal vision, and reported no positive history for speech-language, learning or hearing difficulties. Participants were recruited via the University of Edinburgh PPLS subject pool system, gave informed written consent in line with BPS (British Psychology Society) guidelines, and were awarded course credit for their participation. The study was granted Ethical Approval by the Psychology Research Ethics Committee of the University of Edinburgh (approval no. 157-1415/2).

2.2.2. Materials

Twenty black-and-white line-drawings acted as experimental items, with a further two pictures acting as practice items. We standardized picture dimensions to 705 x 705 pixels. Target picture names were monosyllabic. Each target picture name had within the experimental item set both a rime partner (e.g., C*AP* – T*AP*) and an onset partner (e.g., *CA*P – *CA*N). In the experimental conditions, a typed "distractor" word (lower-case, black Arial font 72) was superimposed in a central position on the picture. We varied the phonological relationship between the word and the picture name so as to generate 4 experimental conditions: *full match* (in which the word and picture name were identical, e.g., CAP – "cap"); *onset overlap* (e.g., CAP – "can"); *rime overlap* (e.g., CAP – "tap"); *mismatch* (e.g., CAP – "tone"). Pictures were also named in a *control* condition, in which no word was superimposed (e.g., CAP – " "). All picture names also appeared within the experiment as distractor words. The experimental design was fully within participants and within items, with all participants

seeing all items in all conditions once across the experiment. Experiments reported in further chapters in this thesis made use of a reduced set of items selected from this item set: in the "reduced" data set target picture names differed from their rime-overlap distractor by only by place of articulation of the onset consonant (results for this reduced data set are reported in table 2.3.3).

2.2.3. Procedure

Participants performed the experiment individually within a sound-attenuated booth. The experiment was presented on a Dell Latitude laptop (model XPS L702) via DmDX software (Forster & Forster, 2003). Participants wore a Sennheiser HMD46 headset (earphones and microphone) throughout the experiment. Spoken responses were recorded directly via the experimental software and externally via a Zoom H2 digital recorder. Participants were trained on the target picture names in order to minimise any effect on response times of uncertainty as to the "correct" picture name. During the training phase each picture name was first auditorily presented, immediately followed by the relevant picture presented on screen for 3000 ms. Pictures were then presented on screen for naming, with the picture name being auditorily presented as a confirmation (or correction) once the picture had been on screen for 5000 ms.

In the experimental phase, participants were instructed that they would again see the pictures, and that they should name these as quickly and accurately as possible. They were informed that sometimes pictures would be accompanied by a written word, but that in all cases they should name the picture. Two practice items were presented in the mismatch condition (i.e., superimposed with an unrelated written word; e.g., DOOR - leaf). Participants then proceeded to the experiment proper. In each trial a fixation cross was presented for 1000ms, following which the picture for naming was presented on screen for 3000ms (alongside the distractor, in relevant conditions). The control and experimental conditions were interleaved. Trial presentation order was randomized within the experimental software with the constraints that no picture or condition could appear more than three times consecutively. All pictures appeared once in each condition. The experiment took around 20 minutes to complete.

2.3. Results

Data were analysed using a mixed-modelling approach implemented in R 3.1.2 via the lme4 package, version 1.1-7 (Bates, Maechler, Bolker, & Walker, 2014). Error rates were analysed using binomial mixed-effects models fit with Laplace estimation. Response time data were analysed using linear mixed-effects models fit by maximum likelihood, using contrast coding.

2.3.1. Error data

In total there were 1400 responses, of which 35 (2.5%) were contextual errors (i.e., phonologically related to the target picture-name or distractor word) and 17 (1.2%) were non-contextual errors (i.e., the participant misnamed the picture, for example RAIL -> "fence"). We analysed contextual error numbers to determine whether these were influenced by the condition in which a picture was named. The analysis included random intercepts for participants and items; the data were insufficient to support a model which included random slopes. Model fit was significantly improved by the inclusion of condition as a fixed effect ($\chi^2_{(4)}$ = 12.76, p = 0.01). The coefficients indicate that participants made significantly more errors in the rime overlap condition than in the control condition (β = 0.96, SE(β) = 0.30, z = 3.169, p < .01), but that error numbers in other experimental conditions did not differ significantly from those in the control condition (all p > .25).

Table 2.3.1: Experiment 1, response type (correct vs. error) by condition

	Control		Match		Rime overlap		Onset overlap		Mismatch	
Correct responses	275	(98.21%)	274	(97.86%)	263	(93.93%)	270	(96.43%)	267	(95.36%)
Contextual error	2	(0.71%)	4	(1.43%)	14	(5%)	6	(2.14%)	8	(2.86%)
Noncontextual error	3	(1.07%)	2	(0.71%)	3	(1.07%)	4	(1.43%)	5	(1.76%)

2.3.2. Response latency data

We subsequently performed an analysis of vocal response latencies for correct responses. Outliers (latencies lower than 200 ms or greater than 2000 ms) were excluded from this analysis, leaving a total of 1288 data points. The analysis included random intercepts for participants and items, and random slopes for participants. Model fit was significantly improved by the inclusion of condition as a fixed effect ($\chi^2_{(4)}$ = 15.44, p < 0.01; see Table 2.3.2 for full model details). The coefficients indicate that participants responded more slowly when the superimposed word and target picture name mismatched (i.e., in onset overlap, rime overlap and mismatch conditions) than when they matched (i.e., in the match condition; β = 79.09, SE(β) = 20.96, t = 3.77). Comparing response latencies in the mismatching conditions, participants were slower to respond when there was no phonological overlap between the word and the target picture name (i.e., in the mismatch condition) than when there was phonological overlap (i.e., in onset overlap and rime overlap conditions; β = 38.56, SE(β) = 16.35, t = 2.358). Response latencies in the onset overlap condition did not differ from those in the rime overlap condition (β = 9.30, SE(β) = 12.45, t = .75). When the superimposed word and target picture

matched, participants were no slower to respond than when the picture was presented in isolation (i.e., control condition; β = 0.37, SE(β) = 23.68, t = 0.02; see Figure 2.3.1).

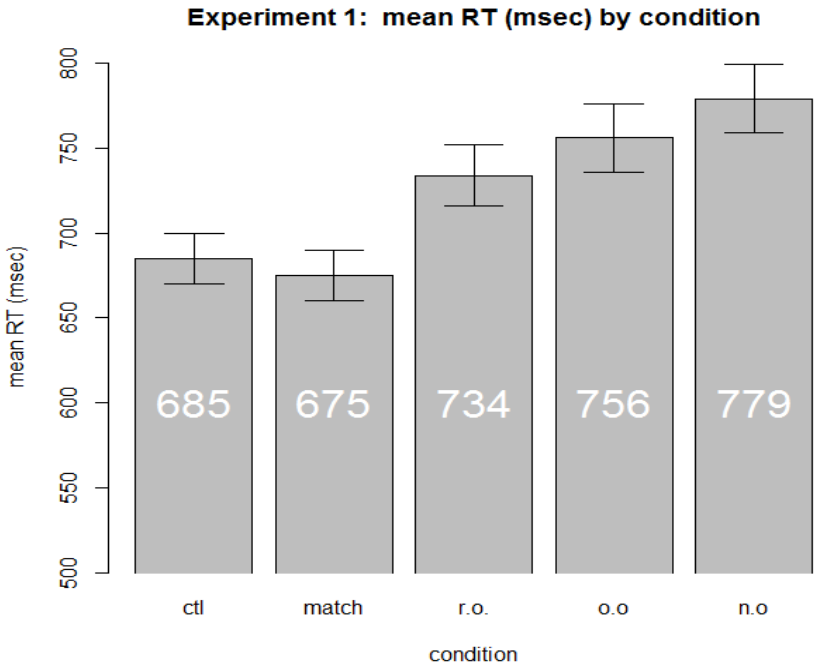

Figure 2.3.1: Experiment 1 mean response latencies by condition for correct responses

Mean RT in each condition is shown in white text on bar. Ctl = control condition, match = full-overlap condition, r.o. = rime-overlap condition, o.o. = onset overlap condition, n.o. = no overlap condition. Error bars indicate confidence intervals.

Table 2.3.2: Experiment 1 model coefficients (in ms) for naming latencies

Fixed Effect	**Estimate**	**SE**	**t**	**Random Effect**		**Variance**
Intercept	748.57	32.93	22.73	Image	Intercept	610.3
Condition				Participant	Intercept	14297.4
Match v Control	0.37	23.68	0.02		MvC	3419.5
All Mismatch v Match	79.09	20.96	3.77		AMMvM	3135.6
Onset v Rime overlap	9.30	12.45	0.75		OvR	1000.3
Mismatch v Rime + Onset overlap	38.56	16.35	2.36		MMvR+O	244.5

The analyses above were performed on a data set that contained eight items that were not included in the subsequent prediction experiments reported in this thesis. We performed a supplementary response latency analysis in which we included only those 12 items that constituted the item set in the main prediction experiments. This allows us to rule out the possibility that any between-experiments differences might arise due to item-set rather than the nature of representations elicited. Model fit again improved over the null model when context was included as a fixed effect ($\chi^2_{(4)} = 18.04$, $p < 0.01$; see Table 2.3.3 for full model details). Participants were quicker to name pictures when the word and picture name fully overlapped than when they did not fully overlap ($\beta = 98.58$, $SE(\beta) = 21.47$, $t = 4.59$). Participants were faster to name pictures when the word and picture overlapped in part (onset overlap/rime overlap contexts) than when they did not overlap at all (no overlap context; $\beta = -27.68$, $SE(\beta) = 13.69$, $t = -2.02$). Response latencies did not differ significantly by type of partial overlap (onset overlap v rime overlap; $\beta = -35.42$, $SE(\beta) = 26.77$, $t = -1.32$). In summary, results obtained for this reduced data set replicated those found for the full data set: Any subsequent by-experiment differences in patterns of results can be understood as arising due to paradigm differences rather than the item set differences.

Table 2.3.3: Experiment 1 (subset) model coefficients (in ms) for naming latencies, no obs = 776

Fixed Effect	**Estimate**	**SE**	**t**	**Random Effect**		**Variance**
Intercept	753.98	34.38	21.93	Image	Intercept	685.3
Condition				Participant	Intercept	15007.2
- Match v Control	-1.89	24.59	-0.08		MvC	1784.9
- All Mismatch v Match	98.58	21.47	4.59		AMMvM	3322.8
- Mismatch v Rime + Onset overlap	27.68	13.69	2.02		MMvR+O	2987.5
- Onset v Rime overlap	-35.42	26.77	-1.32		OvR	1988.8

2.4. Discussion

We obtained typical PWI phonological facilitation effects in a study in which all manipulations were fully within-participants and within-items, and in which all items were monosyllabic and monomorphemic. In keeping with previous findings (e.g., Damian & Martin, 1999; Lupker, 1982; de

Zubicaray & McMahon, 2009; see Appendix A for further details), the phonological facilitation effect manifested as a reduction in interference associated with the presentation of distractor words. That is, naming latencies were shortest in the Match condition, in which the distractor and target fully matched, but shorter when there was phonological overlap between a distractor and target than when there was not. This pattern held true for the "reduced" data set, in which target picture names differed from their rime-overlap distractor by only by place of articulation of the onset consonant. This confirms that the distractor-target contrasts employed in our subsequent investigations of the effect of predicted items are capable of eliciting statistically significant effects when distractor items are perceptually available (i.e. in a typical PWI paradigm).

The degree of phonological facilitation observed in the current study did not differ by overlap type, and therefore our results do not distinguish between the type of phonological overlap employed in articulatory interference experiments and typically included in PWI studies. Our results do, however, confirm that phonological facilitation occurs for monosyllabic picture-names when there is only part-syllable overlap with a perceptually available distractor. This finding is crucial to the premise of subsequent experiments reported in this thesis, in which we investigate whether representations elicited in a top-down manner via prediction produce effects similar to those observed during the bottom-up processing of perceptually available stimuli (as is suggested by, for example, Pickering and Garrod, 2007).

Chapter 3: Prediction elicitation

3.1. Introduction

Our intention in the current thesis is to investigate the speech-motor effects of prediction during comprehension. We elicit predictions by auditorily presenting participants with high-cloze sentences from which the high-cloze sentence-final target has been excised. We elicit speech motor activity by presenting a picture for naming immediately following auditory presentation of the sentence-stem. This approach can be understood as an extended PWI paradigm, in which the "distractor" is predicted rather than perceived. In this chapter we present a study in which we piloted the paradigm, adopting the conventional PWI approach of investigating vocal response latencies (see Chapter 2). We situate the current study by briefly reviewing studies which have employed a similar paradigm to investigate text comprehension, speech comprehension, and speech production processes. We outline adaptations that were necessary in order to allow the paradigm to be employed to investigate phonological-phonetic level representations. We then present our findings and discuss their implications for the design of subsequent studies within this thesis.

Early studies of comprehension-based contextual constraint investigated the processes by which text is read for comprehension (e.g., Stanovich & West, 1979; Tulving & Gold, 1963; see also Goodman, 1967 for theoretical background). The response stimulus was generally a written item presented for either word naming (see examples above) or a lexical decision (e.g., Fischler & Bloom, 1979), with the outcome measure being vocal or manual response latency. The speed with which a word can be processed for reading (aloud) was found to be related to the likelihood of the word being encountered in that context (i.e., its cloze probability), suggesting that in high-cloze contexts participants pre-activate features of anticipated words (see Chapter 1 for further evidence and discussion).

Multi-modal studies of the effects of sentential context on picture (and word) naming have principally been conducted under the framework of a "Competition Model" of language acquisition, comprehension and production (see Bates & MacWhinney, 1989; Li & MacWhinney, 2013; MacWhinney, 2008). This framework, in common with more recent, frameworks that incorporate forward-modelling (e.g., Dell & Chang, 2014; Hickok, 2012; Pickering & Garrod, 2013), treats production and comprehension as highly integrated. However, studies conducted under the Competition Model framework have generally been designed to investigate the effects of semantic and syntactic congruency rather than effects of prediction per se: Sentential context is considered to provide semantic and syntactic cues to potential incoming information (see Appendix B for examples) which affect ease of lexical access during comprehension. Picture or word naming latency is treated as an index of cue-type strength: It is assumed that relatively stronger cues will have a greater impact

on naming latencies than relatively weaker cues, with cue-picture congruence facilitating picture naming and cue-picture incongruence potentially inhibiting naming (see MacWhinney, 2008).

Experiments conducted under the Competition Model framework indicate that the effect of semantic congruency on picture/word naming response times is relatively invariant across modalities (e.g., written versus auditory sentential context; Bentrovato, Devescovi, D'Amico, & Bates, 1999; Bentrovato, Devescovi, D'Amico, Wicha, & Bates, 2003), languages (e.g., Bentrovato et al., 1999; Wicha , Orozco-Figueroa, Reyes, Hernandez, Galvadón de Barreto, & Bates, 2005), and age groups (e.g., Roe, Jahn-Samilo, Juarez, Mickel, Royer, & Bates, 2000). In keeping with the findings of text processing studies, response latencies obtained in a "contextually unbiased" neutral context (e.g., "The teacher told her to repeat the word BOOK ten times"; "Now please say SPOON") are longer than those obtained in a congruent context (i.e., in a context that provides "strong semantic constraints for a target picture name"; Roe et al., 2000, p. 757; e.g., "Peter eats his soup with a SPOON"; see Appendix X for further studies and examples). The effect is relatively invulnerable to repetition effects, suggesting that the facilitation arises of a "relatively automatic process" (Roe et al., 2000, p.763). This interpretation is in keeping with comprehension literature in which facilitation associated with sentential or semantic context is understood to reflect early-stage activation of candidate representations (see Aydelott & Bates, 2004). Such facilitation is argued to arise of an automatic process because it is observed even when participants are not consciously aware of the context material, and at short stimulus onset asynchronies and (for review and discussion see Neely, 1991).

The effect of incongruent sentential contexts appears to be less stable than that of congruent sentential contexts, and is found reliably only when the target item mismatches the context both semantically and syntactically (e.g., Bentrovato et al., 1999; Bentrovato et al., 2003; Wicha et al., 2005; see Appendix B for details of individual studies). Roe and colleagues (2000) generated semantically incongruent trials by re-arranging the pairing of sentence stems and pictures so as to generate unlikely combinations (e.g., Peter eats his soup with a BOOK). They found that participants at the extreme ends of the lifespan (i.e., 3-5 years and 70+ years of age) were slower to name pictures in the incongruent context than in the neutral context (described above), but that this effect was not reliably present in younger adults.

Within the context of the current thesis, it is particularly relevant that Roe and colleagues suggest that the inhibitory effect of incongruous sentential context may be explained in terms of speech production as well as speech comprehension, proposing that a cued item is activated and "intrudes itself into sentence comprehension and/or sentence planning" (Roe et al., 2000, p. 763) thereby evoking competition that slows lexical selection. The inhibitory effect decreased over the course of the experiment, a feature which the authors attribute to sentential interference being, "vulnerable to a build-up of strategies and expectations across the course of the experiment" (Roe et al., 2000, p. 757). They therefore argue that the inhibition effect reflects the operation of strategic as opposed to automatic processing. However, it should be noted that sentence-stems in the neutral condition were

shorter, fewer, and presented more frequently than those in the congruent and incongruent contexts; this being the case, the relative inhibition effect may simply reflect greater processing costs associated with the sentence-stems which provided the incongruous context. Participants encountered each neutral sentence-stem four times over the course of the experiment, whereas constraining sentence-stems were encountered only once in the congruous context and once in the incongruous context. The within sentence-stem predictability of upcoming words was therefore greater in the neutral context than in other contexts. Attentional deployment varies with predictability (e.g., Fischler & Bloom, 1979); the longer reaction times observed in incongruous contexts than neutral contexts may be attributable to this aspect alone, rather than to strategic processes.

The studies described above were concerned primarily with the investigation of comprehension processes. The same paradigm has been used to investigate speech production processes, specifically the relationship between lexical selection and phonological encoding during picture naming (Griffin & Bock, 1998; see also Ferreira & Pashler, 2002; Gollan, Slattery, Goldenberg, van Assche, Duyck, & Rayner, 2011; Piai, Roelofs, & Maris, 2014). Griffin and Bock (1998) argue that the paradigm is appropriate to investigate this aspect of speech production because "although reading sentence frames differs from generating messages, the product of comprehension should be similar to the conceptual representations that speakers normally develop" (Griffin & Bock, 1998, p. 329). They vary sentence fragment constraint (i.e. cloze probability) in order to manipulate picture name "redundancy" (i.e. the extent to which the picture name is uniquely activated by the specifications and constraints of the sentence fragment; high-cloze contexts generate high redundancy). Redundancy is understood to affect picture naming reaction times by facilitating the lexical retrieval process (for discussion and evidence see Griffin & Bock, 1998). Word frequency is understood to affect picture naming reaction times at a phonological encoding level, with higher-frequency forms being retrieved more rapidly than lower-frequency forms. Griffin and Bock report that degree of redundancy (cloze-probability) interacts with lexical frequency, such that in high-constraint contexts the lexical frequency effect was attenuated. This interaction appears to reflect automatic rather than strategic processing, as it is observed even when picture names match the high-cloze 'predicted' item in only 20% of trials. As the lexical frequency effect is understood to originate at a later stage in the word production process than the redundancy effect, Griffin and Bock (1998) argue that the interaction is best explained under a cascaded framework: They propose that high-constraint contexts prime top-down activation of phonological form, but do not consider this to be a prediction process (see Piai, Roelofs, & Maris, 2014, for discussion).

The nature of the relationship between sentential constraint and word frequency effects in picture naming has been further investigated using both eye-tracking and neural imagining approaches (Gollan, Slattery, Goldenberg, van Assche, Duyck, & Rayner, 2011; Piai, Roelofs, & Maris, 2014, respectively). Sentential constraint was found to modulate the power of neural oscillations in the theta band (associated with long-term memory access; Bastiaansen, Oostenveld, Jensen, & Hagoort, 2008; Düzel, Penny, & Burgess, 2010; Nyhus &Curran, 2010) at least 400ms prior to picture presentation

(Piai et al., 2014). This suggests that highly-constrained sentence contexts allow participants to preactivate specific lexical representations. This interpretation is strengthened by the finding within the same study that during the final 500ms of sentence context presentation, theta band oscillatory power was greater when fragments predicted high-frequency targets, than when fragments predicted low-frequency targets (i.e. lexical frequency effects were observable prior to picture presentation). Further, in highly constrained contexts only, oscillatory power in a broad alpha-beta range (8-30Hz) was attenuated during a period extending from 400ms prior to picture presentation onset until 200ms post picture presentation onset. This suggests that when context causes participants to pre-activate specific lexical representations, they also engage in motor preparation activity (presumably associated with preparation for articulation; see de Alegre, Gurtubay, Labarga, Iriarte, Malanda, & Artieda, 2004; Neuper, Wörtz, & Pfurtscheller, 2006; Cheyne, 2013 for evidence that motor preparation and execution are associated with decreases in oscillatory power across alpha and beta bands).

In summary, the sentence-constraint/picture naming paradigm has been employed to study text and spoken language comprehension and speech production processes. It has consistently been found that when a picture (or word) forms the high-cloze target highly-constraining sentence, its naming is facilitated. It is widely accepted that the facilitation effect arises at lexical access level, with researchers proposing a variety of mechanisms by which sentence context favours the activation of the high-target cloze item over potential competitors. Researchers differ in the extent to which they understand this process to be under strategic control, and in how they interpret the time course of activation effects (i.e., concerning whether there is (pre)activation of high-cloze target features prior to picture presentation or whether contextual congruity eases processing post picture presentation). Both semantic and syntactic congruity facilitate picture and word naming compared to naming in a neutral context which does not constrain upcoming words (e.g., Bentrovato et al., 1999; Bentrovato et al., 2003). Under the prediction-as-production framework of Pickering and Garrod (2004; 2007; 2013) it would be expected that phonological congruity should similarly facilitate picture naming (see Chapter 1). Whether picture/word naming is sensitive to phonological constraints associated with lexical level activation has not previously been directly investigated. There is indirect evidence, from lexical frequency studies, that sentential context induces word form level (pre)activation of high-cloze targets prior to picture/word presentation.

In order to address the question of whether highly predictable material is represented at phonological-phonetic level within the listener's speech production system (as proposed by Pickering & Garrod, 2007; 2013), it was necessary to make certain adaptations to the sentential PWI paradigms described above. As we are concerned specifically with processes that underpin spoken language processing, sentential context was presented acoustically: Acoustic presentation has previously been demonstrated to elicit semantic congruency effects (e.g., Roe et al., 2000), but potential form-level effects have been investigated only via serial visual presentation of a sentential context (e.g., Stanovich & West, 1979; Griffin & Bock, 1998; Gollan et al., 2011; Piai et al., 2014). The use of

acoustic material is crucial to address the suggestion that speech motor activation occurs as listeners simulate to-be-heard material via their own speech production system.

Our approach differs from previous instantiations of the sentential PWI paradigm in that all target picture names are monomorphemic monosyllables: As we are concerned with phonological-phonetic activation within the production system, it was necessary to experimentally control the morpho-phonological and phonetic complexity of our target picture names. Picture names were selected so that, within the full set, each name had one onset-overlap partner, one rime-overlap partner and one no-overlap partner (e.g., CAN – CAP; CAN – TAN; CAN – TAPE, respectively; see Chapter 2 and Appendix C for further details). Target picture names therefore feature as the high-cloze sentence-stem targets which act as "distractors". In this way the paradigm is directly comparable to that employed for the conventional PWI study described in Chapter 2 of this thesis, the only difference being that in the experiments reported below the distractor is elicited via prediction arising of sentential context rather than being presented orthographically.

The item set we employ allows us to generate experiments with a fully within items design, in which all participants encounter all items an equal number of times in each experimental condition. This minimizes the potential for confounding effects and allows us to directly compare articulation across conditions. However, sentential context effects have not previously been demonstrated for such a homogeneous set of items. Our use of the paradigm to investigate potential phonological-phonetic effects relies on the assumption that sentential context effects are present (i.e., that distractor words are activated at least to a lemma level). We conducted the experiments reported in this chapter in order to confirm that this was the case.

Participants named pictures/words in three contexts: the *control* context, in which the picture was presented in isolation; the *match* context, in which the picture name formed the high-cloze target of the sentence-stem which preceded it; the *mismatch (rime-overlap)* context, in which the picture name was semantically unrelated to the high-cloze target of the sentence-stem which preceded it. The mismatch context involves rime-overlap with the target picture name; this allows us to investigate whether phonological similarity between targets and distractors does not elicits strategic processing such that the classic semantic congruency effect is no longer observable. If participants activate a lexical representation of the high-cloze sentence-stem completion in a manner similar to that in studies described above, we would expect picture naming to be facilitated in the *match* condition as compared to the *mismatch (rime-overlap)* and the *control* conditions. Response latencies are the primary outcome measure of interest, but we might also expect by-condition differences in error rates: higher error rates in the *mismatch (rime-overlap)* condition than in the *match* and *control* conditions would suggest automatic and involuntary activation of high-cloze sentence-stem completions.

3.2. Method

3.2.1. Participants

10 adults (6 female, 4 male) with a mean age of 25 years (range, 18 – 40 years) participated in the online pre-test. A different cohort of 11 adults (8 female, 3 male) with a mean age of 19.6 years (range 18 – 24 years) participated in Experiment 2a. A further cohort of 22 adults (18 female, 4 male) with a mean age of 19.8 years (range 18 – 26 years) participated in Experiment 2b. Participants were students from the University of Edinburgh research participant pool. All participants spoke English as a first language, had normal or corrected-to-normal vision, and received course-credit for participation. No participant reported a positive history for speech, language, or hearing difficulties. All participants provided written consent in line with British Psychological Society guidelines. The study was granted ethical approval by the Psychology Research Ethics Committee of the University of Edinburgh.

3.2.2. Stimuli

Picture names: Eighteen monosyllabic concrete English nouns of medium frequency were selected for depiction (see Appendix C for a full word list). Each word had a rime pair within the stimuli list (e.g., can-tan), and all words began in either an alveolar or a velar onset in order to conform to requirements of planned follow-up articulatory imaging experiments. Pictures for naming were selected from a public-access internet database of colour photographs. Picture size was standardized to a height of 246 pixels (using the ImageMagick tool).

Prime sentences: For each of the 18 picture names, 3 high-cloze sentence-stems were generated in which the picture-name acted as the high cloze-probability target. All cloze sentence-stems were pre-tested via the online survey in order to determine cloze probabilities. Sentences were presented in the online survey in the same form that they were to be presented in the main experiment. Sentence-stems were recorded at a sampling rate of 44kHz in a sound attenuated recording booth, spoken by a female native speaker of English (mean speaking rate was 3.6 syllables per sec). Prime sentences were designed such that the penultimate word (i.e., the word preceding the high-cloze target) ended in either a vowel or, occasionally, a fricative. This allowed all sentences to be cut at the last steady state acoustic energy of the word preceding the cloze item, thereby eliminating any acoustic cues as to the phonetic onset of the high-cloze target word. Mean sentence length was 10 words, (SD 3). Mean sentence duration was 3.384 seconds (SD 0.984 seconds) (range; 0.972 – 5.991 seconds).

Online pilot norming study: As stated above, we conducted an online pilot study in order to determine: (i) cloze-probabilities for the experimental sentence-stems; (ii) naming agreement for the experimental images; (iii) potential competing onset words elicited by the experimental images. The

online survey was presented via Limesurvey (Carsten Schmitz, 2011). Participants, who did not take part in any of the studies detailed below, were asked to click on an icon to hear each sentence-stem and, after hearing each sentence-stem, to type the word that they thought to be the most likely the "missing word". Subsequently participants were presented with the experimental images to name, one at a time. Participants were instructed to type the single most suitable name for the picture in each case. The order in which sentence-stems and images were presented was randomized within block within the presentation software. Sentence stimuli were included in the main experiment only when cloze probability for the target picture name was .80 or greater (see Appendix C for a tabulated summary of cloze-sentences and naming outcomes). The pre-testing of materials was necessary because it was not possible to use prime sentences from established databases due to the word onset and syllable structure requirements detailed above. The online pre-test also allowed us to determine picture naming agreement and to analyse non-target picture name responses in order to ensure that no image used in the experiment generated a potentially confounding response (e.g., a picture of a "CAN" being named a "TIN" and thereby activating the onset phoneme of the target's rime pair, "TAN").

3.2.3. Procedure

3.2.3.1. Experiment 2a

Experiment presentation was automated using DMdX (Forster & Forster, 2003) and was delivered to participants via a Dell XPS 17 on a 17" screen at 1024x768 screen resolution. Audio was presented via over-ear headphones, and sound capture was performed via the PC-internal mic. Picture-naming response latencies were recorded automatically via the "digitalVOX" function in DMdX, set to trigger at -6dBA. Participants' full responses were recorded using the RecordVocal function, in addition to which the experimental session was separately audio-recorded using an H2 Handy Recorder (Zoom Corporation). At each stage of the experimental procedure, picture presentation order was randomised via DMdX. Participants took no more than 30 minutes to complete the experiment in full.

Participants first encountered each picture during the *familiarisation phase*, which consisted of two stages. In stage one, participants were instructed to "try to remember the picture names". For each item the target picture-name was presented auditorily, immediately after which the relevant picture was presented visually. The picture name was presented below the image in Calibri 40-point black font. In stage two of the learning-phase, participants were instructed to attempt to name each picture when they saw it, before its name appeared on screen. Participants were then presented with each picture in turn for naming. Pictures were presented for 2000ms without text and then for a further 1000ms accompanied by their target name. This allowed participants to confirm that they had named the picture correctly and, if not, presented a further opportunity to learn the picture name. The familiarisation-phase was followed immediately by the experimental phase.

In the first block of the *experimental phase*, participants were presented with each picture to name once "as quickly and accurately as you can". In blocks two and three participants were informed that they would hear a number of sentences, all of which would be missing the final word. Participants were informed that they would also see pictures, and that they should name any pictures that they saw "as quickly and accurately as possible". The stimulus onset asynchrony (SOA) between sentence-stem end and picture presentation was 0ms in all cases. Pictures were presented on screen for 2500ms. In the *match* sentential context the target picture name provided the high-probability cloze for the sentence-stem just heard. In the *mismatch (rime-overlap)* condition the target picture name was the rime-pair of the sentence-stem cloze (i.e. shared a syllable rime with that of the cloze word, but differed in onset. The context in which a sentence-stem was first encountered (*match* v. *mismatch*/ *rime-overlap*) was counterbalanced across participants, with each sentence-stem being presented once in block 2 and once in block 3. The two contexts were represented equally across the two blocks. Order of within-block presentation of items was randomised within DMdX. Each participant therefore named each picture 6 times in total across blocks two and three (3 times in a *match* context and 3 times in a *mismatch/rime-overlap* context). Block 4 was procedurally identical to block 1 (described above) and involved only picture naming. Blocks 1 and 4, in which pictures were named without a preceding sentence provide the baseline control condition against which performance in the sentential contexts was compared.

3.2.3.2. Experiment 2b

In Experiment 2b participants named words rather than pictures. They were therefore instructed to "say the word" rather than to "name the picture". The experimental procedure was otherwise identical to that in Experiment 2a, with the following exception: During the familiarisation phase, when in Experiment 2a the image was presented screen centre with the written word below it, in Experiment 2b the written word was presented screen centre with the image presented beneath.

3.3. Results

3.3.1. A general note on data analysis

Unless otherwise stated, response latency analyses presented in this thesis were conducted using a linear mixed-modelling approach, and were implemented in R 3.1.2 (R Core Team, 2014) via the lme4 package, version 0.999999-4 (Bates, Maechler, & Bolker, 2014). This approach provides estimated, rather than exact, effect sizes, meaning that it is not appropriate to calculate associated p-values exactly. We therefore treat $|t| > 2$ as indicating a statistically significant effect (see Baayen, 2008). Unless otherwise stated, we first constructed a maximal model that allowed slopes and intercepts to vary by participant and item (picture-name) where feasible (see Barr, Levy, Scheepers, & Tily, 2013). We then employed a backward stepwise approach to remove fixed effects that did not

contribute to model fit. The contribution of the remaining fixed effects to model fit was confirmed via forward stepwise testing from a basic model that included only random effects. Details of the confirmed best-fit model are reported in full in each case.

3.3.2. Experiment 2a Findings

3.3.2.1. Error data

Each participant produced 142 experimental picture names (36 in the *control* condition, and 53 in each of the *match* and *mismatch/ rime-overlap* conditions; 1 picture was unintentionally omitted from the experiment in both *mismatch*/*rime-overlap* and *match* conditions). Of the resulting 1562 trials, 27 (1.7%) were errorful responses (either the participant used a non-target name to name the picture or did not respond). Error-rates differed by experimental condition ($\chi^2_{(2)} = 12.73$, $p < .01$), being higher in the *rime-overlap* condition (3.55%) than in the *match* (0.87%) or control (0.51%) conditions. Errorful responses were excluded from further data analyses.

3.3.2.2. Outliers and missing data

Of the remaining 1535 trials, response latency was below 200 msec in 8.14% of cases, over 1499 msec in 0.85% of cases, and was not recorded due to mis-triggering of the voice-key in 7.67% of cases. The proportion of fast, slow, and mis-trigger trials did not differ by experimental condition ($\chi^2_{(2)} < 2.5$, $p > .05$ in all cases). Fast, slow and mis-trigger trials were excluded from further analyses. The remaining 1279 data points were included in the analyses described below.

3.3.2.3. Response latency data modelling

The outcome measure of interest was picture naming latency. Naming condition (*control*, *match*, *rime-overlap*) was experimentally manipulated and was the predictor of interest. We also included *Trial position* within the experiment and *lexical frequency* of the target picture name in the analyses.

The inclusion of *lexical frequency* did not significantly improve model fit compared to a null model ($\chi^2_{(1)} = 0.31$, $p > 0.5$). Lexical frequency was therefore not included as a predictor in subsequent models. The inclusion of *Trial position* significantly contributed to model fit ($\chi^2_{(1)} = 378.77$, $p < 0.001$), as did the further inclusion of *Context* ($\chi^2_{(2)} = 8.24$, $p < 0.05$) and the interaction between *Context* and *Trial position* ($\chi^2_{(2)} = 9.89$, $p < 0.01$). Further including the interaction between Context and lexical frequency did not improve model fit ($\chi^2_{(3)} = 2.12$, $p > 0.05$).The best-fit model therefore included as fixed effects *Context*, *Trial position*, and their interaction, and allowed slopes and intercepts for *Context* to vary by participants and items (i.e. target picture name).

Best fit model

Inspection of the best fit model indicated that response latencies reduced numerically over the course of the experiment (β = -0.27, se(β) = 0.14, t = -1.98). Response latencies in the *match* context were shorter than those in the *control* context (β = -101.34, se(β) = 22.92, t = -4.422). Response latencies in the *Mismatch/ rime-overlap* context were longer than those in the *control* context (β = 47.04, se(β) = 22.32, t = 2.107). The effect of *Context* reduced over the course of the experiment, as indicated by an interaction between *Context* and *Trial position*: Over the course of the experiment, response latencies in the *match* context decreased less than those in the *control* context (β = 0.65, se(β) = 0.21, t = 3.14), whereas response latencies in the *mismatch/rime overlap* context decreased by more (β = -0.44, se(β) = 0.21, t = -2.10). Full details of the best fit model are provided in Table 3.3.1.

Table 3.3.1: Experiment 2a summary of best-fit model for response latency data (effect stated in ms).

Fixed Effect	**Estimate**	**SE**	**t**	**Random Effect**		**Variance**
Intercept	625.26	24.15	25.89	Image	Intercept	272.9
Position	-0.27	0.14	-1.98		MvC	257.6
					RvC	181.3
Context				Participant	Intercept	3885.7
- Match v Control	-101.34	22.92	-4.42		MvC	414.1
- Rime overlap v Control	47.04	22.32	2.11		RvC	189.2
Position x Context						
- Match v Control	0.65	0.21	3.15			
- Rime overlap v Control	-0.44	021	-2.10			

In order to explore whether the duration of the sentential context had an effect on picture naming latency we modelled a subset of the data that included only *match* and *rime-overlap* contexts (i.e., only those contexts in which a sentence-stem preceded picture presentation). We compared a model that contained those predictors included in the best fit model described above to one that that additionally included sentence-stem duration as a predictor. Due to the reduction in data quantity we simplified the random effects structure to allow the model to converge. This was achieved by

removing the by-item random slopes for context. The inclusion of sentence-stem duration contributed significantly to model fit ($\chi^2_{(1)}$ = 108.66, p < 0.01), with response times decreasing as sentence stem duration increased.

3.3.3. Experiment 2b Findings

3.3.3.1. Error data

Each participant produced 144 experimental picture names (36 in the *control* condition, and 54 in each of the *match* and *mismatch/ rime-overlap* conditions). Of the resulting 3168 trials, 26 (0.8 %) were errorful responses (either the participant used a non-target name to name the picture or did not respond). As in Experiment 2a, error-rates differed by experimental condition, ($\chi^2_{(1)}$ = 6.32, p < .05), being higher in the *rime-overlap* condition (18/1188; 1.5%) than in the *match* (5/1188; 0.4%). Only 3 errors were produced in the Control context (representing an error rate of 0.4%). Errorful responses were excluded from further data analyses.

3.3.3.2. Outliers and missing data

Of the remaining 3098 trials, response latency was below 100 msec in 11% of cases and was over 1499 msec in 0.4%. The proportion of fast and slow trials did not differ by experimental context ($\chi^2_{(2)}$ < 3, p > .05 in all cases). Fast and slow trials were excluded from further analyses. We did not identify any mis-trigger trials. The remaining 2730 data points were included in the analyses described below.

3.3.3.3. Response latency data modelling

As in Exp 2a, the outcome measure of interest was naming response latency (in this case being word naming response latency). Naming context (*control*, *match*, *mismatch*/*rime-overlap*) was experimentally manipulated and was the predictor of interest. We also included *Trial position* and *lexical frequency* of the target picture name as predictors in the analyses.

The inclusion of *lexical frequency* marginally improved model fit compared to a null model ($\chi^2_{(1)}$ = 3.3708, p = 0.07). We therefore retained this predictor in the model, in line with the recommendations of Barr and colleagues (2013). The inclusion of *trial position* did not further improve model fit as assessed by a forward step-wise comparison ($\chi^2_{(1)}$ = 1.77, p = 0.18). However, a subsequent backward stepwise comparison from the maximal model indicated an effect for *trial position*; it was therefore retained in the final best-fit model. The inclusion of *context* ($\chi^2_{(2)}$ = 9.57, p < 0.01) further improved model fit, as did the inclusion of the interaction between *context* and *trial position* ($\chi^2_{(3)}$ = 18.88, p < 0.01). Model fit was not improved by including the interaction between *context* and *lexical frequency* ($\chi^2_{(2)}$ = 0.8331, p > 0.5). The best-fit model therefore included as fixed

effects *lexical frequency*, *trial position*, *context*, and the interaction between *context* and *trial position*, and allowed slopes and intercept to vary by participants and items (i.e. target word name) for *context*.

Best fit model

Inspection of the best fit model indicated that response latencies were lower for higher frequency lexical items (β = -12.46, se(β) = 6.42, t = -1.94). Response latencies in the *match* context were shorter than those in the *control* context (β = -29.44, se(β) = 13.76, t = -2.140). Response latencies in the *mismatch/rime-overlap* context were longer than those in the *control* context (β = 60.07, se(β) = 13.68, t = 4.39). As in Experiment 2a, the effect of *context* reduced over the course of the experiment, as indicated by an interaction between *context* and *trial position*: As in Experiment 2a, over the course of the experiment response latencies in the *mismatch*/*rime overlap* context decreased more than those in the *control* context (β = -0.46, se(β) = 0.12, t = -3.72). Numerically, response latencies in the *match* context decreased over the course of the experiment by less than those in the *control* condition. However, unlike in Experiment 2a, this difference was not statistically significant (β = 0.15, se(β) = 0.12, t = 1.20). Full details of the best fit model are provided in Table 3.3.2.

Table 3.3.2: Experiment 2b summary of best-fit model for response latency data acquired in Experiment 2b (effects are stated in msecs)

Fixed Effect	**Estimate**	**SE**	**t**	**Random Effect**		**Variance**
Intercept	513.54	23.87	21.51	Image	Intercept	75.36
Position	-0.21	0.08	-2.58		MvC	40.87
Lexical frequency	-12.46	6.42	-1.94		RvC	14.82
Context				Participant	Intercept	2593.13
- Match v Control	-29.44	13.76	-2.14		MvC	402.08
- Rime overlap v Control	60.07	13.67	4.39		RvC	522.21
Position x Context						
- Match v Control	0.15	0.12	1.20			
- Rime overlap v Control	-0.46	0.12	-3.72			

In order to explore whether the duration of the sentential context had an effect on picture naming latency we modelled a subset of the data that included only *match* and *mismatch/ rime-overlap*

contexts (as for Experiment 2a). We compared a model that additionally included *sentence-stem duration* as a predictor to one that contained only those predictors included in the best fit model described above. The inclusion of *sentence-stem duration* did not contribute to model fit ($\chi^2_{(1)} = 0.94$, $p > 0.1$).

3.4. Discussion

We developed a variant of the sentential context PWI paradigm which allows us to test for phonological-phonetic level speech production effects of prediction during comprehension. We tested whether typical sentential PWI effects are observable within this adapted paradigm. Across two experiments we tested whether the sentential priming effect is present for (a) picture targets and (b) written targets. For both picture and written targets, naming was facilitated when targets matched the high-cloze anticipated sentence-fragment completion (compared to naming in isolation). This result is in line with the findings of previous studies (e.g., Roe et al., 2000) and indicates that the sentence-fragments used in our experiments elicited lexical level activation of anticipated high-cloze completions.

Picture and word name targets matched the high-cloze completion in only 50% of trials, meaning that it was not strategically advantageous to participants to predict the high-cloze completion. As we nonetheless observed facilitation, it appears that high-cloze completions are automatically activated when sentence-fragments are presented auditorily (as when sentence-fragments are presented visually; Stanovich & West, 1979; Griffin & Bock, 1998; Gollan et al., 2011; see Introduction). This interpretation is in keeping with the fact that error-rates were significantly higher when the sentence-fragment completion mismatched the picture/word name taget than when the two matched or the target was named in isolation: The higher error rates appear to be the consequence of the intrusion of involuntarily activated sentence-fragment completions during the picture/word naming production process.

Overall, response latencies decreased over the course of the experiment in both experiments. This is reassuring as it suggests that the experimental procedure engaged participants' attention throughout. An exploration of the interaction between trial position and context reveals that response latencies for both picture and word naming in the *match* context did not decrease over the course of the experiment: It appears that the degree of pre-activation elicited by the high-cloze context may have caused naming latencies in the *match* context to be at floor level. In the *control* condition, latencies to written words were similarly unaffected by trial position (Experiment 2b), whereas picture naming latencies decreased (Experiment 2a). This may reflect the different processes involved in picture and word naming; whilst picture naming requires conceptual and semantic activation, word naming can be achieved via grapheme-phoneme conversion (e.g., Coltheart & Rastle, 1994), and the orthographic representations involved are argued to have priviliged access to articulation (e.g., Costa, Alario, & Caramazza, 2005). Word naming in the *control* context may have been at floor level (for

non pre-activated items) at the beginning of the experiment: Participants would have had many years' experience of linking the written form of our lexically relatively frequent items to the spoken form, whereas links between the specific picture form and the spoken form had to be established within the experimental session. Under this interpretation, the word naming facilitation observed in the *match* context must be understood to arise of preactivation prior to presentation of the reponse stimulus.

For both picture and word naming, response latencies in the *mismatch* context decreased more over the course of the experiment than did those in the *control* context. This is consistent with previous reports that when a mismatch between sentential context and the item to be named elicits inhibition, the effect diminishes over the course of the experiment (e.g., Roe et al., 2000). This interaction may be understood to arise due to the nature of the paradigm either causing participants to strategically adjust their strategies, or to (automatically) update their expectations (see Introduction). Both interpretations assume that participants learn to predict the unpredicted. The experiment exposes participants to incongruent combinations far more frequently than would be encountered in everyday language use; the attenutation of the inhibition effect may reflect participants' becoming accustomed to this. Alternatively, the effect be explained at an attentional level: As participants become more familiar with the speaker and the sentential material over the course of the experiment, the sentential context becomes more predictable allowing greater attentional deployment to the picture/word naming process (see Introduction for reference to Fischler & Bloom, 1979): This interpretation is in keeping with the finding that naming latencies decreased as sentential context duration increased, and that this effect did not interact with context. The experiments reported in this chapter do not distinguish between potential interpretations of the inhibition effect. Further, it may be that the interpretations are not orthogonal but rather reflect the fact that the paradigm has been deployed under a variety of theoretical frameworks.

In general, it appears that sentential PWI effects are not impacted by adaptations that make the paradigm suitable for the investigation of phonological-phonetic level representations. However, we did not observe an interaction between lexical frequency and sentential constraint such as that reported by others and suggested to reflect word-form retrieval effects (e.g., Griffin & Bock, 1998; Piai et al., 2014). Within the context of our study, an interaction of this type would take the form of a lexical frequency effect being greater in the *control* context (no sentential constraint) than in the *match* context (high and appropriate sentential constraint). Our items were deliberately of relatively equal lexical frequency, in order to avoid potential masking effects of target and distractor being of substantially different lexical frequencies. It is therefore unsurprising that we did not observe such an interaction, as the lexical frequency range employed in our study did not elicit a notable lexical frequency effect even in the *control* condition. The lack of a lexical frequency effect across the items employed in our study is reassuring: It eliminates the concern that effects arising of prediction might be masked by lexical frequency effects allowing some item-names preferential access to articulation regardless of their role (target or distractor; see Miozzo & Caramazza, 2005, for a discussion of distractor lexical frequency effects).

In addition to confirming that the paradigm was suitable to investigate effects at a phonological-phonetic level, the experiments reported in this chapter allowed us to compare the relative suitability of picture and word targets. As in previous studies (see Appendix B for details), both picture and word naming were facilitated in congruent contexts and to some degree inhibited in incongruent context. Although findings for word and picture naming generally patterned together, there were some differences: How the interaction between trial position and context was expressed differed between the experiments; this highlighted the fact that semantic processing is not necessary for word naming (see Brysbaert, Fias, & Reynvoet, 2006, for a review of evidence that this is the case). For word naming only, lexical frequency marginally improved model fit; our intention in selecting items of relatively equal lexical frequency had been to avoid this (for rationale see above). These differences cause us to prefer picture naming over word naming as an outcome measure, as the focus of this thesis is on prediction as it occurs during spoken language processing as opposed to text processing. Within the paradigm we employ, picture naming appears to offer increased potential to observe the effects of top-down activation of phonological representations whilst being less susceptible to potentially confounding effects.

Chapter 4: Are listener-generated predictions specified at a speech-sound level?[3]

4.1. Introduction

4.1.1. Abstract to paper

Comprehension of spoken language involves the prediction by the listener of upcoming material. It has been demonstrated that listener-generated predictions of upcoming material can be specified to a phonological level, such that a specific word-onset is anticipated (e.g., DeLong, Kutas, & Urbach, 2005). The current study investigated whether such word-form specific predictions impact picture-naming latencies in a manner similar to that observed when a distractor word is actually presented. Participants were auditorily presented with high-cloze sentence-stems, in order to elicit word-form predictions. Pictures for naming were presented immediately following the sentence-stem. We systematically manipulated the phonological relationship between the predicted word and the picture name. Across three experiments, naming was facilitated when the picture name fully matched the predicted word. However, naming was neither facilitated nor inhibited when the picture name overlapped phonologically with the predicted word. This finding is in contrast to effects of phonological overlap observed when a distractor word is heard or read. Our findings suggest that words which are internally listener-generated (predicted) during comprehension are not robustly specified at a speech sound level.

[3] This chapter constitutes an extended version of a published paper ("Drake, E., & Corley, M. (2015a). Effects in production of word pre-activation during listening: Are listener-generated predictions specified at a speech-sound level?. *Memory & Cognition*, *43*(1), 111-120."). In the current chapter data were analysed in R R 3.1.2 (R Core Team, 2014) via the lme4 package, version 0.999999-4 (Bates, Maechler, & Bolker, 2014), in line with all other analyses reported in this thesis. There are therefore some numerical inconsistencies with data published in the paper; crucially, results do not differ in statistical significance or direction. This chapter includes an extended analysis and discussion, to incorporate findings concerning the effect of mismatch at a lexical level when there is match at a word form level. This material was not included in the original manuscript submitted for review; it is of interest within the context of this thesis but was outwith the remit of the paper.

4.1.2. Main Introduction to paper

The speech production system is active during comprehension. Perceiving speech primes the speech motor system even when no speech output is required. Articulatory muscles are activated when listening to speech sounds but not when listening to non-speech sounds, and such increased excitability of the motor system is accompanied by an increase in activity within Broca's area (Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Watkins, Strafella, & Paus, 2003; see also Pulvermüller et al., 2006). Activation of motor-speech areas during speech comprehension may reflect backward- or forward-looking processes, or both. The production system may be engaged in generating articulatory representations to support the maintenance and decoding of heard material; equally, it may be engaged in simulating upcoming auditory input via the generation of emulations (Watkins & Paus, 2004; Pickering & Garrod, 2004). Recently, there has been an increased focus on the latter possibility. It has influentially been suggested that during comprehension the production system is engaged in generating predictions of upcoming material, thereby reducing processing demands on the comprehension system, by constraining possible interpretations of incoming material (e.g., Pickering & Garrod, 2007; Schiller, Horemans, Ganushchak & Koester, 2009; see also Scott, McGettigan, & Eisner, 2009 for a review). But whereas evidence of speech-motor activation during comprehension is compatible with the notion that the listener's speech-motor system is engaged in generating predictions, it does not constitute proof. In order to confirm such an interpretation, it would be necessary to demonstrate that upcoming material in specific is represented via the speech production system.

Although it has not been empirically demonstrated that predictions during comprehension are achieved via the speech-motor areas, it has been demonstrated that words are predicted at a surface-form level during comprehension (DeLong, Kutas, & Urbach, 2005). When reading sentences which strongly predict a noun (such as *The day was breezy so the boy went outside to fly...*), comprehenders exhibit increased N400 amplitudes upon encountering an indefinite article in a form inappropriate to the predicted noun (*an*, where the prediction is *kite*). The amplitude of this response is correlated with the probability that the predicted noun completes the sentence, as determined by previous offline testing. This effect can relate only to the upcoming word's being specified at a phonological-form level, because the distinction between *a* versus *an* is empty at a semantic and syntactic level, and is based purely on the phonological form of the upcoming word (consonant versus vowel).

In the current study we investigate whether effects of phonological-form prediction in comprehension are observable in a behavioural measure of speech production. If prediction does involve the generation of motor emulations, we would expect to see an effect of prediction in the listener's own speech production system. Previous findings from picture-word-interference (PWI), picture-picture interference (PPI) and sentence-listening paradigms provide some guidance as to the possible nature of such an effect (see Chapters 1 and 2 of this thesis).

Picture-naming is facilitated when the picture is accompanied by a partially-phonologically-overlapping written distractor word (as compared to one with no overlap; Damian & Dumay, 2007; Lupker, 1982). When the distractor word is presented auditorily, picture naming is facilitated only when there is onset-overlap (Meyer & Schriefers, 1991). Overall, pictures are named more slowly in the presence of a distractor word than in isolation (Meyer & Schriefers, 1991). This pattern of findings may result from the production system being habitually and automatically recruited during comprehension, causing presentation of a distractor word to increase the demands on the production system, and thereby leading to a general increase in picture-naming latencies (see, Greene, 1988; Raney, 1993). If so, the phonological facilitation effects described above may be better understood as attenuated inhibition effects, where the inhibition is attenuated as a consequence of the overlap between competing representations in the production system.

In typical PWI experiments the distractor is orthographically represented. In studies in which the distractor is a picture, participants must internally generate the lexical form of the distractor, and effects of phonological overlap are found in some studies (e.g., Morsella & Miozzo, 2002; Navarette & Costa, 2005), but not others (Jescheniak et al., 2009). This difference has recently been attributed to aspects of the distractor pictures used (Opperman, Jescheniak, & Görges 2014), rendering comparison with distractors which are (implicitly) predicted during comprehension difficult. However, one study in which target (rather than distractor) names were internally generated provides evidence that such word forms can be phonologically specified in a way that affects speech production latencies (Humphreys, Boyd & Watter, 2010). In a free-association version of the picture-word interference paradigm, each written word had a single, high-likelihood associate (e.g., "cobweb" -> spider). Participants named the first word that came to mind as an associate of the written word, while ignoring the picture. In phonologically-related trials, the associate and the picture-name shared an onset (e.g., "cobweb" → *sp*ider; *SP*OON), whereas in unrelated trials there was no such phonological overlap (e.g., "cobweb" → *sp*ider; *F*ORK). Response latencies were significantly shorter in the phonologically-related condition than in the unrelated condition, and did not differ from those in the control condition[4] (in which participants did not see any picture). This suggests that there is no need for a word to be perceptually available in order to elicit effects at phonological level, and strengthens the case for suggesting that the locus of any facilitation is in the production system. If prediction during comprehension is production-driven, then we might expect words predicted during comprehension to elicit similar phonological effects to those of pictures implicitly named during viewing.

Previous studies concerning the effect of sentence-stem context on picture naming have generally investigated integration rather than prediction effects, and have therefore focused on the effects of manipulating the semantic and/ or syntactic congruence between the sentence-stem and the picture

4 The statistic for related versus control conditions was not reported, and this conclusion is drawn from Table 1 and Fig. 2.

name (e.g., Roe, Jahn-Samilo, Juarez, Mickel, Royer, & Bates, 2000; see also Griffin & Bock, 1998; Wicha et al. 2005; see Chapter 3 and Appendix B for further details). Pictures are named fastest in a congruent context, more slowly in a neutral context, and most slowly in an incongruent context. This pattern has been interpreted as indicating (prediction-mediated) easing of integration (Griffin & Bock, 1998; Wicha et al. 2005). Of course, the effect is also consistent with a prediction-as-production account. This interpretation merits further exploration, particularly in light of evidence that inhibition of picture-naming may arise from conflict at the production level, rather than from integration costs (Hirschfeld, Jansma, Bölte, & Zwitserlood, 2008; Mahon, Costa, Peterson, Vargas, & Caramazza, 2007; Nozari , Dell, & Schwartz, 2011; Severens et al., 2011).

The possibility that sentence context may elicit phonological representations through a process of prediction-as-production is suggested by the findings of an error-elicitation study (Ferreira & Griffin, 2003): High-cloze sentence-stems were presented orthographically to participants as a means of semantic priming. Immediately following sentence-stem presentation, a picture was presented for naming. Participants were more likely to (erroneously) utter the high-cloze completion in place of the picture name when the high-cloze completion was a semantic competitor of the picture name (e.g., nun – PRIEST) than when the two were unrelated (e.g., nun – HAND). Crucially, with respect to the activation of phonological representations within the speech production system, participants were also more likely to erroneously utter the sentence completion when it was a *homophone* of a semantic competitor of the target picture (e.g., none – PRIEST). In the current study we investigate whether such evidence of interactivity between lemma and word-form levels within the speech production system during comprehension reflects preactivation at a speech-sound level.

In summary, the speech production system is active during comprehension, and comprehension can involve prediction of word-forms at a phonological level. These findings have been linked in the suggestion that prediction during comprehension engages the speech production system. The present study was designed to explore this suggestion directly. In three experiments, participants heard auditory sentence fragments with highly predictable continuations (such as *He managed to fix the drip from the old leaky…*) and named pictures at the offset of the audio. Picture-names were chosen such that they corresponded to the predictable word (tap-*TAP*) or had a partial phonological overlap (tap-*CAP*, tap-*TAN*), or had no overlap (tap-*CONE*). In order to maximise the probability of phonological effects, each mismatching sentence-continuation corresponded to an image name on other trials (cf. Meyer & Damian, 2007). We predicted facilitation of picture-naming in the matching condition (compared to an acontextual control condition), as evidence that participants were making predictions as a consequence of hearing the sentence stems. Of interest was whether there was an effect of phonological overlap, which would confirm that during speech comprehension, predicted items are activated at a phonological level in the production system. We investigated the effects of two types of phonological overlap: segment overlap, such as that associated with the "phonological facilitation effect" typically reported in PWI studies (see Chapter 2), and whole word-form overlap, such as that associated with homophone mediated lexical mis-selection (e.g., Ferreira & Griffin, 2003, see above).

This allowed us to consider the nature of phonological representations that might be active within the speech production system during comprehension-elicited prediction.

4.2. Method

4.2.1. Participants

Twenty-seven adults (10 male) with a mean age of 19 years (range 18–24) participated in Experiment 3a. Twenty-one further adults (7 male) with a mean age of 20 years (range 18–26) participated in Experiment 3b. Finally, a further twenty-one adults (7 male) with a mean age of 20 years (range 18-27) participated in Experiment 3c. Participants were students from the University of Edinburgh, who either received course credit or were paid for participation. One participant in each of Experiments 3a and 3b identified themselves as multilingual subsequent to the recordings; data from these participants was excluded from the analyses. All remaining participants were monolingual speakers of English. No participant reported relevant language or visual impairments. Written consent was provided, in line with British Psychological Society guidelines.

4.2.2. Stimuli

We used identical sets of sentence-stems and of pictures across the three experiments. Sentence-stems were chosen such that they strongly predicted the following word; depending on the condition in which they were encountered, pictures had names which either corresponded to the predictable word, overlapped with it phonologically, or were unrelated. The experimental items comprised twelve of the pictures used in the experiments described in Chapter 3. We reduced the experimental items from the 18 used in the previous experiments in order to: (i) avoid the experiment being so lengthy that participants would begin to feel fatigue and/ or exhibit attentional drift; (ii) ensure that phonological overlap was as extensive in word initial position as it was in word final position (i.e., 2 segments in each context; see Appendix C for details). A further 12 pictures acted as filler items.

4.2.2.1. Filler items

In order to protect against strategic response bias arising of experimental items being encountered in a sententially unpredicted (incongruent) context twice as often as they were encountered in a sententially supported (congruent) context, we introduced filler items that were encountered in a supportive context only, thereby balancing the supportive and non-supportive contexts encountered within participants across the experiment. This allowed us to investigate the possibility that the sentential context in which an item has recently been encountered leads to an updating of the lexical activation of that item.

Target names for filler pictures were monosyllabic and did not overlap with experimental picture names at syllable onset or rime. Filler items were included in order to minimize participants' conscious attention to phonological aspects of the picture names, and to maintain an even balance between trials in which pictures corresponded to predicted or unpredicted names. Filler items were selected from cloze normed sentences reported by Block and Baldwin (2010). The cloze probability of all filler sentences was .8 or greater, all cloze items were imageable and target cloze items were preferred picture names in British English (the original sentences reported in Block and & Baldwin were tested on an American-English speaking population). All filler target cloze words were of a CVC structure in order not to be distinguishable in this respect from experimental items.

4.2.2.2. Experimental items

Experimental pictures had been pre-tested online for name-agreement (see Chapter 2 and Appendix C: median agreement = .8; mean = .7) and their names were all monosyllabic concrete nouns of medium frequency (mean log10CD = 2.93, SD = 0.41, range = 2.07-3.91; SUBTLEX-US database, Brysbaert & New, 2009). There were 36 experimental sentence-stems, 3 of which predicted the name of each of the 12 experimental pictures, as determined by pre-test (see Chapter 3 and Appendix C: all cloze-likelihoods > .8). Similarly, 3 of 36 filler sentence-stems predicted the name of each of the 12 filler pictures. Sentence-stems were recorded by a female native speaker of British English in a sound attenuated room at the University of Edinburgh. Mean speech rate of the experimental sentence-stems was 3.9 syllables per second, and mean duration was 3.2 seconds (see Appendix C for full details). In experimental trials, we manipulated the phonological relationship between the word predicted by the sentence-stem, and the accompanying picture presented for naming. In Experiment 3a, we included matching (e.g., tap-*TAP*), onset-overlap (e.g., tap-*TAN*), and rime-overlap (e.g., tap-*CAP*) conditions. We compared naming latencies in these conditions to those in a control condition where the picture was named following backwards speech (maintaining the 'speech-like' qualities of the sentence-stems but ensuring that there were no linguistic cues). In Experiment 3b we replaced the onset-overlap condition with a no-overlap condition (e.g., tap-*CONE*). In Experiment 3c we replaced the rime-overlap condition of Experiment 3b with an onset-overlap condition, allowing for direct comparison of the onset-overlap and no-overlap conditions; response times differed between these two conditions in the word association study of Humphreys et al. (2010; see above for details). In each experiment, all experimental pictures occurred in all conditions for all participants, allowing each picture to act as its own control in a fully within-participants and within-items design.

4.2.3. Procedure

As in Experiments 1 and 2, the stimuli in all experiments were presented using DMdX (Forster & Forster, 2003). Each experiment began with a familiarisation phase, during which participants saw each of the 24 experimental and filler pictures accompanied by its printed name, and named the picture aloud. Each picture and corresponding name appeared three times in total. In each experiment

the familiarisation procedure was followed by 5 experimental blocks. Blocks 1 and 5 were included principally in provided a control condition comparable to that used in Experiments 1 and 2 (see previous chapters). For the purposes of the current study these blocks allowed us to confirm that participants were able to correctly name all pictures when these were presented in isolation. Each picture was presented on its own, and named aloud. Participants viewed a fixation point in the centre of the screen for 2.9 seconds (which was the mean duration of the sentence-stem recordings used in other blocks), immediately prior to the presentation of the picture-to-be-named. Participants were instructed to name each picture as quickly and accurately as possible (as they had practised during the familiarisation phase). In blocks 2, 3, and 4 participants again viewed a fixation point, but this time while listening to a sentence-stem. The picture-to-be-named was presented immediately at the offset of the last word of the sentence-stem. Participants were again instructed to name pictures as quickly and accurately as they could. In each of blocks 2, 3, and 4 each experimental picture was presented 4 times; once in each experimental condition and once in the backward-speech control condition. Each sentence-stem was heard by each participant once per block, and once per condition across the three blocks. We manipulated the condition in which a picture was encountered by altering the sentence-stem with which it was presented. Item presentation within each block was fully randomised so that all conditions were interleaved with one another and with filler items.

Picture-naming response latencies were automatically recorded, together with full auditory responses, by the experimental software. We also made an independent audio-recording of each full session. Participants took no more than 50 minutes to complete an experiment.

4.3. Results

Analyses were conducted using the lme4 package, version 1.1-7, in R 3.1.2 (Bates, Maechler, Bolker, Walker, Christensen, Singmann, & Dai, 2014; R Core Team, 2014). As throughout this thesis, for response times we used linear mixed-effects models fit by maximum likelihood; error responses were analysed using binomial mixed-effects models, fit using Laplace estimation. Response time models included *log trial position* as a predictor to account for practice effects throughout the experiment. Error models did not include this predictor as, due to the low number of errors produced, models including *log trial position* did not converge. In each case we included the effects of *context* (matching, rime-overlap, onset-overlap [Experiment 3a], matching, rime-overlap, no-overlap [Experiment 3b], matching, onset-overlap, no-overlap [Experiment 3c], backward speech) on the response variable of interest. The *context* predictor was orthogonally coded, as detailed below for each experiment. Following suggestions made by Barr, Levy, Scheepers, and Tily (2013), each model was 'maximally specified', with both intercepts and slopes, as well as their correlations, allowed to vary by participants and, where possible, by items. In linear analyses we treated *t* value of 2.00 or above as significant, due to complexities in estimating the degrees of freedom associated with predictors (Baayen, 2008).

4.3.1. Experiment 3a Findings

In Experiment 3a there was a total of 4368 recorded experimental responses (including 624 responses in the simple picture naming context and 938 in each of the experimental contexts), of which 70 (1.6%) were errors, either because of lexical intrusion of the word predicted by the context, or because of other factors (such as failure to record a response). Table 4.3.1 provides a summary of error numbers by condition. We first modelled likelihood of producing an error, using a subset of the data which included all experimental trials but not the control trials. Errorful trials were then removed from the data, leaving 4298 data points, allowing us to perform an analysis of response latencies to correct picture names. In each analysis we used orthogonal contrasts for the context predictor, such that we first compared observations for the match condition to those for all other conditions; second, we compared the two conditions in which a word different to the picture was predicted by the context (mismatch conditions) to the backward control condition; and third, we compared the two mismatch conditions (rime-overlap vs. onset-overlap).

4.3.1.1. Error data 3a

To avoid empty cells, we analysed total error numbers, rather than contextual errors alone. The model did not include a random effect of context by items, due to a failure of a model containing this specification to converge. Including a fixed effect of context significantly improved model fit ($\chi^2_{(3)}$ = 27.5, $p < .001$). The coefficients showed that there were significantly fewer errors in the match conditions than in the other three conditions combined (β = -0.32, SE(β) = 0.12, z = -2.72, $p < .01$), and that there were more errors in the two overlap conditions than in the backward control condition (β = 0.41, SE(β) = 0.13, z = 3.17, $p < .01$). There was no difference between the two overlap conditions ($z < 1$). Table 4.3.3 gives details of the model coefficients.

Table 4.3.1: Recorded errors in Experiments 3a-c: Figures refer to total errors/numbers of errors in which distractor was produced in error (percentages in brackets).

Exp	**Context**					
	Control	Match	Rime overlap	Onset overlap	No overlap	Backward
3a	5/-- (0.8%)	5/-- (0.5%)	23/18 (2.5%/1.9%)	29/24 (3.1%/2.6%)	--	8/-- (0.5%)
3b	3/-- (0.6%)	4/-- (0.6%)	23/20 (3.1%/2.8%)	--	16/14 (2.2%/1.9%)	8/-- (1.1%)
3c	1/-- (0.2%)	1/-- (0.1%)	--	18/17 (2.7%/2.6%)	12/10 (1.8%/1.5%)	6/-- (0.9%)

Table 4.3.2: Fast and slow responses excluded from further analyses in Experiments 3: Figures indicate raw number of responses that were below 100ms ('fast') or above 2499ms ('slow'), and (in brackets) percentages of non-error data points in each condition.

	Context											
Exp	Control		Match		Rime overlap		Onset overlap		No overlap		Backward	
	fast	slow	fast	slow	fast	slow	fast	slow	fast	slow	fast	slow
3a	48 (7.7%)	2 (0.3%)	79 (8.4%)	1 (0.1%)	76 (8.3%)	0 (0%)	89 (9.6%)	0 (0%)	--	--	88 (9.4%)	2 (0.2%)
3b	8 (1.7%)	1 (0.2%)	17 (2.4%)	7 (1.0%)	6 (0.9%)	6 (0.9%)	-- --	-- --	14 (1.9%)	12 (1.7%)	12 (1.7%)	8 (1.1%)
3c	43 (0.9%)	0 (0%)	61 (0.9%)	0 (0%)	-- --	-- --	63 (0.9%)	0 (0%)	65 (1%)	0 (0%)	63 (0.9%)	0 (0%)

4.3.1.2. Response latency data 3a

The model of response times included *log trial position* and *context* as predictors, as well as fixed and random intercepts, and the random effects of log trial position and context by-participants, and of context by-items. The fixed effect of context significantly improved model fit ($\chi^2_{(3)} = 44.7$, $p < .001$). Model coefficients showed that participants' responses became faster as the experiment progressed ($\beta = -11.7$, SE(β) = 4.8, $t = 2.42$), and that participants were faster to correctly name pictures in the match context compared to the other three contexts ($\beta = -18.2$, SE(β) = 2, $t = -9.15$). There were no differences within the three non-matching contexts ($ts < 1$). Table 4.3.4 gives details of the model coefficients.

4.3.2. Experiment 3b Findings

In Experiment 3b, twenty participants produced 3360 responses (480 in the original control context, and 720 in each of the experimental contexts). Fifty-four (1.6%) (51) of these responses contained errors. Errors are summarised in Table 4.3.1. In analyses of response times and of error rates, we used orthogonal contrasts similar to those for Experiment 3a, such that the matching condition was compared to all other sentential contexts, the two contexts which elicited mismatching predictions were compared to the backward control contexts, and finally, these two contexts (rime-overlap vs. no-overlap) were compared to each other.

4.3.2.1. Error data 3b

For Experiment 3b we again analysed total error numbers. As for Experiment 3a, the analysis did not include trial number; nor did it include a random effect of context by items or participants, due to a failure to converge. Model fit was significantly improved by the addition of a fixed effect of context ($\chi^2_{(3)} = 18.147$, $p < .001$). Inspection of the coefficients revealed that there were fewer errors in the match condition than in the other conditions ($\beta = -0.33$, SE(β) = 0.13, $z = -2.47$, $p < .05$), and that there were more errors in the mismatch contexts than in the backward control context ($\beta = 0.30$, SE(β) = 0.13, $z = 2.28$, $p < .05$). There was no difference between the rime-overlap and no-overlap contexts ($z = 1.14$). See Table 4.3.3 for the model coefficients.

4.3.2.2. Response latency data 3b

Other than the differences in contexts, details of the response latency model construction were identical to that for Experiment 3a. The fixed effect of context again improved model fit ($\chi^2_{(3)} = 27.4$, $p < .001$). Participants again responded faster as the experiment progressed ($\beta = -15.8$, SE(β) = 4.0, $t = 3.96$), and were fastest to respond in the match condition compared to the other conditions ($\beta = -19.7$, SE(β) = 2.4, $t = 8.33$). There were no differences within the three non-matching conditions (ts < 1.29). The model coefficients are given in Table 4.3.4.

4.3.3. Experiment 3c Findings

One participant in Experiment 3c made 22 errors early in the experiment, and was excluded from all analyses. The remaining 19 participants made 37 errors over 2736 responses (1.4%). Errors are summarised in Table 4.3.1. For our analyses we once again used orthogonal contrasts, this time comparing the matching context to all others, the two mismatching contexts to the backward control, and finally, the mismatching contexts (onset-overlap vs. no-overlap) to each other.

4.3.3.1. Error data 3c

Other than the differences in contexts, the models were constructed in the same way as for the two previous experiments. An analysis of total error numbers did not include either trial number or a random effect of context by items. Adding a fixed effect of context significantly improved model fit ($\chi^2_{(3)} = 20.9$, $p < .001$). There was a statistically significant difference between the match context and the other contexts combined, with the match context containing fewer errors ($\beta = -0.603$, SE(β) = 0.25, $z = -2.370$). The two mismatching conditions resulted in more errors than the backward control ($\beta = 0.31$, SE(β) = 0.15, $z = 2.02$). The onset-overlap and no-overlap conditions did not differ (z = 1.1).

4.3.3.2. Response latency data 3c

When response latencies were modelled, once again the fixed effect of context improved model fit ($\chi^2_{(3)} = 20.3$, $p < .001$). Participants were quicker to respond as the experiment progressed ($\beta = -15.5$, SE(β) = 5.0, $t = 3.13$), and responded fastest in the match condition ($\beta = -14.1$, SE(β) = 2.3, $t = 6.06$). No other differences were significant (ts < 1; see Table 4.3.4 for full model details).

Table 4.3.3: Experiments 3a-c model coefficients (in logits) for the likelihood of producing an error

Fixed Effect	Estimate	SE	z	Random Effect		Variance
Experiment 3a						
Intercept	-4.77	0.32	-15.04	Participant	Intercept	0.77
Context					Context M v O	--
- Match vs. Others	-0.32	0.12	-2.72		Context MM v BC	--
- Mismatch vs. Backward Control	0.41	0.13	3.17		Context R v OO	--
- Rime vs. Onset Overlap	-0.12	0.14	-0.86	Item	Intercept	0.26
Experiment 3b						
Intercept	-4.53	0.41	-12.82	Participant	Intercept	0.69
Context					Context M v O	--
- Match vs. Others	-0.33	0.13	-2.47		Context MM v BC	--
- Mismatch vs. Backward Control	0.30	0.13	2.28		Context R v NO	--
- Rime vs. No Overlap	0.19	0.17	1.14	Item	Intercept	0.00
Experiment 3c						
Intercept	-4.97	0.37	-13.45	Participant	Intercept	0.54
Context					Context M v O	--
- Match vs. Others	-0.60	0.25	-2.37		Context MM v BC	--
- Mismatch vs. Backward Control	0.31	0.15	2.02		Context R v NO	--
- Onset Overlap vs. No Overlap	0.21	0.19	1.11	Item	Intercept	0.00

Table 4.3.4: Experiments 3a-cModel coefficients (in ms) for naming latencies

Fixed Effect	Estimate	SE	t	Random Effect		Variance
Experiment 3a						
Intercept	600.20	22.57	26.59	Participant	Intercept	7373.3
log(trial pos^n)	-11.66	4.82	-2.42		log(trial pos^n)	320.4
Context					Context M v O	17.0
- Match vs. Others	-18.19	1.99	-9.15		Context MM v BC	167.5
- Mismatch vs. Backward Control	-0.29	3.89	-0.07		Context R v OO	193.2
- Rime vs. Onset Overlap	-5.59	7.44	-0.75	Item	Intercept	232.9
					Context M v O	35.2
					Context MM v BC	24.2
					Context R v OO	334.6
Observations = 3344				Residual		33080.0
Experiment 3b						
Intercept	732.45	14.09	51.99	Participant	Intercept	1759.2
log(trial pos^n)	-15.66	4.39	-3.57		log(trial pos^n)	240.5
Context					Context M v O	33.5
- Match vs. Others	-19.69	2.35	-8.36		Context MM v BC	82.3
- Mismatch vs. Backward Control	1.57	3.70	0.43		Context R v NO	293.5
- Rime vs. No Overlap	-8.56	6.70	-1.28	Item	Intercept	766.9
					Context M v O	20.4
					Context MM v BC	59.1
					Context R v NO	203.7
Observations = 2746				Residual		17905.5
Experiment 3c						
Intercept	593.59	29.71	19.98	Participant	Intercept	13232.5
log(trial pos^n)	-15.53	4.96	-3.13		log(trial pos^n)	289.5
Context					Context M v O	46.4
- Match vs. Others	-14.15	2.33	-6.07		Context MM v BC	182.9
- Mismatch vs. Backward Control	-1.03	4.03	-0.26		Context OO v NO	31.1
- Onset Overlap vs. No Overlap	-2.44	4.42	-0.55	Item	Intercept	132.2
					Context M v O	1.9
					Context MM v BC	11.7
					Context OO v NO	4.5
Observations = 2447				Residual		20918.4

4.3.4. Lexical (homophone) Analyses

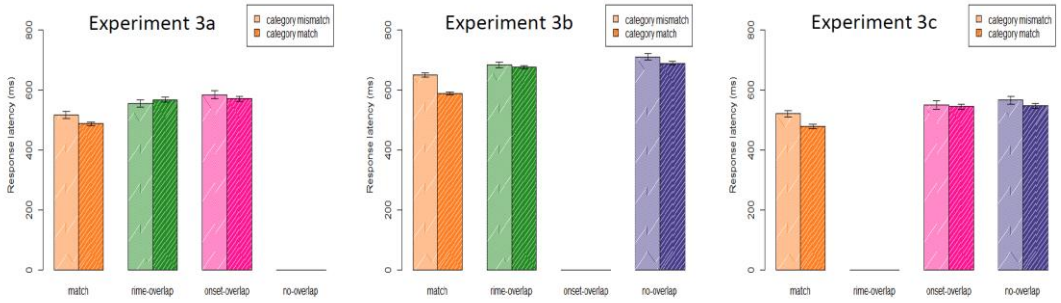

Figure 4.3.1: Experiments 3a-c response latencies to pictures, reported by category (category mismatch = homophone in full-overlap condition, category match = full match in full-overlap condition)

The primary manipulation of interest across the experiments reported in this chapter was, as stated in the Introduction, the nature of the phonological overlap between the predicted word and the target picture name noun. However, some sentence-stems predicted homophones of picture name target nouns rather than the nouns themselves (e.g., *"There's no such word as can't. You have to believe that you…"* CAN; "*They raised their glasses in a…*" TOAST). In such cases the predicted item mismatches the picture-to-be-named at a lexical/word-class level (forming a homophone) even where there is full phonological overlap between predicted item and the picture name (i.e., in the match context).

Given the controversy within the literature concerning the precise circumstances of homophone activation during spoken language comprehension and production (e.g., see Rose, Spalek & Abdel Rahman, 2015, for a recent brief review), we wished to confirm that the effects reported above were observed across both homophone and full-match sentence-fragment types. The ensuing analysis is exploratory, in that it was not planned into the experimental design. It can, however, provide some insight as to whether facilitation observed in the match condition might be explained solely at a lemma level, or whether phonological level activation contributes to the facilitatory effect observed in the match context: If the facilitation effect occurs only for full lexical match sentence-fragment types, we might assume that that only semantically appropriate pictures can benefit from comprehension-elicited prediction. Conversely, if facilitation effects are observable for homophones, we can remain confident in our assumption (detailed above) that predictions are active in some way at a word-form level.

We therefore performed supplementary analyses to investigate the effect of lexical mismatch. These analyses included data acquired in the phonological overlap contexts only (i.e., the backward context was excluded as we assume it could not elicit lexical predictions). For each sentence-stem, we determined whether the high-cloze target had the same meaning as the picture it was encountered with in the match condition (e.g., *"The fire alarm's gone off again; someone must have burnt the …"* TOAST; referred to as "full match") or whether it shared only phonological form (e.g., *"They raised their glasses in a …"* TOAST; referred to as "homophone"). Where the sentence fragment predicted the picture name lemma in the match condition, the trial was considered to be a "full match"; where only the phonological form matched, the trial was considered to be a "homophone". Of the 36 experimental sentence-fragments, 11 (31%) fell in the homophone condition[5].

We would expect any effect of category (mis)match (full match v. homophone) to be observable only in the full phonological overlap context: The full overlap context is the only context in which the relationship between the predicted item and the target picture name differs according to sentence-stem category (full match/ homophone). This is because only the full overlap context allows the predicted item to match the picture name at a lexical level. We retained all contexts in the homophone analyses reported below. This allows us to address whether any effect observed is a property of the relationship between the item predicted by the sentence-fragment and the picture, as opposed to some other property of the sentence-fragments themselves. If the former is the case, an effect would be observable only in the full overlap context, whereas if the latter is the case we would expect to observe the effect across contexts.

We modelled the effect of category (mis)match (full match v. homophone) on response latencies following the principles described above (see Results). Slopes and intercept for Context were allowed to vary by participant for Experiments 3a, 3b and 3c, and by item for Experiments 3b and 3c (allowing the slope to vary by item for Experiment 3a caused the model not to converge). Including trial position as a predictor led to model non-convergence in a number of cases: Trial position was therefore not included in any of the models reported below (for effects of Trial position see main analyses above). For each experiment we modelled the effects on response latency of phonological Context, Category (full match v. homophone), and their interaction. We report details of the best-fit model in each case (see Table 4.3.5. for full details; see Figure 4.3.1 for plots of condition means).

4.3.4.1. Experiment 3a homophone findings

As above, the fixed effect of *Context* improved model fit over a null model ($\chi^2_{(2)} = 36.8$, $p < .001$); participants were quicker to respond in the Match context than the Rime-overlap and Onset-overlap

[5] Three of these represented not only a lemma mismatch but also a syntactic mismatch, in that the picture name was a noun and the sentence-fragment predicted a verb. This contrast is of theoretical interest but the numbers involved in this study were too low to justify a statistical analysis. Previous studies suggest that syntactic congruency interacts with semantic congruency (e.g., Bentrovato et al., 2003; see also Appendix B)

contexts (β = - 24.4, SE(β) = 2.9, *t* = -8.29). Again, as expected from the results presented above, response latencies did not differ significantly between the two partial overlap contexts (Rime-overlap and Onset-overlap: (β = -5.9, SE(β) = 5.3, *t* = -1.1). Including *Category* (i.e., full match v. homophone) and its interaction with *Context* did not further improve model fit ($\chi^2_{(1)}$ = 1.2, *p* > 0.1; $\chi^2_{(3)}$ = 4.9, *p* > 0.1 respectively).

4.3.4.2. Experiment 3b homophone findings

The fixed effects of *Context*, *Category* (full-match v. homophone), and their interaction each contributed to model fit (respectively, $\chi^2_{(2)}$ = 30.0, p < .001; $\chi^2_{(1)}$ = 9.1, p < .01; $\chi^2_{(2)}$ = 10.1; < .01 in all cases). As in previous analyses, participants were quicker to respond in the Match context than in the Rime-overlap and No-overlap contexts (β = - 16.6, SE(β) = 4.8, t = -3.4). Under the current analysis, participants were quicker to respond in the Rime-overlap context than in the No-overlap context (β = -19.1, SE(β) = 9.2, t = -2.1). Overall, participants were quicker to respond in the Match category than the Mismatch category (β = -23.5, SE(β) = 7.4, t = -3.2). The effect of *Category* was greater in the Match context than other contexts (β = -14.7, SE(β) = 4.8, t = -3.0), and did not differ between the Rime and No-overlap contexts (β = 15.3, SE(β) = 9.5, t = 1.6). See Figure 2.3.1 for a plot. This suggests that, in cases where the picture name fully matched the prediction at a phonological level, picture naming latencies were shorter when the sentence fragment predicted the picture name as a lexical item rather than simply a homophone. For example, the image TOAST would be responded to more quickly following the sentence fragment *"The fire alarm's gone off again; someone must have burnt the …",* than following the sentence fragment *"They raised their glasses in a …"*.

4.3.4.3. Experiment 3c homophone findings

As for Experiments 3b, the fixed effects of *Context* and *Category* (full-match v. homophone) contributed to model fit (respectively $\chi^2_{(2)}$ = 19.4, *p* < .001; $\chi^2_{(1)}$ = 7.5; *p* < .01 in both cases). The interaction of *Context* and *Category* marginally contributed to model fit ($\chi^2_{(2)}$ = 5.7, *p* = 0.06) and was therefore retained in the best-fit model reported in Table 4.3.5. Participants were quicker to respond in the Match context than in the Onset-overlap and No-overlap contexts (β = - 11.5, SE(β) = 4.9, *t* = -2.4). Response latencies did not differ significantly between the Onset- and No-overlap contexts (β = -11.6, SE(β) = 7.7, *t* = -1.5). Overall, participants were quicker to respond for sentence-stems that in the Match context predicted the same lexical item as the picture name (β = -21.8, SE(β) = 7.7, *t* = -2.8). The effect of Category (full-match v. homophone) was greater in the Match Context than other contexts (β = -10.6, SE(β) = 5.2, *t* = -2.0), and did not differ between the Onset- and No-overlap contexts (β = 12.2, SE(β) = 9.1, *t* = 1.3).

The results above indicate that naming facilitation seen in the Match (i.e., full phonological-overlap) context was reduced when the target picture name Category did not match that predicted by the sentence-stem. That is, the response latency facilitation effect seen for picture names that

phonologically matched the predicted item was reduced for homophones compared to words that fully matched the prediction at a lexical as well as a phonological level. In order to confirm that the Context effect remained even when Category did not match (i.e., for homophone responses), we modelled response latency for each Category condition separately (full-match and homophone). Data obtained in each experiment were subsetted by the Category condition in which they were obtained (full-match/ homophone). For each subset a model was constructed in which Context acted as a fixed predictor. Intercepts were allowed to vary by participant and item; slopes were allowed to vary by participant. In each case, including Context as a fixed predictor improved model fit over a null model (in all cases $\chi^2_{(2)} > 6.5$, $p < 0.05$). In all cases participants responded faster in the full-overlap context than in any other context, with no other between context differences being statistically significant (see Table 4.3.6 for full details of models). This finding that the facilitation effect is present for homophones suggests that the experimental paradigm did elicit activation of predicted items at a word-form level, if not a phonological segment level, and that the picture does not need to be semantically congruent with the sentence-fragment in order for facilitation to be observed.

Table 4.3.5: Experiments 3a-c model coefficients (in ms) for naming latencies (sentence contexts only)

Fixed Effect	Estimate	SE	t	Random Effect		Variance
Experiment 3a						
Intercept	545.11	11.27	48.37	Participant	Intercept	2485.6
Cloze.match	NA	NA	NA			
Context					Context M v O	52.15
- Match vs. Others	-24.36	2.94	-8.29		Context RO v OO	215.44
- Rime vs. Onset Overlap	-5.88	5.35	-1.10			
				Item	Intercept	215.80
Observations = 2506				Residual		33500.8
Experiment 3b						
Intercept	677.29	17.87	37.90	Participant	Intercept	4560.1
Cloze.match	-23.53	7.378	-3.19			
Context						
- Match vs. Others	-16.62	4.83	-3.44		Context M v O	103.1
- Rime vs. No Overlap	-19.10	9.23	-2.07		Context RO v NO	282.7
Cloze.match*Context	-8.56	6.70	-1.28	Item	Intercept	667.5
- * Match vs. Others	-14.70	4.843	-3.03		Context M v O	26.8
- * Rime vs. No Overlap	15.33	9.53	1.61		Context R v N	154.7
Observations = 2054				Residual		18853.0
Experiment 3c						
Intercept	538.06	29.71	19.98	Participant	Intercept	8638.0
Cloze.match	-21.81	4.96	-3.13			
Context						
- Match vs. Others	-11.48	4.87	-2.36		Context M v O	98.1
- Onset v No Overlap	-11.64	7.68	-1.52		Context OO v NO	34.6
Cloze.match*Context				Item	Intercept	72.9
- * Match vs. Others	-10.57	5.19	-2.04		Context M v O	19.1
- * Onset vs. No Overlap	12.23	9.15	1.34		Context OO v NO	4.5
Observations = 1832				Residual		20854.0

Table 4.3.6: Experiments 3a-c model coefficients (in ms) for naming latencies (sentence contexts only, data sub-setted by category type; full match and homophone)

Fixed Effect	Estimate	SE	t	Random Effect		Variance
Exp 3a – full match						
Intercept	541.87	11.20	48.4	Participant	Intercept	2492.9
Context				Item	Intercept	132.9
- Match vs. Others	-27.87	3.18	-8.62	Participant	Context M v O	22.37
- Rime vs. Onset Overlap	-1.48	7.35	-0.21		Context RO v OO	63.88
Observations = 1743				Residual		31694.1
Exp 3a – homophone						
Intercept	550.74	13.65	40.35	Participant	Intercept	3003.5
Context				Item	Intercept	247.3
- Match vs. Others	-18.40	5.55	-3.31	Participant	Context M v O	131.0
- Rime vs. Onset Overlap	-13.90	9.88	-1.41		Context RO v OO	247.3
Observations = 763				Residual		36352.7
Exp 3b - full match						
Intercept	652.64	17.87	39.14	Participant	Intercept	4287.8
Context				Item	Intercept	608.1
- Match vs. Others	-30.98	4.83	-8.17	Participant	Context M v O	156.8
- Rime vs. No Overlap	-3.57	5.00	-0.71		Context RO v NO	88.4
Observations = 1432				Residual		18041.1
Exp 3b - homophone						
Intercept	679.08	18.58	36.55	Participant	Intercept	4928.1
Context				Item	Intercept	714.7
- Match vs. Others	-17.60	4.67	-3.77	Participant	Context M v O	19.9
- Rime vs. No Overlap	-13.44	12.18	-1.10		Context R v N	1359.6
Observations = 622				Residual		21033.2
Exp 3c – full match						
Intercept	516.12	21.29	24.25	Participant	Intercept	8239.9
Context				Item	Intercept	39.8
- Match vs. Others	-22.46	3.15	-7.14	Participant	Context M v O	41.5
- Onset vs. No Overlap	0.55	5.45	0.10		Context OO v NO	112.4
Observations = 1266				Residual		19624.1
Exp 3c - homophone						
Intercept				Participant	Intercept	9418.7
Context				Item	Intercept	268.7
- Match vs. Others	-11.51	5.66	-2.03	Participant	Context M v O	200.6
- Onset vs. No Overlap	-12.65	8.35	-1.51		Context OO v NO	36.7
Observations = 566				Residual		2355380

4.3.5. Filler v Experimental Item Analyses

Across the three experiments, filler items appeared in only the match and control contexts, whereas experimental items appeared in contexts in which the target picture name was semantically incongruent with respect to the context in which it was presented. When filler items were predicted by the sentence-stem they were always presented for naming. When experimental items were predicted by the sentence-stem they were presented for naming in one-third of cases. If mismatch between a predicted item and item presented for naming affected future lexical activation levels during prediction or naming on subsequent trials containing the same stimuli we would expect to see a reduced effect of trial number on experimental stimuli (encountered in all contexts) compared to filler stimuli (encountered only in predicted or neutral contexts). We tested for this using the subset of the data from Experiments 3a, 3b and 3c that included only the contexts in which both filler and experimental items appeared (i.e., the acontextual picture naming and full phonological overlap contexts). We modelled response latency as a function of trial number, item type (filler v. experimental), and their interaction. Model coefficients indicated again that, overall, participants became quicker to name pictures as the experiment progressed (β = - 0.15, SE(β) = 0.06, t = -2.62). Participants were generally quicker to name filler items than to name experimental items (β = -27.4, SE(β) = 9.1, t = -3.01). It is likely that this difference arose due to item properties. This interpretation is supported by the fact that the interaction between trial position and item type indicated that the effect of trial position was smaller for filler items than for experimental items (β = 0.16, SE(β) = 0.06, t = 2.82).. It appears then that in the current experimental paradigm mismatch between a predicted item and a presented item did influence subsequent prediction strength in an item-specific manner.

4.4. Discussion

In three experiments, we showed that the time to name a picture after hearing a sentence fragment which strongly predicted a given distractor word was faster if the picture name coincided exactly with the distractor than in any other condition. Counter to expectations, there were no response latency differences between the remaining conditions: Naming times did not differ whether the predicted word overlapped at the onset with the picture name, or at the rime, or where there was no overlap at all between distractor and picture name. Moreover, the naming latencies for these conditions were the same as for a backward control condition in which no specific word could have been predicted from the auditory context. These findings stand in contrast to picture-word interference paradigms in which a distractor is explicitly presented, either in writing (Damian & Dumay, 2007; Lupker, 1982; Chapter 2 of the current thesis), auditorily (Meyer & Schriefers, 1991), or pictorially (Humphreys et al., 2010). In each of these cases, phonological overlap has been shown to facilitate picture-naming relative to no-overlap conditions. An initial interpretation of the present findings is therefore that, although

upcoming material is predicted during comprehension, any involvement of the production system stops short of a speech sound level of representation.

A potential objection to this interpretation is that the present experiment may not have induced participants to predict specific words at all: Instead, the faster naming of pictures which happened to match predictable words may have been due to the ease with which those picture names could be semantically integrated with the preceding context. In cases of phonological overlap, there would still be a semantic mismatch, and therefore relative difficulty of integration. However, this explanation seems unlikely, because each picture was named several times throughout each of the present experiments. Since the picture names were also used as the words predicted by sentence fragments, it is highly likely that specific words would have been activated at each trial, due to extensive repetition. In line with this suggestion, the majority of recorded errors (60% and 62% in Experiments 3a and 3b respectively) were cases where the words predicted by the sentence fragments were accidentally produced in lieu of the picture names. The fact that distractor names were overtly produced suggests that these names were fully specified at the lexical, and hence the phonological, level. This interpretation is further supported by the findings that: (i) the naming of homophones of predicted lexical items was facilitated, and; (ii) participants were significantly more errorful in the phonological overlap conditions than in other conditions, suggesting an effect of phonological representation at a monitoring level (see Severens et al., 2008). Moreover, the delay in naming pictures in the backward control condition relative to the match condition militates against a purely integration-based account, since in this condition there is no semantic context with which the picture name can be integrated.

Given that distractor words are predicted, the question arises of whether the naming latency differences between the match and other conditions are due to inhibition of mismatching picture names, or to facilitation of names that match. Of note is the fact that participants made relatively few errors in the backward-speech control condition, despite the fact that the associated response latencies patterned with the various overlapping conditions. This suggests that, at least in the control condition, specific words were not predicted and could not have led to interference. Moreover, in the experiments reported in this chapter, words predicted in the no-overlap condition were as vulnerable to semantic inhibition as were words in the partial phonological overlap conditions (i.e., rime-overlap and onset-overlap). When vulnerability to inhibition was kept constant in this way, partial phonological overlap between the distractor and the picture name did not affect response times, ruling out a suggestion that inhibition and facilitation effects may have been masking each other in the experiments reported in Chapter 3. Taking these considerations together, it seems that facilitation, but not inhibition is implicated in the present experiments; and, importantly, that facilitation only occurs when a word which is predicted exactly matches at a form level the name to be produced for a picture.

4.4.1. General implications

Over three experiments, we found no evidence to suggest that comprehension-associated predictions gave rise to inhibition within the production system, although we did find evidence of a facilitatory effect. However, this effect does not appear to extend to phonological representations accessible specifically to the speech-production system at segment level: Causing participants to produce words which overlap phonologically with words predictable through comprehension does not give rise to facilitation of production, unless the overlap is complete (and therefore the picture-name matches the predicted word at levels other than the phonological segment). On the assumption that phonological assembly precedes articulation in production, this further implies that the speech-motor system is not involved in making specific predictions.

One way to reconcile these findings with evidence that the speech-motor system is activated during comprehension may be to suggest that this activity is associated with performance rather than with content (e.g., Rothermich, & Kotz, 2013; Scott, McGettigan, & Eisner, 2009). According to such a view, presentation of an auditory sentence-stem would automatically engage prediction processes. One aspect of prediction would be estimation of the timing of the speaker's production, and preparation to respond (related to the so-called "how" pathway), reflected in speech-motor activity. Predictions at the lexical-phonological level (the "what" pathway) would not rely on the speech-motor system (see Hickok, 2012). In the context of the present experiments, the auditory contexts would consistently alert listeners to the likely moment at which a picture might appear. The facilitation effect seen in the full-overlap condition would relate to the "what" channel, and would not involve motor-speech activation.

This interpretation is not inconsistent with the conclusion drawn in a previous study, that phonological-form expectations are generated during the comprehension of high-cloze sentences (DeLong et al., 2005). It is in no way implicit in that conclusion that such phonological expectations would be accessible to the speech-motor system. The way our findings differ from those of DeLong and colleagues recalls the suggestion that phonological representations incorporate separable levels of representation (e.g., Hickok, 2012; Goldrick & Rapp, 2007): Predictions of another's speech may be specified at a phonological level that is driven by lexical or auditory representations, but not necessarily implemented at a phonological-articulatory level. Of note in this respect is the finding that phonological word-form effects arising of sentential constraint are associated with central processing costs, whereas phonological overlap effects observed in a conventional PWI paradigm are not (Ferreira & Pashler, 2002). Ferreira and Pashler argue that this difference indicates that phonological overlap effects arise at an execution stage, whereas word-form effects arise at a selection stage. Under that interpretation, the current finding that comprehension-elicited prediction facilitates the naming of homophones but not partially overlapping phonological forms suggests that predictions are selected at an abstract lexical level but are not represented at a speech-motor level.

If auditory sentence-stems do enable prediction via a "how" pathway, the question remains as to why picture-naming latencies were not relatively delayed in the backward-speech control condition: Backward speech may not offer sufficient cues at either a semantic prosodic level from which timing can be estimated (Brown et al., 2012; Londei et al., 2010). Of course, failure to find an effect cannot be interpreted as evidence that there is no such effect. However, this pattern of findings does raise the question of whether speech-motor activity would be recorded in this condition, and if not, whether the benefits of "how" prediction might be relatively small, at least in the context of the present rather artificially constrained task.

Of course, the possibility remains that response latencies do not reliably reflect activation levels through the production system to articulation (although such a suggestion would effectively undermine a substantial body of work). If, however, we align ourselves with the picture-word interference literature and accept that the time taken to name a picture is likely to be influenced by pre-activation of relevant phonological representations, the present experiments strongly suggest that "what" prediction during comprehension does not appear to occur at a phonological-articulatory level, and thus that the speech-motor activation associated with language comprehension does not reflect a detailed prediction of upcoming content.

Chapter 5: Exploring speech as action: An ultrasound imaging approach

5.1. Introduction

In the first part of this thesis we adopted a response latency approach in order to address the question of whether predictions elicited during comprehension are represented within the listener's speech-motor system. We perceptually classified responses as "correct" or "errorful", with only correct responses being included in the response latency analyses. Results suggested that items predicted during comprehension are not represented at a speech-motor level. However, we note that vocal response latencies do not provide a means by which to directly observe motor speech behaviour, and perceptual categorisation of responses as correct or errorful does not capture fine grained variation in motor-speech execution (see Gafos & Goldstein, 2012 for a review of how the "choice of observables" has impacted theoretical conclusions). Such variation in motor speech execution, observable via articulatory imaging, has been shown to systematically reflect the activation of higher-level representations (e.g., Davidson, 2005; McMillan & Corley, 2010; Pouplier, 2007). In the following chapters (Chapters 6 and 7) we report studies in which we employed ultrasound tongue imaging to capture and analyse speech as an action. This provided a means to observe directly if, and how, predictions elicited during comprehension impact motor speech execution.

In this chapter we briefly outline the limitations associated with the use of speech reaction time data to address topics concerning speech as action. We review electromyographic evidence concerning motor activity observed to occur during the response latency period in manual response studies. We consider the difficulties that arise when EMG methodology is applied to speech, as opposed to manual, motor activity, and we review alternative articulatory imaging approaches that have been adopted in order to explore the relationship between cognition and action during speech production. We consider how the findings of articulatory imaging studies might relate to claims concerning the involvement of the speech production system in prediction during comprehension. We then introduce the ultrasound imaging and analysis approach that we employ in this thesis in order to investigate the involvement of the speech production system in prediction during comprehension.

5.2. Motor activity during response making

Voice-key response latency records the point at which the sound level associated with a vocal response exceeds a pre-determined threshold, as opposed to the onset of articulation. It therefore

records an acoustic rather than an articulatory event. The acoustic event that acts as a voice key trigger is preceded by innervation of muscles involved in producing speech (for discussion and evidence see Kessler, Treiman, & Mullennix, 2002; Rastle, Croot, Harrington, & Coltheart, 2005; Rastle & Davis, 2002). Acoustic response latency therefore does not provide direct information about speech as an action (although acoustic material following voice key triggering may provide indirect information concerning articulation; e.g., Kello, Plaut, & MacWhinney, 2000; Damian, 2003). Under an assumption that motor variability during speech production is "motor noise", this lack of information about speech as an action is unproblematic. However, studies of manual response latencies have revealed that motor activity during response making can relate systematically to the psychological and/or environmental conditions under which a response is made (e.g., Hasbroucq, Possamaï, Bonnet, & Vidal, 1999; Coles et al., 1985; Eriksen et al., 1985; Smid, Mulder, & Mulder, 1990). If this is also the case for speech responses, voice-key methodology such as that employed in this thesis up to this point may not be an optimal approach to investigate whether the speech motor system is implicated in prediction during comprehension.

The use of thumb muscle electromyography (EMG) has allowed manual response latencies recorded in button push experiments to be fractionated into an initial, purely cognitive, phase (a "pre-motor response time") and a subsequent phase during which muscle activity is observable (a "motor response time"; Boulinguez, Jaffard, Granjon, & Benraiss, 2008; Tandonnet, Burle, Vidal, & Hasbroucq, 2004; see also Klapp, 1996, for discussion of response time fractionation into central and peripheral components). Fractionation of manual response times has revealed that the motor response component is longer under conditions where there is conflict between potential responses than in conditions where there is no potential conflict. This is the case even when by-condition differences in stimulus presentation are not strictly relevant to the response required of the participant, for example in Simon and Eriksen flanker tasks (Simon, 1990; see also Lu & Proctor, 1995; for EMG evidence see Hasbroucq, Possamaï, Bonnet, & Vidal, 1999; Coles et al., 1985; Eriksen et al., 1985; Smid, Mulder, & Mulder, 1990). This finding confirms that variability in motor execution during response making is not simply "noise", but can relate systematically to the psychological conditions under which the response is made.

In addition to revealing systematic variability in the duration of motor response making, EMG studies indicate differences in the nature of motor response making. During the motor response period, it is sometimes possible to observe EMG activity associated with a response that would be incorrect under the circumstances. Such activity may be associated with an overt error (i.e., depressing the wrong button), but it can also be observed in cases where the overt response is correct. Such responses are referred to as "incorrect-correct" responses. Response latencies in trials involving "incorrect-correct" are typically longer than those in trials involving "correct-correct" responses. Incorrect-correct responses are more commonly observed in incongruous trials than congruous trials, suggesting that conflict between a stimulus dimension and a response requirement expresses itself at a motor level.

Findings of the type described above have led to the conclusion that incompatibility between an irrelevant stimulus dimension and a response dimension slows stimulus evaluation and evokes response competition (Fournier, Scheffers, Coles, Adamson, & Abad, 1997). Longer pre-motor response times are thought to arise due to slowed stimulus evaluation, whereas longer motor response times are thought to reflect response competition between two simultaneously active response channels, observable in EMG activity. The fact that response competition is observable in motor output has been central to the proposal that, during human information processing, information can cascade from one processing level to another in a continuous manner (for a review see Van 't Ent, 2002). This suggests that if we are to investigate the nature of motor representations during speech processing via a conflict-inducing approach, it may be necessary to investigate motor activity during response making. If anticipated input is represented within the speech motor system during comprehension, there will potentially be response conflict when the picture-to-be-named is incongruous with the sentential context. Evidence of conflict at an articulatory level would support the proposal that prediction during comprehension involves the production system at a speech-motor level.

An attempt has previously been made to apply an EMG approach in order to fractionate speech reaction times into pre-motor and motor elements (Riès, Burle, & Alario, 2012). It is considerably more complex to apply this technique to the investigation of speech responses than to the investigation of manual thumb push responses: Thumb press measures are taken from the flexor pollicis brevis muscle. This muscle is directly involved in the thumb flexion process required to achieve the button press. EMG signals from the facial muscles involved in lip movement during speech production can be captured but, as Riès and colleagues (2012) note, these muscles are not necessarily those involved in producing a vocal response as captured via a voice key. The relationship between EMG activity and voice key triggering is therefore considerably less direct than that between EMG activity and button push triggering. Further complexities are introduced by the fact that it is particularly difficult for participants to attain the relaxation of facial muscles prior to stimulus presentation necessary for successful EMG fractionation of the response time. Interpretation of lip movement data is also more complex that of thumb movement data; each thumb has its own musculature, whereas lip movement associated with the production of differing sounds shares a muscle system within which muscle fibres are interdigitated (see Mowrey & MacKay, 1990, for examples and details).

Although techniques to allow fractionation of speech response times are in their infancy, informational cascade during speech production has been investigated using alternative approaches. Early evidence of informational cascade from phonological to articulatory levels was obtained through the use of intra-muscular EMG to examine muscle fibre motor-unit activation during production of error-eliciting tongue twisters (Mowrey & MacKay, 1988; Mowrey & MacKay, 1990). This allowed the researchers to investigate whether speech activity contains traces of conflict at a cognitive level comparable to that revealed by the "incorrect correct" button push responses described above. Due to the invasive nature of the intra-muscular recording technique, the researchers themselves acted as the

two participants. EMG records indicated anomalous activity in some perceptually correct productions (analogous to the "incorrect-correct" push button responses described above). This behaviour was noted particularly when there existed competition between two potential responses (in this case, competition between active representations of the speech sounds /s/ and /ʃ/, as in the tongue twister "She sells sea-shells on the sea shore"). Anomalous activity was gradient in nature, falling within a range between that observed in typical ("canonical") tokens and that observed in perceptually errorful tokens. In line with the manual EMG data described above, this finding indicates that, when environmental or psychological conditions cause multiple representations to be simultaneously active, this activation can cascade to a motor execution level. In the case of tongue-twisters, the simultaneously active representations are generally considered to be abstract segments specified at a phonological level: It has been proposed that such cascading from higher to lower levels during speech processing is a form of forward modelling of the type proposed under the Pickering and Garrod model (e.g., Dell, 2013; see Chapter 1 for further details on forward modelling and the comprehension-as-production account).

5.3. Speech imaging approaches and findings

The invasive EMG methodology employed by Mowrey and MacKay does not lend itself to use with typical participant groups. However, cascade during speech production has been further explored by coupling the use of error elicitation techniques with speech imaging approaches that allow dynamic recording of the location of the tongue during speech production. Imaging techniques employed include electropalatography (EPG; McMillan, Corley, & Lickley, 2009), electromagnetic articulography (EMA, e.g., Pouplier, 2003; Pouplier & Goldstein, 2010; Slis & Van Lieshout, 2013) and ultrasound speech imaging (sometimes referred to as ultrasound tongue imaging; UTI; McMillan & Corley, 2010). All three approaches provide information about tongue location over time; this information is central to an understanding of speech production as the tongue is the primary active articulator involved in speech production. EPG records contact between the tongue surface and the hard-palate at a temporal resolution of up to 100 Hz. EMA records tongue and lip flesh point location at a temporal resolution of up to 400 Hz (see Kim, Lammert, Ghosh, & Narayanan, 2014, for details). Ultrasound imaging typically records a mid-sagittal sector of the oral cavity at a temporal resolution of ~100 Hz: the image obtained can provide an estimate of the location of the tongue-surface within the oral cavity (see below for further details). EMA and EPG are partially-invasive techniques that require the instruments of data collection to be within the speaker's oral cavity throughout the course of data collection. Ultrasound imaging is non-invasive; in this respect the ultrasound data capture process is less likely to interfere with typical speech production processes (e.g., kinaesthetic feedback) than are EMA and EPG approaches

Competition during speech planning has been experimentally invoked in articulatory imaging studies through the use of tongue twister, word order competition (WOC) and "Spoonerisms of laboratory

induced Predisposition" (SLIP) paradigms. In keeping with their use of error elicitation paradigms, the articulatory imaging studies referred to above have been primarily concerned with the way speech is realised rather than the period that elapses between stimulus presentation and acoustic response. The use of articulatory imaging approaches allowed these studies to investigate whether competition within the speech production system leaves a "trace" in speech motor execution. Two measures obtainable from EMA, EPG and UTI data that have been particularly prominent in addressing questions concerning such cascade are: (i) the location of maximal constrictions (i.e., consonant realisations) at a given time-point defined from the acoustic record (e.g., Goldstein et al., 2007; Pouplier, 2003), and; (ii) the degree of global similarity between different tokens and a canonical token (e.g., McMillan, Lickley, & Corley, 2009; McMillan & Corley, 2010). The term "canonical token" in this context refers to a perceptually error-free token produced in a competition free environment; for example, a token produced in response to a single picture or word presented for naming in isolation or for repeated iteration (e.g., "cop cop cop cop" as opposed to "cop top cop top"; see Pouplier, 2007; McMillan & Corley, 2010).

Tongue twister experiments employing speech imaging techniques have provided evidence that when alternating word onsets compete during speech production, competing and target onsets gestures can be co-produced (for articulatory evidence see Boucher, 1994; Goldstein, Pouplier, Chen, Saltzman, & Byrd, 2007; McMillan & Corley, 2010; Mowrey & MacKay, 1990; Pouplier, 2008; for acoustic evidence concerning voicing as opposed to manner or place of articulation see; Frisch & Wright, 2002; Goldrick & Blumstein, 2006; McMillan & Corley, 2010). Maximal constriction measures taken via EMA indicate that, when alveolar and velar gestures compete (for example in the tongue twister "cop top cop top"), co-productions involve initial simultaneous movement toward closure at both target and competitor location (i.e., for both /k/ and /t/). In tokens that are perceptually categorised as "correct", movement toward the competitor location is attenuated compared to movement toward the target (e.g., Pouplier & Goldstein, 2010). This attenuated movement is not typically identifiable within the acoustic speech signal: although gestures are coproduced at movement onset (with synchronous onset of the overlapping gestures confirming that the coproduction originates in planning rather than monitoring), the gesture production develops into a typical target release phase (Pouplier & Goldstein, 2010). In keeping with this finding, global similarity measures obtained via ultrasound recording during tongue twister production indicate that traces of the competing phoneme cause perceptually correct tokens produced in competing conditions to be more similar to the competing phoneme than those produced in non-competing conditions (e.g., McMillan, Corley, & Lickley, 2009).

Co-productions have similarly been observed in WOC and SLIP task studies (Pouplier, 2007b; McMillan et al., 2009). Both WOC and SLIP tasks exploit expectancy in order to evoke coactivation of competing word onsets. Unlike tongue twisters, SLIP and WOC tasks do not rely on multiple iterations of target sounds to elicit errors. This militates against a purely articulatory-level explanation of any coproduction phenomena observed. In the WOC task, participants are required to name pairs of

(non)words that have just been presented orthographically on screen (e.g., "gope doof"). (Non)word pair presentation is followed immediately by the presentation of an arrow onscreen, which is the cue to start naming. A rightward pointing arrow indicates that the words are to be named in the order that they appeared onscreen (i.e., as read from left to right; "gope doof"). A leftward pointing arrow indicates that the words are to be named in reverse order to that in which they appeared onscreen (i.e., the word presented on the right of the screen is to be named first; "doof gope"). All catch trials involve a leftward pointing arrow, but these are outnumbered by filler and foil trials which involve a rightward pointing arrow. Naming order expectancy is therefore generated by orthographic presentation, and reinforced by trial-type weighting. Competition arises when the expectancy (to name the leftmost word first) is not congruent with the response requirement (to name the rightmost word first).

McMillan and colleagues (2009) employed EPG to record tongue-palate contact during the WOC task. Critical trials involved nonwords that contrasted in word-onset tongue-palate contact location (velar stop v alveolar stop, as in the example "gope doof"). Tongue-palate contact over the period of complete stop closure (as indicated by articulatory records) was analysed[6]. Articulatory records associated with perceptually "correct" responses produced under incongruent conditions revealed instances of "non-canonical" productions. These non-canonical productions involved double articulations, in which tongue-palate contact was observable both at the intended target location and the competitor location. For example, the articulation of "doof" in response to the stimulus "gope doof" ← might show contact in the alveolar region associated with the target /d/ but also in the velar region associated with the competitor "gope". This finding is in keeping with the anomalous tongue EMG activity reported by Mowrey and MacKay (1990), and confirms that when a response conflicts with a stimulus dimension this conflict can be expressed at a speech motor level. Overall, perceptually correct responses in the study by McMillan and colleagues were more similar to the competitor phoneme when the non-word target had a lexical competitor than when it did not (e.g., "gope" has the lexical competitor "dope" whereas the equivalent competitor for "gofe" is the nonword "dofe"). This lexically conditioned effect suggests that, when a speaker anticipates their own speech production, information cascades from at least a lexical level to an articulatory level. If this is also the case for prediction during comprehension we would expect to see similar articulatory evidence in the studies described in the second part of this thesis.

5.4. Speech imaging to study prediction-as-simulation

In the studies reported in Chapters 6 and 7 we employ speech imaging to investigate whether speech planning during comprehension-elicited prediction, as opposed to during production, produces traces in articulation. This allows us to explore whether predictions are represented at a speech motor level,

[6] In this way the articulatory activity analysed was that recorded during the silent stop portion of the respective onset consonants

as Pickering and Garrod have proposed. As outlined in Chapter 1 of this thesis, the proposal that prediction during comprehension invokes the speech motor system was developed with reference to a broad range of evidence concerning: (i) the relationship between bodily motor activity and body movement prediction and; (ii) the activation during listening of neural networks associated with speech motor activity. In the current thesis we seek to bridge the gap between these two sources of evidence by using articulatory imaging to specifically explore speech activity associated with prediction during spoken language comprehension. We investigate this proposal within in the mixed comprehension-production paradigm developed over the experiments reported in Chapters 3 and 4. The mixed paradigm is used in light of argumentation that prediction is most likely in a production setting (see Huettig, 2015).

In its most recent exposition[7], the prediction-as-production framework invokes a production command as the source of the efference copy which informs the forward production model on which prediction-by-simulation is argued to be predicated (Pickering & Garrod, 2013; see specifically Figure 5 and relevant discussion). The authors state that prediction-by-simulation need not involve an "action implementer", although efference copies are generally understood to be copies of a movement-producing signal sent to an action implementer (e.g., Mathalon & Ford, 2008; Sun et al., 2015), This apparent incompatibility remains to be resolved, with one possibility being that the forward models proposed by Pickering and Garrod are generated at level more abstract than that of the motor-speech command and its associated efference copy (for example, see Sun et al., 2015, for evidence that in humans the efference copy may originate within the premotor cortex; see also Niziolek, Nagarajan, & Houde, 2013 for discussion of possible sources of the efference copy in human speech production). Whilst this issue is undoubtedly pertinent to the development of a biologically plausible, empirically testable model, we note that it is not problematic for the interpretation of studies reported in the current thesis. This is because, as described above, previous articulatory imaging studies have investigated the speech motor consequences of competition arising within the speech production system at a relatively abstract level of representation (usually referred to as a phonological level). Given that articulatory consequences are observed in these studies, we can expect to observe articulatory consequences of comprehension-elicited prediction if it similarly involves representation at a speech sound level within the production system, regardless of whether or not the "efference copy" to which Pickering and Garrod (2013) refer originates at a premotor or motor level.

As detailed above, studies that use a tongue-surface approach to determine points of maximal constriction have provided evidence that competing gestures can be "coproduced", with two constrictions being produced at a single time-point where only one constriction would be produced in a canonical realisation (e.g., Goldstein et al., 2007; Pouplier, 2008). Studies that use a global difference measure have provided evidence that perceptually error-free articulations realised in a competing environment are more similar to the competing gesture than are those produced in a non-

[7] Published during the period that research for this thesis was being conducted

competing environment (e.g., McMillan & Corley, 2010). Findings across the approaches are potentially consistent: the presence of an "intruding" constriction gesture (as reported by Pouplier, 2007, for example) would increase overlap between the target gesture realisation and the competing gesture realisation, thereby increasing similarity on a global measure (as reported by McMillan & Corley, 2010, for example). In the current thesis we adapt the analysis procedure developed by McMillan and Corley in order to make it possible to explore the time-course of global differences. This allows us to investigate articulatory movement during the response latency period. We outline the approach below.

5.5. The ultrasound imaging approach

Ultrasound imaging allows dynamic recording of the movements of the tongue during speech, and has been valuable in providing information about many aspects of articulation, including those concerning lexical and phonological level effects on articulation (see above, see also Stone, 2005, for a comprehensive introduction to the technique; see Davidson, 2005, for an example of a study in which the technique was used to measure co-articulation in order to address a phonological question). Articulator imaging is achieved by placing a Doppler transducer probe (similar to that used in foetal imaging) against the under-surface of the participant's chin. The transducer emits and receives very high-frequency sound waves (inaudible to humans). The sound waves fan out along the midsagittal plane, and are reflected at points where substance impedance changes (primarily at the tongue surface). The transducer receives the reflected echoes, from which it is possible to determine the location co-ordinates of the surface boundary at which a reflection took place. The location coordinates are then converted into a visual image of the oral cavity in midsagittal section.

In the following experiments we employ grey-scale images. The intensity of reflections from any given location is represented on a scale from black (no reflection) to white (total reflection). The tongue surface appears as a bright contour on screen, with the tongue root typically pictured on the left of the screen and the tongue tip on the right. Changes in tongue position, for example those associated with changes in the sound being articulated, are visible as movements of this contour. Sampling rates greater than 300 frames per second (fps) can be achieved. In the current thesis, data were acquired at a rate of 100 fps but were processed at a video rate of ~30 fps for reasons of tractability. This sampling rate allowed us to examine tongue position in relation to key time points determined from the auditory data (e.g., the onset of acoustically available speech), and also to investigate frame-to-frame change in tongue position during the response latency period.

The ultrasound technique is well suited to psycholinguistically-motivated studies, in that it provides a non-invasive and relatively low-cost way to capture tongue movements during speech. However, ultrasound data are notoriously noisy, and are both time-consuming and complex to process. Typically, the processing of speech ultrasound data requires considerable manual labour to determine the location of the tongue surface (as opposed to other reflective surfaces) at any given point during an

utterance. Although tongue surface contour tracking can be semi-automated, the algorithms which permit this generally require guide information obtained through visual inspection and manual annotation of the image by the researcher (for further description and an example, see Pouplier, 2008). This increases the potential for researcher subjectivity and error to impact findings, and, perhaps more significantly, limits the quantity of data which can reasonably be processed. This constraint means that, although data is captured dynamically, analysis tends to be conducted on only a single frame per token.

In the current experiment we were not concerned with the absolute position of the tongue, but with whether articulation varies systematically as a function of the relationship between the predicted word and the articulated word. This meant that we were able to use and extend a semi-automated analysis approach which does not rely on tongue contour tracing (the Delta approach: McMillan & Corley, 2010; McMillan, 2008). This approach allows each token to be represented by multiple frames, allowing the dynamics of articulation to be examined and compared across conditions. In previous implementations the technique has been employed to allow time-adjusted pairwise comparisons, which capture within a single measure the position of the tongue over the full time course of a given target articulation. In this thesis we adapt the approach in order to allow analyses of speech-motor behaviour over time (comparable to analysis approaches typically adopted for eye-tracking and EEG data). In addition to making our analyses more directly comparable to those found within the comprehension literature, this adaptation allows us to integrate aspects of the two key traditions within speech production research: In keeping with error research we examine spatial aspects of articulation; in keeping with response latency research we examine temporal aspects.

Chapter 6: Articulatory Imaging Implicates Prediction During Spoken Language Comprehension[8]

6.1. Introduction

As reviewed in previous chapters, it has been suggested that the activation of speech-motor areas during speech comprehension may, in part, reflect the involvement of the speech production system in synthesising upcoming material at an articulatorily-specified level. In this experiment we seek to explore that suggestion through the use of articulatory imaging. We investigate whether, and how, predictions that emerge during speech comprehension influence articulatory realisations during picture-naming.

6.1.1. Paper abstract

We elicited predictions by auditorily presenting high-cloze sentence-stems to participants (e.g., "When we want water we just turn on the…"), as in Experiments 2 and 3 (see Chapters 3 and 4). Participants named a picture immediately following each sentence-stem presentation. Pictures either matched (e.g., TAP) or mismatched (e.g., CAP) the high-cloze sentence-stem target. Throughout each trial participants' speech-motor movements were recorded via dynamic ultrasound imaging. This allowed us to compare articulations in the match (full-overlap) and mismatch (rime-overlap) conditions to each other and to a control condition (simple picture-naming). Articulations in the mismatch condition differed more from the control condition than did those in the match condition. This difference was reflected in a second analysis which showed greater frame-by-frame change in articulator positions for the mismatch compared to the match condition around 300-500 ms before the onset of the picture name. Our findings indicate that comprehension-elicited prediction influences speech-motor production, suggesting that the speech production system may be implicated in the representation of such predictions.

[8] This chapter constitutes an extended version of a published paper ("Drake, E., & Corley, M. (2015b). Drake, E., & Corley, M. (2015). Articulatory imaging implicates prediction during spoken language comprehension. *Memory & Cognition*, *43*(8), 1136-1147").

6.1.2. Prediction within the production system

When we listen to another person speaking, our own motor system is activated (Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Pulvermüller et al., 2006; Watkins & Paus, 2004; Wilson, Saygin, Sereno, & Iacoboni, 2004; for reviews see Gambi & Pickering, 2013; Scott, McGettigan, & Eisner, 2009). This motor activation appears to reflect two levels of representation; referential resonance and communicative resonance (Fischer & Zwaan, 2008; Willems & Hagoort, 2007). Referential resonance describes activation elicited by the linguistic content of the listened-to-material, and involves the representation or simulation of motor acts referred to by the speakers (e.g., hearing "kick" activates leg areas: Hauk, Johnsrude, & Pulvermüller, 2004; Tettamanti et al., 2005). Communicative resonance describes activation related to the phonetic content, and involves representation or simulation of the motor activity involved in speech production itself (e.g., hearing /k^{h}ɪk/ activates areas involved in the articulation of that sound stream: Fadiga et al., 2002; Pulvermüller et al., 2006). This experiment is concerned with the speech-motor activation associated with communicative resonance. We employ articulatory imaging to investigate the suggestion that, as well as reflecting the bottom-up processing of auditory material as it is encountered, communicative resonance additionally indexes the top-down prediction of to-be-heard material (e.g., Pickering & Garrod, 2007; Schiller et al., 2009). If activity in the speech-motor system were shown to be related to prediction, the speech production system would be implicated, as suggested by Pickering and Garrod (2007).

Predictions would need to be made at least at the level of phonological-phonetic speech sounds for relevant activation of the speech-motor system to ensue. The prediction of phonologically-specified representations has been demonstrated during written language comprehension: In an RSVP reading study, participants displayed N400-indexed surprisal upon encountering the indefinite article (*a/an*) in a phonological form that was inappropriate given the predicted upcoming word (e.g., encountering *an* when strongly constrained to anticipate a noun with a consonant onset such as *kite*; DeLong, Urbach, & Kutas, 2005). However, our response latency findings suggest that the phonological form of a predicted word does not facilitate picture naming in the way that is seen in typical PWI experiments where the "distractor" word is actually presented (see first part of this thesis). In contrast to findings from PWI studies, including our own, phonological overlap between predicted words and picture names was not found to have an effect on response times: Participants were no quicker to name a picture when its name partially overlapped with the predicted word (TAN) than when it didn't (COAT: Chapter 4; see also Severens, Ratinckx, Ferreira, & Hartsuiker, 2008). It therefore remains open to question whether speech-sound predictions are generated during spoken language comprehension, and, if so, whether the speech-motor system and the production system more generally, are implicated.

A reasonable interpretation of the evidence by the response latency studies would be that speech sounds are not routinely predicted in the production system during spoken language comprehension,

at least not to the extent that they affect the timing of responses. However, the time taken to name pictures may be an inappropriate measure to base such a conclusion on: In order to complete the task, participants had to decide *when* to speak, and they may have been able to make use of prosodic and timing cues from the spoken sentence fragments in order to do so (e.g., Wilson & Wilson, 2005; see also Rothermich & Kotz, 2013). To the extent that participants' speech timing was governed by extrinsic as well as intrinsic factors, subtle differences in naming latencies may have been difficult to detect, in contrast to PWI studies, where no extrinsic timing information is available.

Another reason for treating the behavioural evidence with caution is that there does appear to be evidence which implicates motor areas in prediction more generally. Such evidence derives primarily from studies of representational momentum in the perception of human movement (e.g., Verfaillie & Daems, 2002; Miall & Wolpert, 1996; Miall & Reckess, 2002; Huber & Krist, 2004; see Pickering & Garrod, 2007; 2013, for discussion with respect to language comprehension). In order to directly investigate the involvement of the speech-motor system in the prediction of upcoming sounds, a more appropriate source of evidence than speech timing may be the articulatory movements that are the product of activation in the speech-motor areas on an ongoing basis. If this activation reflects, in any part, the prediction of upcoming speech sounds, then we should be able to find evidence for the activation in perturbations of speech-sound movements made during the time in which such predictions are active.

We recorded the responses of eight new participants in an experiment which was closely related to the response latency experiments reported in Chapter 4. Predictions were elicited using high-cloze sentence-stems, each of which strongly predicted a specific word (cf. DeLong et al., 2005). Presentation of the sentence stems was as in the previous experiments: following each auditorily presented sentence-stem (e,.g., *When we want water we just turn on the…*), participants named a picture which either matched the predicted word (i.e. full-overlap; TAP), or differed in onset (i.e., rime-overlap; CAP), in a fully counterbalanced design. As previously, we used pictures rather than written words because it has been suggested that written words have privileged access to articulation (e.g., Costa, Alario, & Caramazza, 2005). We anticipated that, in cases where participants were anticipating *tap* but naming a CAP, activation of the speech-motor system related to prediction would affect the articulation of *cap* , such that its onset would be 'less /k/-like' than in the case where *cap* was predicted (*On his head he wore the school…)*. To investigate this, we measured the articulations of the same picture names where there was no sentence stem and therefore no potential interference from a predicted word. By calculating the differences between the articulations of picture names in experimental and control conditions, we were able to establish whether articulation varied more from the control when participants anticipated that they would hear a mismatching word than when a matching word was predicted. By calculating the degree of movement over time in the matching and mismatching conditions, we were able to investigate whether there were specific periods during articulation where there was more movement in one experimental condition relative to the other.

6.2. Method

6.2.1. Participants

Eight participants (7 female) aged between 21 and 40 years took part in the study. All participants were monolingual speakers of English, had normal or corrected-to-normal vision, and reported no positive history for hearing or speech-language difficulties. Participants were recruited from research pools at Queen Margaret University and the University of Edinburgh, were paid for their participation, and gave written informed consent in line with BPS guidelines. The study was granted ethical approval by the Psychology Research Ethics Committee of the University of Edinburgh (approval no. 14-1213/1).

6.2.2. Materials

Items were the experimental items used in Experiment 3 (Chapter 4; see Appendix C for details). Twelve pictures were used as experimental items; a further two pictures were used as practice items. Picture names were single-syllable and represented the 6 rimes /-æn, -æp, -eɪp, -eɪk, -əʊn, -əʊst/, each paired once with the onset /t-/ and once with the onset /k-/ (can, tan, cap, tap, cape, tape, cake, take, cone, tone, coast, toast). As in the previous experiments, the final, high-cloze item was omitted from all sentence-stem recordings.

6.2.3. Procedure

Participants wore an ultrasound probe throughout the experiment. The probe was secured directly against the under-surface of the chin using a proprietary helmet (Articulate Instruments: http://www.articulateinstruments.com/ultrasound-products/). This allowed us to record the movement of the tongue within the oral cavity during each trial (the tongue is the key active supralaryngeal articulator). Ultrasound images were recorded at a rate of ~30 frames-per-second, with acoustic data being simultaneously captured via Articulate Assistant Advanced (Articulate Instruments, 2012; for details, see Wrench & Scobbie, 2008).

As in previous experiments reported in this thesis, the experiment was presented on a laptop, using DMdX software (Forster & Forster, 2003). Participants were trained on the correct name for each picture prior to the experiment to ensure that any articulatory differences could be ascribed to competition between the predicted word and the picture name, rather than to uncertainty concerning the name of the picture. All participants named the pictures with 100% accuracy by the end of the training phase (which consisted of 3 exposures to each picture).

In all blocks, trial presentation was randomized via the presentation software. In the first experimental block, participants named each picture aloud once. Participants viewed a fixation point in the centre of the screen for 2.9 seconds immediately prior to the presentation of each picture-to-be-named. Participants were instructed to name the pictures as soon as they could, but to make sure that their naming was accurate.

In blocks 2 and 3, participants again viewed a fixation point immediately prior to the presentation of each picture for naming, but this time while listening to an auditory sentence-stem. In all trials the picture was presented immediately after the end of the auditory sentence-stem. Sentence-stems and pictures were paired together within trials so that in half of the trials the sentence-stem predicted the picture-name (i.e., match/ full-overlap condition, e.g., *On his head the boy wore the school…* CAP): In the other half of the trials the sentence-stem predicted a name that rhymed with the name of the picture presented for naming (i.e., mismatch/ rime-overlap condition, e.g., *Jimmy used a washer to fix the drip from the old leaky…* CAP). All sentence-stems were presented once in each experimental condition. The condition in which a sentence-stem was first encountered was counterbalanced across participants. Participants encountered an equal number of match and mismatch trials in each of blocks two and three.

Block four was identical to block one (i.e., simple picture naming following a fixation point). Trials from blocks one and four formed the control condition. Each participant named each picture 8 times in total (twice in the control condition, three times in the match and three times in the mismatch condition). In all blocks participants followed the same instruction; to name the picture as quickly and accurately as they could. Including setup, the experiment lasted approximately 30 minutes.

6.2.4. Data treatment and analysis approach

The ultrasound data for each token recorded consisted of a sequence of black and white video frames. For each frame, there were 141,824 pixels (517 x 277) which ranged in luminance from 0 (black) to 255 (white). To make the analysis tractable, we first calculated the average luminance of each 8 × 8 grid of pixels, resulting in a 2,240-pixel "pixelized" image.

In order to analyse the pixelized ultrasound images, we first inspected the relevant audio file independent of the visual data and blind to the experimental condition, using Audacity (http://audacity.sourceforge.net). We identified two key moments during the participants' productions of each word: the *acoustic release* of the onset consonant, and the *end of the vowel*. The *acoustic response latency* was taken to be the time between stimulus (picture) presentation and the acoustic burst of the onset consonant of the picture name (i.e. the onset of a target-specific acoustic signal visible in the waveform). It was possible to determine this period for 7 of the 8 participants[9]. Time-

[9] The onset of picture presentation was determined as being the point at which the acoustic presentation of the sentence-stem stopped. In the case of the one participant excluded from the

points based on the audio recordings were also used to select portions of each video for further analysis. Different portions of video were used for different purposes, as described in the relevant sections below. Once a portion of video had been extracted, it was expanded or contracted to a standardised number of video frames, using an averaging algorithm. This allowed us to control for slight differences in video frame rate and in articulation timings.

Differences between standardised video sequences were calculated using the Delta technique (McMillan & Corley, 2010). The pixels in each frame were represented as a 2,240-dimensional vector, with each dimension taking values between 0 (black) and 255 (white). Differences between pairs of frames were calculated as the Euclidean distances between vector representations; and differences between sequences of frames were calculated as the average by-pair Euclidean distance. This quantity, in arbitrary units, is referred to as the *Delta distance*.

When analysing the ultrasound video, we initially generated a data quality metric by calculating '*discrimination scores'* for the data recorded in each session (see "Recording Quality" below). These discrimination scores were then used as weighting factors in a series of regressions examining the effects of the experimental manipulations (with the consequence that observations with higher discrimination scores had more influence on the reported outcome). First, we examined the degree to which participants' productions were affected by auditory sentential context, by comparing the degrees to which their experimental articulations varied from control articulations in matching (full-overlap) and mismatching (rime-overlap) conditions (see "Differences Between Conditions"). Second, we examined the degree of movement over the time-course of each articulation, allowing us to examine the time-course of articulatory differences due to context (see "Time-Course of Differences"). We begin with an illustration of issues with recording quality and the derivation of the discrimination score, before presenting each of the experimental analyses in turn.

6.2.5. Recording Quality

A problem with ultrasound recordings of articulatory movements is that they can vary greatly in quality, due to individual differences in the tongue and oral cavity, noise in the recordings, and ultrasound probe slippage, among other factors. However, such differences are difficult to detect at recording time.

In the present study we reduced the impact of this issue by generating discrimination scores. We conceptualised recording quality as the ability to discriminate between the six different CV onsets used throughout the present paradigm (/kæ/, /keɪ/, /kəʊ/, /tæ/, /teɪ/, /təʊ/). We used ultrasound video beginning 0.1s before consonant onset (acoustic burst), and ending at the offset of the steady-state vowel. The relevant section of each video was digitised and quantised into 8 frames, each of which

response time data, the recording of the sentence-stem presentation was not loud enough to permit reliable annotation.

represented approximately 33 ms of recorded time. For each participant, we then created a table of the Delta distances between each possible pair of articulations. Initially, we used multidimensional scaling (Gower, 1966; Mardia, 1978) over two dimensions to visualise the relationships between a participant's recordings. Figure 6.2.1 shows data from the participants we judged, by visual inspection, to have produced the 'best' and 'worst' recordings (least and most noisy recordings). Whereas the left-hand plot clearly shows that ultrasound analysis using the Delta approach is capable of distinguishing articulations, the right-hand plot shows that this capability is at the mercy of the noise that is inherent in ultrasound recordings.

In order to deal with this problem, we generated a 'discrimination score' for each participant and each CV onset, designed to calculate how well a given CV such as /keɪ/ could be discriminated from the other CVs in the experiment (here, /kæ/, /kəʊ/, /teɪ/, /tæ/, /təʊ/). These calculations were based on articulations from the control conditions only, since we predicted additional variability in articulation in the experimental conditions. Using the tables of Delta distances calculated above, we divided the mean distance between control articulations of words which *didn't* share a given CV onset by the mean distance between words which *did* share that onset. The more discriminable the words sharing an onset were from the other words, the higher the discrimination score was. Discrimination scores ranged from 1.25 to 2.43 (mean 1.62; SD 0.29). Table 6.2.1 shows the discrimination scores calculated for the participants shown in Figure 6.2.1.

Table 6.2.1: Discrimination scores calculated for the participants shown in Figure 6.2.1

CV	CV(IPA)	Participant A	Participant B
ke	(/keɪ/)	2.39	1.3
ko	(/kəʊ/)	2.43	1.42
ka	(/kæ/)	2.32	1.32
te	(/teɪ/)	2.00	1.36
to	(/təʊ/)	1.97	1.28
ta	(/tæ/)	2.14	1.25

Scores are calculated from the control articulations only, and represent the degree to which articulation of a given CV can be distinguished from other CVs in the ultrasound recordings.

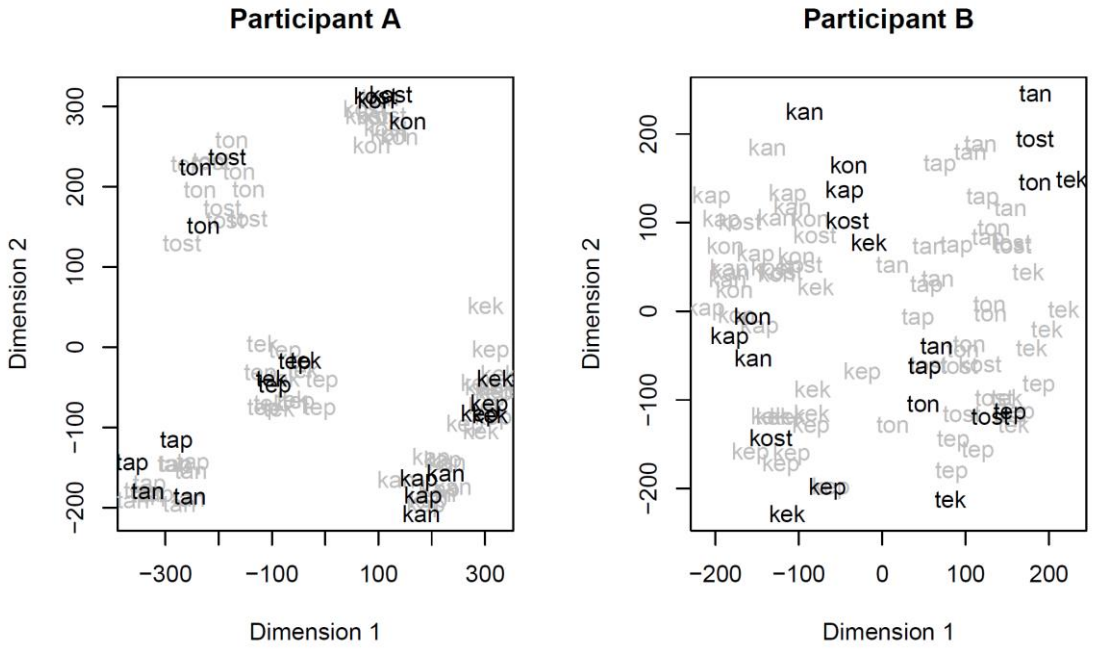

Figure 6.2.1: Multidimensional scaling of Delta differences between all articulations produced by each of two participants, measured from 0.1s before consonant onset to vowel offset

The plots show (left) that the Delta technique is highly capable of distinguishing articulations, but that (right) ultrasound recordings can be subject to noise. Words in black represent recordings from control conditions, and words in grey correspond to experimental conditions.

In keeping with the approach to response latency analysis reported in previous chapters of this thesis, analyses of the treated data were all conducted using linear mixed effects models with maximally specified random effects structures (following Barr, Levy, Scheepers, & Tily, 2013).

6.3. Results

Individual audio and ultrasound recordings were obtained of each participant's articulatory movements during each trial, and were digitized to video. Each participant produced 96 picture names (24 in the control conditions, and 72 in the experimental conditions). Of the resulting 768 recordings, 27 (3.5%) were discarded because of failures either to record audio, or to properly register ultrasound. There was no difference between conditions in the proportions of recordings removed ($\chi^2_{(2)} = 2.62$, $p = .27$). The remaining 741 recordings were used in all subsequent analyses, including the calculation of discrimination scores described in the Methods section above. Note that the portion of a recording used to calculate the discrimination score (acoustic onset forward) differed from that used in the main analyses reported below (acoustic onset backward).

6.3.1. Response Latencies

When participants named pictures in the match condition (mean RT = 515 ms; se = 11 ms) the acoustic burst occurred sooner than in the mismatch condition (mean RT = 632 ms; se = 12 ms) or control condition (mean RT = 606 ms; se = 15ms). We conducted a mixed-effects regression analysis of the effect of sentential context (match, mismatch, or control) on RTs. The model included both intercepts and slopes which could vary by participant and by picture name. Random effects for intercepts and slopes were allowed to correlate. This constitutes the maximal justified random effects structure, in line with recent recommendations for confirmatory hypothesis testing (Barr, Levy, Scheepers, & Tily, 2013). Using orthogonal contrasts, the model confirmed that response latencies in the match condition were significantly shorter than in the mismatch and control conditions (β = 110, se = 28, t = 3.96), and that response times did not differ between mismatch and control conditions (β = 20, se = 31, t = 0.63). This pattern replicates the patterns for the relevant conditions (rime-overlap and control) reported in the response latency studies in which this paradigm was employed (see Chapter 4).

6.3.2. Acoustic durations

We had manually segmented the acoustic portion of picture name responses in order to provide time-locking points for the ultrasound analyses (see Method, above). This enabled us to investigate whether there were systematic by-condition differences in the acoustic durations of; (i) the consonant burst; (ii) the vowel segment; (iii) the whole syllable to the final consonant. The effect of Context on these measures was investigated by, in each case, comparing a null model that contained only random effects (intercepts and slopes by context for participant; intercepts only by context for item) to a model in which Context was included as a fixed effect. In no case did including Context as a fixed effect improve model fit ($\chi^2_{(2)}$ = 3.64, p = 0.16; $\chi^2_{(2)}$ = 2.56, p = 0.28; $\chi^2_{(2)}$ = 1.76 , p = 0.4143 for consonant burst, vowel segment, and whole syllable durations respectively).

6.3.3. Ultrasound Analysis

All regression models reported here were weighted (Carroll & Ruppert, 1988), using the CV-specific *discrimination scores* described in the Methods section. To avoid misrepresenting the effective power of the experiment, discrimination scores were scaled to a geometric mean of 1. This allowed recordings which were better able to capture relevant differences between control articulations to have greater influence on the outcomes of the analyses, without arbitrarily excluding recordings which may have been of poorer quality. In this context, it should be noted that the discrimination measure is independent of within-cell variation about the mean (correlation: r = −0.01).

6.3.3.1. Differences Between Conditions

The effect of context on articulation was investigated by comparing articulations in experimental conditions to reference articulations from the control condition. Here, we were primarily interested in the production of the onset consonant /k/ or /t/, since the rimes in picture names never differed from the rimes predicted by context. Accordingly, we extracted ultrasound video starting half a second before the consonant onset and ending at the consonant release (approximately 17 frames of video at 30 fps). All recordings were averaged into 17 frames; for each participant, we then proceeded as follows. First, we created participant-specific reference articulations of the onset consonants /k/ and /t/, by averaging the luminance of each of the 2,240 pixels frame-by-frame for all of 17-frame sequences representing control articulations of words beginning with /k/ or /t/ respectively. We then calculated a *Delta score* for each individual articulation produced in the experimental conditions, representing the (mean frame-by-frame Euclidean) difference between a particular onset articulation and that participant's mean control articulation of the same onset (see McMillan & Corley 2010).[10]

The *Delta scores* thus obtained were subjected to a mixed-effects regression analysis, examining the effects of onset (/k/ or /t/) and of context (match or mismatch) on deviance from mean control articulation. Together with these fixed effects and their interaction, our model included intercepts which could vary randomly by participant and by picture name. As in previous response latency analyses, the slopes associated with each fixed effect and the interaction could vary by participant, and the slope associated with context could vary by picture name (item). Random effects for intercepts and slopes were allowed to correlate. This model therefore includes the maximal justified random effects structure (Barr et al., 2013). Predictors were centred about their means prior to analysis. We considered coefficients to differ reliably from zero where $|t| > 2$ (see Chapter 3). Because our conclusions were based on model coefficients, we fit models using restricted maximum likelihood, to reduce the probability of Type I errors.

Discrimination score weighted regression

Discrimination score weighted regression showed that a numerical tendency for participants to produce /t/ onsets which differed more from the participant-specific /t/ controls than their /k/ productions differed from the /k/ controls failed to reach significance ($\beta = 9.98$, $SE(\beta) = 5.77$, $t = 1.73$). Participants were, however, affected by sentential contexts, such that onsets produced in the mismatching condition differed more from their controls than did those produced in the matching

[10] Due to the nature of ultrasound recordings, a number of pixels in each frame are more-or-less randomly grey. However, pixels at clear physiological junctures tend to be more deterministically coloured, and there are likely to be similarities in luminance patterns across frames for similar tongue positions within a given speaker. Similarities between pixels will tend to reduce Delta values, allowing us to distinguish signal from noise.

condition (β = 10.89, SE(β) = 5.05, t = 2.15). The effect of context did not differ by onset consonant (t = 0.58)[11] . Table 6.3.1 gives full details of the regression model.

Table 6.3.1: Model coefficients (in Delta units) for differences from Control condition: Details of Context by Onset model.

Fixed Effect	**Estimate(β)**	**SE(β)**	**t**	**Random Effect**		**Variance**
Intercept	286.87	8.99	31.90	Participant	Intercept	589.61
Context (mismatch vs match)	10.89	5.05	2.15		Context	39.37
Onset (/t/ vs /k/)	9.98	5.76	1.73		Onset	38.03
Context * Onset	5.62	9.63	0.58		Context*Onset	82.64
				Item	Intercept	39.16
					Context	63.73
Observations = 559				Residual		2159.63

6.3.3.2. Time-Course of Differences

In order to investigate the time-course of articulation we extracted standardised ultrasound videos corresponding to the period from 1 second before consonant onset to consonant release (approximately 32 frames of video at 30fps). Using the same vectorisations as for *Delta* calculations, we then calculated Euclidean distances between successive frames of standardised ultrasound video, producing a sequence of 31 inter-frame values which represent moment-by-moment degree of movement for a particular articulation. These values are related to 'speed' of articulatory movement rather than 'velocity', since they do not include information on the direction of movement. However, it is possible to determine at which points in time participants' tongues tend to be moving quickly between frames, and at which points they are more stationary; these values are then used to provide plots of the speed of tongue movements over time in different experimental conditions.

Euclidean distances between successive frames of ultrasound were compared by experimental condition using a series of mixed-effects regression analyses, investigating the effects of onset, context, and their interaction at each time point. Models were fit using restricted maximum likelihood

[11] Regression without weights showed the same general pattern of results, although the difference between onsets reached significance: /t/s differed more from their controls than did /k/s (β = 10.99, SE(β) = 5.28, t = 2.08); mismatching onsets differed more than matching onsets (β = 10.94, SE(β) = 5.26, t = 2.08); there was no interaction (t = 0.64).

and included maximally justified random effects, as described in the previous section. Models were weighted by CV-specific *discrimination scores*. Effects were considered reliable where |t| > 2.

There were no interactions between Onset and Context at any time-point. Effects of Onset were found from approximately -217 to -117 ms and -83 to -50 ms, reflecting more frame-to-frame movement for /k/ at the earlier epoch and more movement for /t/ just prior to consonant release. Effects of Context were found from approximately -483 to -283 ms and from -50 to -17 ms, reflecting more frame-to-frame movement in the mismatch (rime-overlap) condition in each case.

Because of the cumulative risk of a Type 1 error associated with multiple independent tests of this nature, we do not consider isolated significant differences further, but instead focus on time-points when there are clusters of differences. Figure 6.3.1 illustrates the differences between conditions over the time-course of articulation.

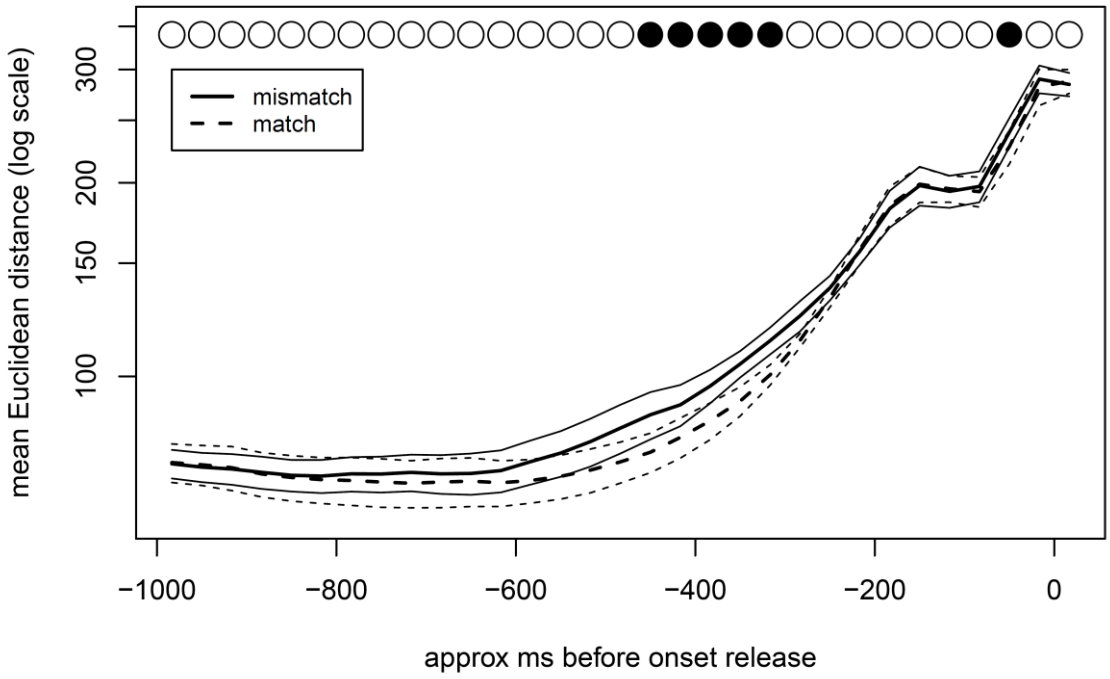

Figure 6.3.1: Articulatory movement over time in producing the onsets of picture names which match or do not match predictions from the sentence stem.

The x-axis represents time milliseconds prior to the acoustic response (i.e. the response latency period between picture presentation and acoustic picture naming). The y-axis represents the amount of change in the ultrasound image between two consecutive frames of the ultrasound record (averaged by condition). This can be *loosely* interpreted as the extent of articulatory movement between frames. Time zero represents the release of the /t/ or /k/ onset. Lines show the Euclidean distances between vectors of pixel intensity for successive (normalised) frames of ultrasound video, together with by-participant standard errors. *y*-axis is log-scaled to help with viewing of differences. Filled circles correspond to transitions at which there is a significant difference (at |t| > 2 for mixed models weighted by discrimination score, with participants and words as random effects) between mismatched and matched onset productions.

6.4. Discussion

We recorded ultrasound images of tongue movements while participants named pictures. In one experimental condition, the name of the picture matched the most likely continuation of a spoken sentence stem that the participant had just heard; in another, the picture name was a mismatching name which began with a different consonant. We used two approaches to compare articulation across matching and mismatching conditions, and found in both cases that articulation prior to the consonant release differed between conditions. The by-condition difference confirms that predictions made as a listener can affect production. More specifically, the finding demonstrates that prediction from another's speech affects the motor execution of one's own speech, suggesting that top-down prediction may involve simulation of the motor activity involved in speech production.

In the first analysis, we compared summarised articulatory movements directly, and found that participants' articulations in the mismatch (rime-overlap) condition differed more from their average articulations in a control condition when a mismatching word was predicted than when the prediction was of a matching (full-overlap) word. One potential account of this process might be that, compared to the faster matching condition, articulation is simply slowed in the mismatching conditions. Under this hypothesis, apparent differences in variability would, in fact, be due to differences in the timings of articulatory gestures. However, two aspects of the data militate against this view. The first is that the differences in naming latencies between match, mismatch, and control conditions do not align with the differences in articulations. Both the mismatch and control conditions result in slower naming latencies than the match condition; if articulation speed explains the differences in the first analysis, then the mismatch articulation should be more similar to the control than the match articulation. In fact, the opposite is the case.

The second argument against an account based on speed of articulation comes from the second analysis, in which we inspected the frame-by-frame degree of movement involved in each articulation. We achieved this by measuring the differences between consecutive frames of each ultrasound video in experimental conditions. The resultant measurements encompassed the second or so leading up to the acoustic release of each onset consonant, and showed that where there were differences between conditions, the mismatch condition showed greater movement. Again, these findings are inconsistent with the view that differences in the mismatch condition can be ascribed to generally slower articulatory movements associated with a longer response latency. Taken together, the analyses provide prima facie evidence for an influence of linguistic prediction on the manner, rather than the timing, of articulatory movements when the person making the prediction has to speak.

The time-course analysis additionally reveals that the period during which the frame-to-frame change is greater in the mismatch condition is relatively early in the articulatory gesture, at around 500-300ms

before the onset of the picture name. After this period the articulatory trajectories in the two conditions converge, and are statistically indistinguishable by 280ms prior to the acoustic release, and for the remainder of the articulation. This reflects the facts that the onset of articulatory movement in the mismatch condition occurs significantly earlier in relation to acoustic release than in the match condition; and that less articulatory movement is required overall to achieve the acoustic target in the match condition than in the mismatch condition. In other words, articulation in the match condition is ultimately more efficient than in the mismatch condition. To produce words that mismatch a prediction generated as a listener is not only more demanding at a cognitive level (pace response latencies), but also more demanding at a motor-execution level.

Findings from the visuo-motor control literature suggest that the 'inefficiency' seen in the mismatch condition may be due to articulation in that condition involving a movement toward the (incorrect) predicted target. Perturbation of movement towards an incorrect target has been observed to be the case when two stimuli "try to control the same speeded motor response" (Schmidt & Schmidt, 2009, p.595; in that case, of movements with a stylus towards a location that either matches or mismatches the location of a masked prime). As upcoming predicted lexical items can be specified at least as early as presentation of the preceding word (DeLong et al., 2005; see Introduction), it is possible that in the current experiment both the predicted item and the item-to-be-named were "trying to control" the motor response. The analyses employed in the present study do not allow us to directly address this possibility; it may be feasible to address the question more directly in future studies, given advances in clarity of ultrasound recordings.

Whatever the specifics of the influences on participants' articulations, articulations are qualitatively affected by the presence of lexical representations which have been generated entirely endogenously; the 'competing' predicted words were the product of the participants' semantic prediction systems, having a endogenous rather than an exogenous origin. We were able to observe anticipatory speech-motor consequences associated with predicting from another person's speech. To that extent, the current study directly implicates the listener's speech-motor system in the top-down prediction of upcoming material at the level of communicative resonance.

It appears that anticipatory activation in the speech-motor system is largely outside strategic control: Prediction of the upcoming item was not beneficial to overall performance in the experimental context, and experiments reported in Chapter 4 indicate that mismatching predictions do not generally produce temporal inhibition. Although likely to be automatic, the activation may be specific to situations in which the listener anticipates their own role as a speaker (as one does in dialogue: see Rommers, Meyer, Piai, & Huettig, 2013, for evidence that the neural processing of linguistic material differs depending on whether one expects to be required to speak or not).

Having considered how the data inform our understanding of the issue that the study was specifically designed to address, we turn briefly to a more general issue: The time-course of articulator

movements in the current study strongly suggests that stimulus-related lingual movement occurs well before the acoustic response onset, at a point when cognitive processing would be expected to be ongoing. This finding is perhaps surprising in light of psycholinguistic models of picture naming, which generally involve a sequence of at least four processes *prior* to the initiation of articulation (for a brief recent review see Strijkers & Costa, 2011). According to mappings of the time course of picture name production processes determined via meta-analyses of neuroimaging studies (Indefrey & Levelt, 2004; Indefrey, 2011; see also Laganaro, Python & Toeppel, 2013), motor programming and execution occur only in the final 150 ms prior to acoustic onset of the target picture name. However, the current experiment indicates that articulation starts much earlier, in line with electromyelographic (EMG) data presented by Riès and colleagues (2012) which showed that speech-associated muscular innervation is observable around 380 ms prior to acoustic response onset (Riès et al., 2012; see also Schuhmann, Schiller, Goebel & Sack, 2012). This study confirms the conclusion drawn from Riès et al. (2012) that if we are to further understand the processes involved in speech production it will be necessary to consider effector activity as an important observable outcome and time-course marker, in addition to the acoustic onset more typically used as a time-locking point.

The generalizability of the findings reported here may be impacted by the relatively low number of participants tested. In fact, due to pragmatic difficulties with data collection, this is a common issue with speech-motor studies (comparable numbers of participants are reported by Davidson, 2005; Pulvermüller et al., 2006; Pouplier, 2008; Watkins & Paus, 2004; Watkins, Strafella, & Paus, 2003). In the case of the current study, this concern may be partially mitigated by the fact that the pattern of response latencies was in keeping with that reported in previous chapters, in which participant numbers were in keeping with those typically employed in psycholinguistic research.

Given this caveat, this experiment has demonstrated the importance of articulatory measurement in two ways. As discussed above, muscle activation and motor movements associated with articulation appear to start much earlier than supposed in existing psycholinguistic models. This suggests that the use of articulatory information may be important if we are to develop greater insight into the processes of speech production. For example, in previous work the present authors investigated the acoustic onset times to name pictures in a paradigm very similar to that employed here. On the basis that there were no facilitatory or inhibitory effects when the to-be-named picture partially overlapped with the predicted word, we concluded that "prediction during comprehension [did] not appear to occur at a phonological-articulatory level" (Drake & Corley, 2014). The current study indicates that this was far from the final word on the matter: The second consequence of using articulatory measurement is that we are now able to conclude that there clearly *is* an effect of prediction on articulation to be found, if you know where to look.

Chapter 7: Are Articulatory Effects of Prediction During Spoken Language Comprehension Speech Sound Specific?[12]

7.1. Introduction

Speech comprehension incorporates the generation of predictions and evokes neural activity within speech-motor areas (DeLong, Urbach, & Kutas, 2005; Fadiga, Craighero, Buccino, & Rizzolati, 2002; Kutas, DeLong, & Smith, 2011). It has been suggested that the activation of speech-motor areas during speech comprehension may, in part, reflect the involvement of the speech production system in synthesising upcoming material at an articulatorily-specified level (e.g., Pickering & Garrod, 2007). Evidence presented in the previous chapter confirms that predictions evoked during speech comprehension have an impact at a speech motor execution level. In the current chapter we extend the scope of the experiment reported in Chapter 6 in order to explore whether the speech motor effects reflect speech sound specific predictions.

The development of experimental methodologies such as eye-tracking, EEG, MEG, and fMRI has revolutionised our understanding of the processes involved in language comprehension. It is now clear that comprehension involves the top-down prediction of upcoming material as well as the bottom-up processing of perceived input (Altmann & Kamide, 1999; DeLong et al., 2005; Federmeier & Kutas, 1999; Kamide, Altmann, & Haywood, 2003; see Federmeier, 2007, for a review). There is considerable evidence to show that listening to spoken language evokes activity within the listener's motor system (Hauk, Johnsrude, & Pulvermüller, 2004; Pulvermüller et al., 2006; Watkins, Strafella, & Paus, 2003; see Scott, McGettigan, & Eisner, 2009 and previous chapters for reviews). Such motor activity may reflect what is being conveyed (e.g., hearing "kick" activates motor areas associated with leg movement: Hauk et al., 2004; Tettamanti et al., 2005). It can also reflect the process by which it is conveyed (e.g., hearing /khIk/ activates areas involved in the articulation of that sound stream: Pulvermüller et al., 2006). In the previous chapter we presented findings which implicated speech

[12] The work presented in this chapter has been presented and published in a peer-reviewed conference paper (Drake, E., Schaeffler, S., & Corley, M., 2015, Articulatory consequences of prediction during comprehension. *Proceedings of the 18th ICPhS, Glasgow*, 0618.)

motor system involvement in the prediction of upcoming input during language comprehension. The evidence we presented demonstrated that changing the relationship between the word predicted and picture named has an impact on articulation during picture naming. The aim of the present experiment is to investigate whether the speech motor effects reported in Chapter 6 of this thesis (also Drake & Corley, 2015b) reflect motor representation of predicted speech sounds, or whether they arise at a more general level.

7.1.1. Somatotopic activation during listening

Listening to speech-like syllables leads to activation in the speech-motor cortex (S. M. Wilson, Saygin, Sereno, & Iacoboni, 2004). This activation is somatotopic: Passive perception of labial onset syllables activates the part of the pre-central gyrus associated with the production of lip movements, whereas listening to alveolar onset syllables activates the region associated with tongue movements (D'Ausilio et al., 2009; Pulvermüller et al., 2006). Similarly, electromyelographic activity in a given articulator is raised when a speech sound involving that articulator is presented (Fadiga et al., 2002; Watkins & Paus, 2004; Watkins et al., 2003). Such concomitant speech-motor activation can be accounted for in terms of resonance between the comprehension and production systems. However, recent findings suggest a causal role for speech-motor activation during spoken language listening: TMS stimulation of specific areas of the speech-motor and premotor cortex been demonstrated to affect the process of speech comprehension (Schomers, Kirilina, Weigand, Bajbouj, & Pulvermüller, 2014): People are quicker to match heard words to pictures when the place of articulation of the word onset is consistent with a speech-motor area activated via TMS than when the area activated relates to a competing articulator.

7.1.2. Somatotopic activation as an aspect of prediction during listening

As has been reviewed in previous chapters, a number of frameworks include the proposal that neural regions traditionally considered to form part of the speech production system are involved in predicting the linguistic input which is currently being attended (e.g., Pickering & Garrod, 2007; Schiller, Horemans, Ganushchak, & Koester, 2009; Dell & Chang, 2014). In response to findings that; (i) comprehension can involve prediction to a speech-sound level (e.g., DeLong at al., 2005), and; (ii) speech listening invokes activation of speech motor regions (see above), Pickering and Garrod (2007) proposed that speech motor activation during listening might, in part, reflect synthesis of upcoming speech at a speech-motor level. This suggestion was situated in evidence and theory from the broader visuo-motor control literature, indicating that our own motor systems are involved in predicting the action of others through a process of emulation (e.g., M. Wilson & Knoblich, 2005; for a recent review of the evidence see Colling, Thompson, & Sutton, 2014).

7.1.3. Predictive emulation

Predictive emulation of others' actions is thought to engage processes more typically associated with planning of one's own actions, specifically the generation of forward models (e.g., Grush, 2004). The notion that forward modelling contributes to our ability to predict other's movements is supported by the finding that performing actions interferes with simultaneous action prediction, and that the degree of interference is modulated by the degree of overlap between the executed and predicted actions (Springer et al., 2011). One source of evidence that has been cited in support of the view that the production system is implicated in prediction during comprehension is that listeners are affected by metrical and phonological mismatches to predictable information at a suprasegmental level. Metric foot structure allows listeners to form temporal predictions about when upcoming salient syllables will occur (Lidji, Palmer, Peretz, & Morningstar, 2011). When temporal predictions generated during sentence listening are violated, activation levels are raised within an area of the left inferior frontal gyrus which is associated with phonological processing (Rothermich & Kotz, 2013). The localisation of this effect is distinct from that of effects observed when semantic predictions are violated, and implicates subcortical sensorimotor areas in comprehension and prediction.

Event-related brain potentials also reveal that predictions generated during reading comprehension can be specified to a speech-sound level (DeLong et al., 2005; Foucart, Martin, Moreno, & Costa, 2014; see Chapter 1 for discussion). An N400 effect is observed when readers encounter an indefinite article in a form inappropriate to the anticipated upcoming noun (e.g., 'an' when the anticipated noun is kite). Because the indefinite articles differ only as a function of the phonological onset of the following word, this effect can only be located at a form level (phonological, orthographic, or both), again consistent with the view that prediction implicates the production system. We should note however that there remains an open question as to whether it is correct to conclude from prediction of orthographic material during written language comprehension that the motor activation observed during spoken language comprehension reflects specific predictions of upcoming speech sounds.

In this thesis we investigate the relationship between comprehension and production during prediction by causing participants to comprehend, predict, and produce language simultaneously. Of interest is whether prediction results in (motor) activation within the production system, and whether this activation impacts the subsequent production of spoken words. By a similar logic to that of Springer et al. (2011), we might expect performance on the production task to be modulated by the degree of overlap between what is predicted and what is produced. We conducted three experiments designed along these lines (Drake & Corley, 2015a; see Chapters 3 and 4 of the current thesis). In contrast to findings from PWI studies in which the distractor word is physically presented in written or auditory form (e.g., Meyer & Schriefers, 1991; see also Chapter 2 of the current thesis), phonological overlap between predicted words and picture names was not found to have an effect on response times: Participants were no quicker to name a picture when its name partially overlapped with the predicted

word (e.g., TAP-cap) than when it did not (e.g., TAP-cone; Drake & Corley, 2015a; see also Severens, Ratinckx, Ferreira, & Hartsuiker, 2008).

However, the time at which the response becomes acoustically available is only one source of evidence concerning the speech production system. In the same way that visual world eye-tracking can provide information on the ongoing interpretation of an utterance, and on the prediction of what is likely to be mentioned (e.g., Altmann & Kamide, 1999; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), measures which capture the movements of the articulators during the production of an acoustic signal can provide further insight into ongoing speech-motor activation. In Chapter 6 we report a study in which we employed ultrasound tongue-imaging to such an end. The findings indicated that where there is a phonological contrast between a linguistically predicted item and an item presented for naming, the articulatory path is affected. Such interference at a motor-speech level suggests that the predicted item is in some way represented at this level.

The additional movement observed in mismatching conditions suggests that to produce a mismatching picture name places additional demands on the articulatory system, as well as on the cognitive system (as evidenced by increased reaction times). There are at least two possible accounts of the increased movement found in the mismatching condition. One is that it is specifically driven by a motor representation of a competing (predicted) onset. Perturbation of movement towards an incorrect target has been observed to be the case when two stimuli "try to control the same speeded motor response" (Schmidt & Schmidt, 2009, p. 595), and in the speech error elicitation paradigms reviewed in Chapter 5. An alternative view is that it is the existence, rather than the specific nature, of a competing prediction which gives rise to increased movement. On this view, response competition, rather than the specific details of a competing representation, may be the cause of perturbations in movement (cf. Dhooge & Hartsuiker, 2010).

The two views outlined above can be distinguished by comparing the articulations of picture names which match predictions (When we want water we just turn on the. . . "TAP") to those with matching rimes ("CAP") and those with matching onsets ("TAN"). The first two conditions replicate those investigated in Chapter 6 of this thesis (see also Drake & Corley, 2015b). The third condition is an addition: Because the predicted word in the third condition shares an onset with the picture name, any speech-motor activation associated with prediction should support the production of the required /t/ onset. Perturbations in articulation should therefore be reduced in this third condition relative to those for the matching rime (mismatching onset) case. If, on the other hand, perturbations in articulation reflect response competition, then the phonological specifics of the predicted word should not be relevant; rather, any mismatching prediction should result in greater perturbation to articulation. Below, we present the ultrasound study designed to test these hypotheses.

7.1.4. Current Experiment

Participants named pictures with monosyllabic names directly after hearing high-cloze sentence fragments. The fragments predicted either; (i) the name of the picture (match/ full-overlap condition); (ii) a word with a matching rime (mismatch/rime-overlap condition), or; (iii) a word with a matching onset (mismatch/ onset-overlap condition). On the view that prediction during comprehension invokes specific speech-motor activation, picture names with different onsets from the predicted word should show articulatory perturbation (predicting "tap" and naming CAP would show perturbation; predicting "tap" and naming TAN would minimise perturbation due to predictive activation of /ta/). Alternatively, on the view that perturbations in articulation are the consequence of response competition, picture names which differ in any respect from the predicted word should show articulatory perturbation (i.e., CAP and TAN would show comparable perturbation relative to TAP, when "tap" was the high-cloze sentence target).

7.2. Method

7.2.1. Participants

Twelve monolingual English speakers (9 female) aged between 20 and 28 years took part in the study. Participants were recruited from research pools at Queen Margaret University and the University of Edinburgh, were paid for their participation, and gave written informed consent in line with BPS guidelines. The study was granted ethical approval by the Psychology Research Ethics Committee of the University of Edinburgh (approval no. 14-1213/1).

7.2.2. Materials

Materials were identical to those used by Drake and Corley (2015b; see Chapter 6 of current thesis), with the amendment that in the present study we additionally paired sentence stems with pictures which had overlapping onsets. Twelve pictures served as experimental items. Each picture corresponded to a single-syllable name which paired one of two onsets (/t-/ or/k-/) with one of six rimes (/-æn, -æp, -eɪp, -eɪk, -əʊn, -əʊst/). A further two pictures were used as practice items. We used high-cloze 3 sentence stems which predicted each of the picture names (can, tan, cap, tap, cape, tape, cake, take, cone, tone, coast, toast). The 36 sentence-stems were recorded as spoken by a female speaker of British English (mean speaking rate = 3.92 syllables/second; mean sentence stem duration = 3.10 seconds, range = 1.90–5.29 seconds; see Appendix C).

7.2.3. Procedure

As in previous prediction experiments reported in this thesis, the experiment was presented on a laptop, using DMDX software (Forster & Forster, 2003). Subsequent to training on picture names (all participants achieved 100% accuracy), the experiment proceeded in five blocks. In blocks 1 and 5, pictures appeared in isolation following a fixation point which was displayed for 2.9 seconds, and participants named each picture aloud once. In blocks 2, 3, and 4, the fixation point was accompanied by the auditory presentation of a sentence stem (i.e., participants would hear "When we want water we just turn on the…" while the fixation point was visible). Immediately following the sentence stem, the fixation point was replaced with the picture to be named. In a third of cases in each block, the correct name matched the predicted word (e.g., "tap"- TAP); in a third of cases the name had an overlapping rime (e.g., "tap" - CAP); and in a third of cases the onset overlapped (e.g., "tap"-TAN). In each of blocks 2, 3, and 4, all 36 sentence stems were used, and each picture was presented 3 times. Combinations of sentence stems and pictures were counterbalanced such that each picture was only combined with a given sentence stem once across the 3 experimental blocks. Throughout the experiment, participants were instructed to name each picture as soon as they could, without sacrificing accuracy. We used Articulate Assistant Advanced (Articulate Instruments Ltd, 2012) to capture an acoustic record of participants' responses and a time-locked ultrasound record of movements of the midsagittal tongue contour). Ultrasound images were obtained via a probe secured directly against the under-surface of the chin using a proprietary headset (Articulate Instruments Ltd, 2008), at a rate of ~100 frames per second (see Chapter 5).

Acoustic data were manually tagged in Praat (Boersma, 2001) to indicate; (i) the onset of picture presentation; (ii) the acoustic burst of the onset consonant during picture-naming; (iii) the onset of voicing striation (i.e., vowel onset), and; (iv) the offset of the steady-state vowel.

7.3. Results

Recordings from 2 female participants were not analysed further. In the case of one participant, this was due to a (technical) failure to record sufficient numbers of ultrasound records. In the other, it was not possible to determine relevant timings from the acoustic record (due to experimenter error during microphone set-up at the beginning of the experiment). Each of the remaining 10 participants produced 132 picture names (24 in the control and 108 in the experimental conditions). Of the resulting 1320 recordings, 38 (2.9%) were discarded because of failures to properly record either audio or ultrasound (there were no differences across conditions in the proportions of recordings dropped: $\chi^2_{(3)} = 3.8$, $p = .28$). The remaining 1282 recordings were digitized to video at a rate of ~30 frames per second for further analysis.

7.3.1. Response latencies

We determined naming latencies manually, by measuring the time between the onset of the picture presentation and the acoustic burst associated with the onset of the picture name. When participants named pictures in the match (i.e., full-overlap) condition (mean RT = 479 ms; se = 25 ms) they were faster than in either the rime-overlap condition (mean RT = 693 ms; se = 38 ms) or the onset-overlap condition (mean RT = 674 ms; se = 33 ms). Mixed-effects regression using orthogonal contrasts with the maximal justified random effects structure (Barr, Levy, Scheepers, & Tily, 2013) showed that the mismatch (i.e., partial overlap) conditions resulted in significantly slower naming latencies than the match (i.e., full-overlap) condition (ß = 67.6; se = 8.9; t = 7.6). However, there was no reliable difference in naming latencies between the rime-overlap and onset-overlap conditions (ß = 6.8, se = 11.5; t = 0.6). This replicates the differences found previously for the relevant conditions (Drake & Corley, 2015a; 2015b; see also Chapters 3, 4, and 6 of the current thesis)

7.3.2. Ultrasound Analysis

The ultrasound analysis closely follows that reported in Chapter 6. First, we identified key moments during the participants' productions of each word by inspecting the audio channel using Praat (Boersma & Weenink, 2014). These were the acoustic release (burst) of the onset consonant, and the end of the steady-state vowel (striation) which followed. These time points were used to extract portions of each video for further analysis. Video extracts were expanded or contracted to a standardised number of frames using an averaging algorithm (see McMillan & Corley, 2010) to control for slight differences in video frame rate and in articulation timings. Each resultant frame of ultrasound video constituted a 512 × 277 grid of pixels. Pixels ranged in luminance from 000 (black) to 255 (white). In order to achieve data tractability, we processed each frame so that luminance was averaged over blocks of 8 × 8 contiguous pixels (see McMillan & Corley, 2010). A vector was generated from each frame, with each 8 × 8 pixel block assigned a specific position in the vector. Each vector ran from bottom left to top right of the video frame. Vectors formed the basis for analyses, which were performed by calculating and comparing Delta scores (McMillan & Corley, 2010): i.e., the Euclidean distances between individual vectors (frames).

In order to minimise the effects of noise in the ultrasound images (see Wrench & Scobbie, 2008) we performed a preparatory analysis in order to determine the quality of the data acquired from each participant for each CV onset. This analysis was performed on ultrasound data acquired between the acoustic burst and the end of the steady-state vowel for each token; subsequent analyses were performed on data acquired prior to the acoustic release of the onset consonant. We used multidimensional scaling (Mardia, 1978) to calculate how well the Delta scores distinguished tokens of a given CV onset from tokens of all other CV onsets produced by that participant. This was

achieved by determining the mean Euclidean distance of a given vector from: (a) all vectors representing different CV onsets; (b) all vectors representing the same CV onset. The Discrimination score for each onset for each participant was equal to (a)/(b). Therefore the higher the score the better the data discriminated between a given CV onset and others in the picture-name set, and the less noisy the data. This information was used to geometrically weight the contribution of each participant's data to subsequent analyses (Carroll & Ruppert, 1988). In this way we were able to avoid arbitrarily discarding poor quality data, whilst accounting for the great by-participant variability known to be associated with ultrasound articulatory data.

We used a mixed modelling approach, implemented in R via the lme4 package (Bates, Maechler, Bolker, & Walker, 2014). In line with common practice we report effects as significant where $|t| > 2$. In all analyses we modelled Delta scores as the outcome variable, and included Condition (Match, Onset Overlap, Rime Overlap) and Onset consonant (/k/, /t/) as fixed effects, and Participant and Picture-name (i.e. item) as random effects.

7.3.2.1. Location of articulation analysis

This analysis was performed on articulatory data acquired between -500ms and 0ms of the consonant acoustic burst. Data acquired during this period were collapsed to produce one average-luminance vector per token. For each item, a reference vector was generated by averaging across all Control productions of that item. This allowed us to calculate Delta scores which expressed the degree to which tokens produced in the experimental conditions (Full Overlap, Onset Overlap, Rime Overlap) differed from those produced in the Control condition.

Differences between individual articulations and mean control articulations were then modelled as the response variable in a linear mixed model (for details of predictor variables and model structure see above). Delta scores in the Full Overlap condition were significantly lower than those in the partial overlap conditions ($\beta = 1.687$, $t = 2.108$). Delta scores in the two partial overlap conditions (Onset and Rime) did not differ significantly ($\beta = 0.506$, $t = 0.364$). This indicates that, in line with our prediction, articulation in the Rime Overlap condition differed from that in the Full Overlap condition. However, contrary to the prediction of a speech-sound specification hypothesis, articulation in the Rime Overlap condition (where there was onset competition) did not differ from that in the Onset Overlap condition (where there was not onset competition).

In order to investigate whether tokens in the Rime Overlap condition exhibited traces of articulatory interference, we generated Delta scores that expressed the degree to which tokens produced in the experimental conditions differed from the Control reference vector for their onset competitor (i.e., rather than comparing tokens of "take" to the reference vector for "take" as above, we compared them to the reference vector for "cake"). These Delta scores did not differ by condition (in all cases $|t| < 1$): We did not observe a phoneme-specific articulatory interference effect.

7.3.2.2. Time-course analysis

This analysis was performed on all ultrasound frames acquired between -1000ms and 0ms of the onset consonant acoustic burst for each token (i.e., 31 frames per token). Within each token we calculated Delta scores for all inter-frame transitions over the time-course. Higher delta scores indicated greater frame-to-frame change associated with greater change in tongue configuration. Delta scores were automatically averaged and plotted by condition (Full, Onset, and Rime Overlap; see Figure 7.3.1).

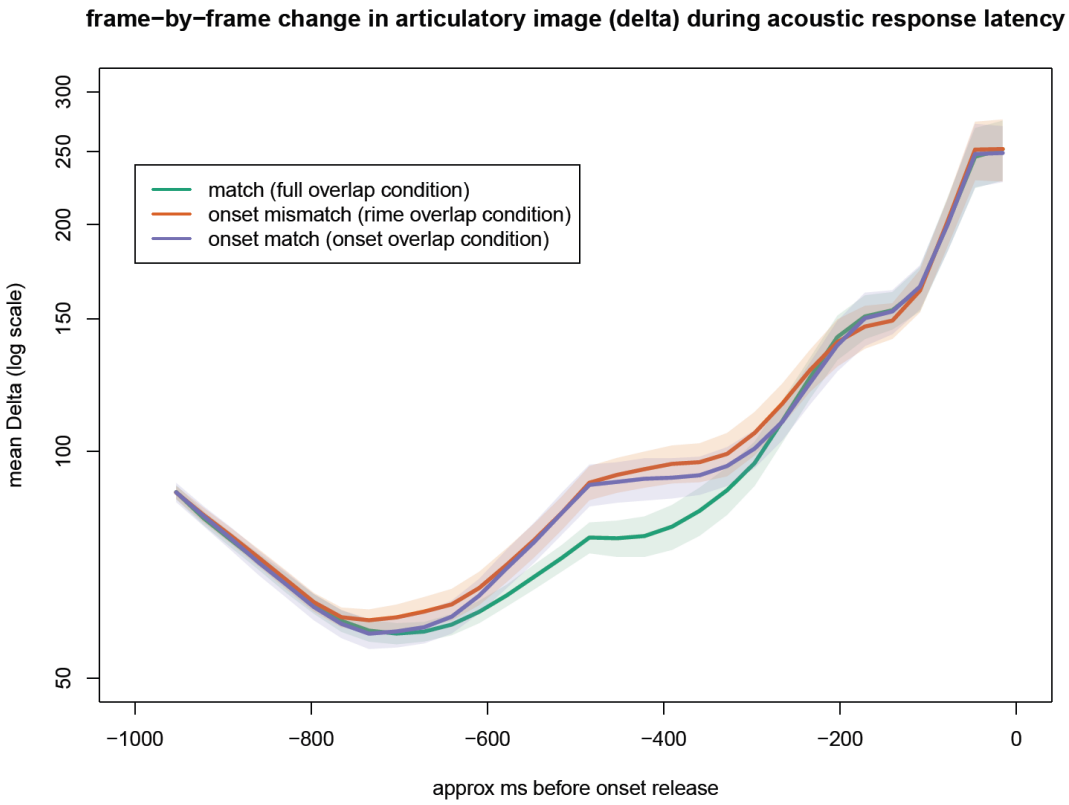

Figure 7.3.1: Frame-to-frame change in ultrasound tongue image during pre-acoustic articulation

The x-axis represents time milliseconds prior to the acoustic response (i.e. the response latency period between picture presentation and acoustic picture naming). The y-axis represents the amount of change in the ultrasound image between two consecutive frames of the ultrasound record (averaged by condition). This can be *loosely* interpreted as the extent of articulatory movement between frames. Solid lines indicate the by-condition means at each time point. Faint colour bands indicate 95% confidence intervals for the by-condition means.

We modelled Delta scores at each inter-frame transition via mixed-effects models comparing productions in the Onset and Rime Overlap conditions to those in the Full Overlap condition. We treated effects as significant only when they clustered across three or more consecutive time-points (Lage-Castellanos, Martínez-Montes, Hernández-Cabrera, & Galán, 2010). Effects of condition were

found to be statistically significant (i.e., |t| > 2) for both consonant onsets (i.e., /k/ and /t/) between -500ms and -300ms. Effects were also significant for /t/ onset items in the time window -700ms to -500ms. Significant effects were observed over a longer time period in the Rime Overlap condition than in the Onset Overlap condition, but the two conditions pattern similarly, with consistently greater frame-to-frame movement in these conditions than in the Full overlap condition. This means that, as in the Location of Articulation analysis, we observed an articulatory effect of mismatch between the lexical prediction and the picture name. We did not find evidence that this effect was confined to situations in which there was onset competition between the predicted word and the picture name.

7.3.3. Lexical analysis

The primary manipulation of interest across the experiments reported in this chapter was, as stated in the Introduction, the nature of the phonological overlap between the predicted word and the target picture name noun. However, as in the response latency experiment reported in Chapter 4, some sentence-stems predicted homophones of picture name target nouns rather than the nouns themselves (e.g., *"There's no such word as can't. You have to believe that you…"* CAN; "*They raised their glasses in a…*" TOAST). As discussed in Chapter 4, in such cases the picture name overlaps fully with the prediction at a phonological level but not at a semantic level, and is therefore a homophone of the prediction rather than a full match.

In Chapter 4 we reported that participants were quicker to name pictures when they were a full match for the prediction than when they were merely a phonological match (i.e., a homophone). However, homophones were still named more quickly than pictures that overlapped partially at a phonological level (i.e., in the rime and onset overlap conditions, referred to in this chapter as the mismatch and onset overlap conditions respectively). This confirmed that predictions are active in some way at a word-form level. Results reported in the current chapter suggest that articulatory differences between the phonological match and mismatch conditions are not affected by the type of segment level overlap between a predicted word and a to-be-named picture. It therefore appears that articulatory differences reflect conflict at whole syllable level and/or at a more general item level. Homophones conflict with predictions at a general level, but do not conflict with predictions at a syllable level. We further analysed the articulatory data acquired in the study reported in the current chapter in order to investigate whether articulation of homophones differed from that when pictures fully matched the prediction. Evidence of greater articulatory movement in the homophone condition would suggest that the conflict we observed in articulatory records from both partial overlap conditions arose, at least in part, at a general (semantic) level.

We performed a supplementary analysis to investigate whether there was an articulatory difference between the homophone and full match conditions. This analysis included data acquired in the full overlap context only (i.e., where the picture name fully overlapped with the predicted item at a phonological level; see Figure 7.3.2), although we include plots of data acquired in the mismatch and

onset overlap conditions for comparison. As in Chapter 4, for each sentence-stem, we determined whether the high-cloze target had the same meaning as the picture it was encountered with in the match condition (e.g., *"The fire alarm's gone off again; someone must have burnt the …"* TOAST) or whether it shared only phonological form (e.g., *"They raised their glasses in a …"* TOAST). Where the sentence fragment predicted the picture name lemma in the match condition the trial was considered to be a "name"; where only the phonological form matched the trial was considered to be a "homophone". Of the 36 experimental sentence-fragments, 11 (31%) formed category mismatches. Only the full overlap context allows the predicted item to match the picture name at a lexical level.

The analysis approach is identical to that described above to compare articulation by condition, with the exception that all data included now came from the full-overlap condition, and the comparison was between "homophones" (phonological match) and "names" (full match).

We modelled Delta scores at each inter-frame transition via mixed-effects models comparing productions in the Homophone condition to those in the Name condition. As described above, under this approach we treat effects as significant only when they clustered across three or more consecutive time-points. Movement was found to be statistically greater (i.e., $|t| > 2$) in the Homophone condition than in the name condition in only one time-point model ($\beta = 4.01$, $se(\beta) = 1.91$, $t = 2.10$), with the following time-point model approaching significance ($\beta = 3.78$, $se(\beta) = 1.90$, $t = 1.99$). These models include data covering a time period from approximately 810ms to 750ms prior to the acoustic onset (refer to Figure). The failure to achieve significance over 3 or more time points means that we do not interpret this as indicative of a reliable difference in mean articulatory movement between Homophone and Name conditions. This data set therefore does not provide evidence that articulatory interference arises as a result of mismatch at a general semantic level.

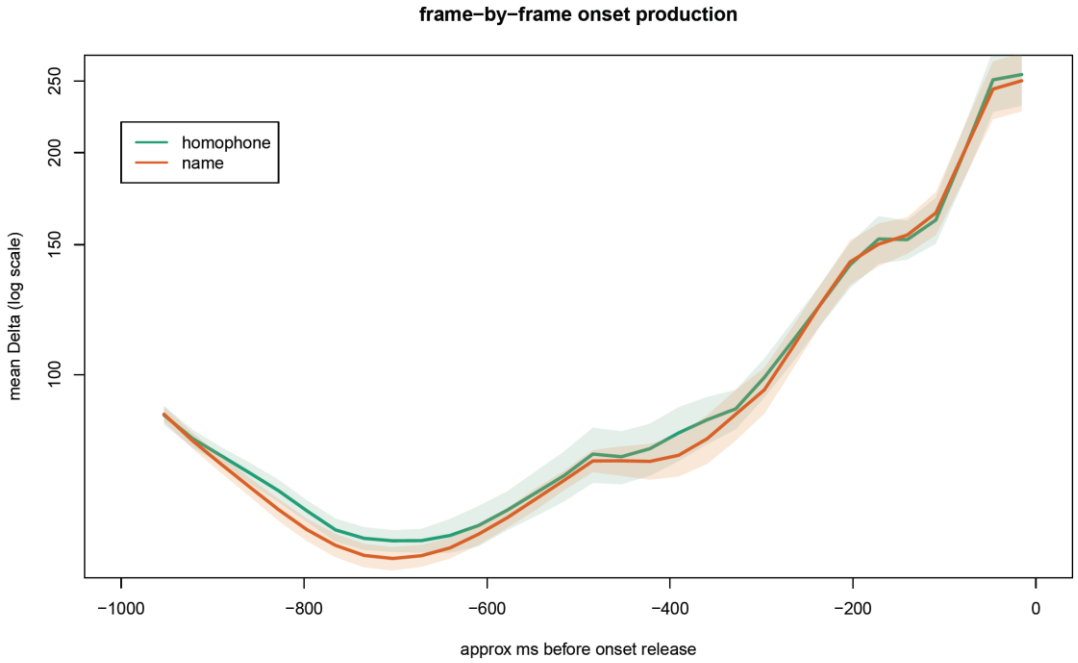

Figure 7.3.2: Frame-to-frame change in ultrasound tongue image during pre-acoustic articulation in full-overlap condition (homophone analysis)

The x-axis represents time milliseconds prior to the acoustic response (i.e. the response latency period between picture presentation and acoustic picture naming). The y-axis represents the amount of change in the ultrasound image between two consecutive frames of the ultrasound record (averaged by condition). This can be *loosely* interpreted as the extent of articulatory movement between frames. Solid lines indicate the by-condition means at each time point. Faint colour bands indicate 95% confidence intervals for the by-condition means.

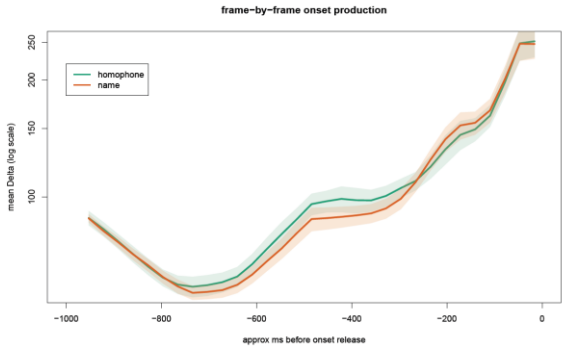

Figure 7.3.3: Frame-to-frame change in ultrasound tongue image during pre-acoustic articulation rime-overlap condition (homophone analysis)

The x-axis represents time milliseconds prior to the acoustic response (i.e. the response latency period between picture presentation and acoustic picture naming). The y-axis represents the amount of change in the ultrasound image between two consecutive frames of the ultrasound record (averaged by condition). This can be *loosely* interpreted as the extent of articulatory movement between frames. Solid lines indicate the by-condition means at each time point. Faint colour bands indicate 95% confidence intervals for the by-condition means.

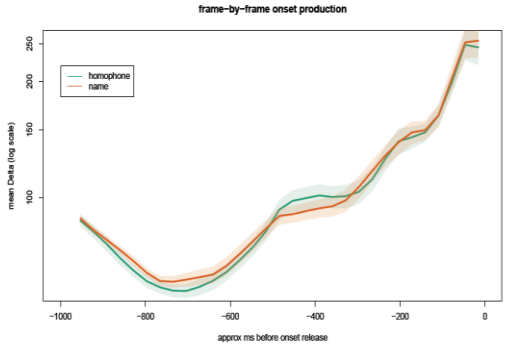

Figure 7.3.4: Frame-to-frame change in ultrasound tongue image during pre-acoustic articulation onset-overlap condition (homophone analysis)

The x-axis represents time milliseconds prior to the acoustic response (i.e. the response latency period between picture presentation and acoustic picture naming). The y-axis represents the amount of change in the ultrasound image between two consecutive frames of the ultrasound record (averaged by condition). This can be *loosely* interpreted as the extent of articulatory movement between frames. Solid lines indicate the by-condition means at each time point. Faint colour bands indicate 95% confidence intervals for the by-condition means.

7.4. Discussion

We reported a study in which we used an automated approach to the analysis of ultrasound tongue imaging data in order to investigate whether comprehension-elicited lexical predictions have articulatory consequences. Participants named pictures in one control condition and three experimental conditions. The experimental conditions differed with regard to the extent that the picture name overlapped with a predicted word at a phonological level. Lexical predictions were elicited by auditorily presenting participants with high-cloze sentence-stems (i.e. via comprehension). Of specific interest was whether comprehension-related predictions elicit cascade from a phonological to a motor-speech level as do representations activated during speech production and speech listening.

Effects of prediction were observed in articulatory data acquired prior to the acoustic onset of picture naming. When data were collapsed over the 500ms preceding acoustic onset, articulations were more similar to the control productions when there was full overlap between the predicted word and the picture name than where there was only partial overlap. This indicates that lexical predictions elicited via comprehension can have articulatory consequences.

The findings of previous error-elicitation studies led us to predict that if motor-speech activation during comprehension reflects the representation of upcoming material at an abstract gestural level we would observe interference from a competing representation at an articulatory level. We therefore investigated whether tokens produced in the Rime Overlap condition, where the lexical prediction

would activate a competing onset representation, were more similar to articulations of the competing word than were those in the Onset Overlap. We did not find evidence to support this interpretation.

We used a time-course analysis to further investigate the nature of the effects on articulation of comprehension-elicited predictions. This analysis approach revealed that effects are seen only at a relatively early stage during the pre-acoustic phase. During this early time-window, frame to frame change is greater in conditions where the picture name does not match the predicted word. However, the degree and pattern of frame to frame change did not appear to differ according to whether there was phonological conflict at word onset: Articulation in the Onset Overlap condition (in which there was no conflict at word onset) differed from articulation when there was a full-word match but not from articulation when there was conflict at word onset. This suggests that articulatory effects may arise at a whole-word level. It should be noted, however, that picture names were all monosyllabic so it is not possible to distinguish between effects arising at a whole syllable level and those arising at a whole word level.

The suggestion that articulatory effects arise at a whole word level is compatible with the early time-window at which effects are observed, and with the failure to find evidence of phonological interference effects. The apparent non-specificity of the articulatory effects observed raises the possibility that the articulatory consequences of the lexical predictions reflect general conflict monitoring and resolution processes (possibly an articulatory correlate of neural error-related negativity).

Chapter 8: Conclusion

This thesis reports a series of studies that explore how the production system might be involved in prediction during comprehension, specifically during prediction at a speech-sound level. The first part of the thesis investigates whether the effects on speech production of sententially-induced lexical predictions are comparable to the effects of perceptually presented distractors. This question is central because studies implicating the production system in comprehension at a speech sound level have so far investigated neural speech-motor activation only in *response* to *perceptually presented* stimuli. If speech motor representations subserve prediction, they must be evoked *prior* to the point at which the relevant sensory stimulus is presented (see Pickering & Garrod, 2013, p.106). Their effects should therefore be observable at a speech motor level in the absence of an external stimulus. Whilst EEG data has provided evidence consistent with the view that participants predict upcoming material at a speech sound level (e.g., DeLong et al., 2005), this has not been shown to relate to speech motor representations. The response time study reported in Chapter 2 of this thesis used a PWI paradigm to confirm that our materials elicit the typical speech sound effects expected when distractors are made perceptually available. However, in the studies reported in Chapters 3 and 4, where distractors were predicted rather than presented, response latencies did not reveal evidence of similar speech sound effects.

In the second part of this thesis, we reported two studies in which articulatory imaging was employed in order to directly investigate whether speech-sound predictions are represented at a speech-motor level. Speech motor movement during the response latency period was captured via ultrasound imaging. Data were analysed using an extension of the Delta method approach (McMillan & Corley, 2010). The extended Delta approach developed in this thesis allowed speech production to be investigated as a dynamic motor process rather than as single-point acoustic goal realisation. Speech motor movement was tracked over the response latency period. This approach allowed investigation of *how* the acoustic goal associated with voice-key triggering is realised as well as when. As stated above, the response latency findings reported in this thesis did not provide evidence that segment-like[13] representations are activated within the speech production system during comprehension-based prediction. The articulatory findings indicate that predictions elicited during spoken language comprehension do have speech-motor consequences. However, it appears that these speech-motor consequences are not specific to overlap between a predicted sound segment and the segment-to-be-articulated.

In the current chapter we consider our response time and articulatory findings with reference to recent exegeses of the prediction-as-simulation framework proposed by Pickering and Garrod (2004; 2007; 2013; see also Gambi & Pickering, 2016). We refer to neurobiologically-inspired and phonologically-

[13] we use the term "segment" to refer to representations specified at a conventional phoneme level

informed models of speech processing which incorporate the notion of forward models and multiple levels of phonological representation. In particular, we consider our findings with reference to Hickok's instantiation of a Hierarchical State Feedback Control model of speech production (Hickok, 2012; 2014).

With respect to prediction as a production activity during comprehension, the Pickering and Garrod framework assumes that the listener can covertly imitate the speaker's utterance, thereby deriving a production command associated with that utterance. The derived production command then allows the listener to use the forward modelling process that s/he uses when producing her/his own speech in order to predict upcoming (speaker generated) sensory input via the forward model (Gambi & Pickering, 2016; Pickering & Garrod, 2013). This process is referred to as "prediction-by-simulation"[14]. PWI and sentence completion studies are cited by Pickering and Garrod (2013) as sources of evidence that comprehension and production are finely interwoven and that a split between the two should be challenged. Combining features of the two paradigms in order to create a picture prediction interference paradigm allowed us to investigate whether the predictive mechanism employed during sentence listening resembles the prediction-by-simulation process described by Pickering and Garrod. If participants do indeed use a production command to predict during comprehension, we would expect such predictions-of-another to affect production in the same way as do other competing representations within the production system, such as those evoked during error elicitation and PWI tasks.

As reviewed and demonstrated in Chapter 2 of this thesis, the picture-word interference effect is reduced when the distractor shares initial or final speech-sounds with the picture-to-be-named; the relative reduction in naming latencies is generally referred to as a phonological facilitation effect. The effect indicates that distractors are subject to representation at a speech-sound level in a form that contacts the speech production system. If predictions were also represented at a speech-sound level within the speech production system we would expect to see a comparable effect. We did not see such an effect. We consider this finding with reference to the primary model relating to reaction-time data, WEAVER++ (see Strijkers et al., 2013 for comment on Pickering and Garrod's reference to this model).

Under WEAVER++ (Roelofs, 1997; 2002), when a segment is retrieved as part of a lemma's associated phonological code during word *production*, activation spreads to all the syllabic scores with which it is associated. For example, during naming of a picture CAP, the morpheme <cap> is activated following selection of the lemma *cap* (along with its relevant syntactic diacritics, e.g., noun, singular). Activation of the morpheme <cap> elicits retrieval of the segments /k, a, p/, which are

[14] With respect to prediction of another's speech, the framework provided by Pickering and Garrod (2013) differs from earlier proposals made by the same authors: Earlier work (e.g., Pickering & Garrod, 2007) proposed that a comprehender produces fully-specified predictions representing what she would say herself if she were the speaker.

concatenated to form the syllable score /kap/. However, retrieval of the segments /k, a, p/ also causes activation to spread to other syllable scores that link to one or more of those segments (e.g., /kan/, /kat/, /kab/, /tap/, /map/, /lap/). In this way, retrieval of a phonological code during production planning will always lead to the activation of multiple syllable scores (although under a discrete model only one phonetic score is forwarded to articulatory control processes; see Cholin et al., 2006, p. 209).

Under WEAVER++, phonological facilitation in the PWI paradigm is achieved indirectly, via activation within the *perceptual* system: The distractor word is processed within the perceptual network, activating relevant phonological segment nodes. This leads to the activation of corresponding segment nodes within the production network. When these corresponding segment nodes are components of the target picture name, its phonological code spell out is speeded up. This ultimately reduces the time lapse between picture presentation and syllable program retrieval. To take an example from the study reported in Chapter 2: a distractor word "cap" activates the segments /k, a, p/ within the perceptual system. This leads to the activation of corresponding segments within the production network. The activation of these segments leads to the activation of the syllable progams to which they link (for example, /kan/, /kat/, /kab/, /tap/, /map/, /lap/). Picture naming latency to such phonologically overlapping words is therefore reduced, although articulation can be initiated only once all the syllabic scores of an intended phonological word have been retrieved. The finding of phonological facilitation reported in Chapter 2 would be accounted for in this way.

WEAVER++ also allows a direct production route to phonological facilitation, which, unlike the indirect route, does not require that a stimulus be perceptually available. This route was developed to account for evidence that the phonological segments associated with multiple candidate lemmas can be activated in parallel within the speech production system (Peterson & Savoy, 1998; see also Jescheniak & Schriefers, 1998; see also Miozzo, 2002; Navarette & Costa, 2005; Meyer & Damian, 2007 for evidence from picture-picture naming paradigms). Under Pickering and Garrod's prediction-as-simulation framework, predictions are generated within the production system; phonological facilitation would therefore be expected to arise via a direct production route. However, as reported in Chapters 3 and 4, we did not observe a phonological overlap effect when distractor words were predicted rather than presented.

Although predictions did not elicit phonological facilitation, homophone naming was facilitated. This suggests that lexical predictions generated during comprehension are in some way represented at a speech form level within the production system, although it does not in itself speak to the issue of whether speech motor representations are involved. Homophone facilitation in the absence of other phonological facilitation may be accounted for by feedback from a speech sound to a lexical level. Alternatively or additionally, it may indicate that the proximal planning unit involved in prediction differs from that involved in perception-production: prediction may involve syllable-sized units whereas perception allows segment-size units. Pickering and Garrod's speech processing framework

does not specifically address the possibility that speech sounds may be represented at multiple levels. However, the interpretation suggested above is in line with the speech processing architecture outlined in Hickok's Hierarchical State Feedback Control model of speech processing (HSFC; Hickok, 2012; 2014). We return to this matter following discussion of our articulatory imaging findings.

As reviewed in Chapters 5 and 6, articulatory imaging has been previously been employed to investigate the effects of phonological competition during speech *production*. Studies have focused on competition that arises as participants plan and execute their own speech (as opposed to the potential "planning" of others' speech suggested under the prediction-as-simulation proposal). The phonological competition investigated via error elicitation paradigms is understood to arise of the same process that underlies the direct-route phonological facilitation modelled in WEAVER++; segments associated with multiple candidates are simultaneously active at a phonological level. Activation at lexical and phonological levels cascades to a speech motor level, resulting in multiple candidates being simultaneously active during response execution (see McMillan, 2008; McMillan & Corley, 2010; Goldrick & Chu, 2014). It has been suggested that such cascade from a planning to execution involves forward modelling of the type invoked under the prediction-as-simulation proposals made by Pickering and Garrod (2013; Dell, 2013).

In the speech imaging studies reported in this thesis, the control condition required participants to name pictures presented in a non-predictive context, whereas experimental conditions required participants to name pictures following a sentence fragment that strongly predicted a specific lexical item. When the prediction and picture name competed at word onset (e.g., cap TAP) articulation differed more from the control condition than when the prediction and picture name matched (e.g., cap CAP). This articulatory interference, reported in Chapter 6, is in keeping with previous findings concerning *production* induced competition, and suggested that predictions generated during comprehension do involve speech motor level representation as proposed under Pickering and Garrod's prediction-as-simulation account. However, the design of the experiment reported in Chapter 6 did not allow us to distinguish whether articulatory interference during the response latency period reflected conflict between onset segments (as in previous studies of production induced competition) or between whole syllable units (as suggested by the facilitation of homophone naming revealed in the response latency study reported in Chapter 4). In Chapter 7 we conducted a further study, in which we extended the design to include predictions that matched the picture name at word onset (e.g., cap – CAN), allowing us to investigate whether articulatory interference reflected competition between onset segments: If predictions were represented as phonological segments encoded within the production system, we would expect articulatory interference to be reduced in the onset overlap condition (in which prediction and picture name match at onset) compared to the rime condition (in which prediction and picture name compete at word onset). In fact, statistical comparisons of articulatory movement over time revealed that articulation in the onset overlap condition was indistinguishable from that in the rime overlap condition, with both conditions showing greater articulatory movement than the full match condition.

Our evidence therefore suggests that predictions elicited during comprehension are not represented at a segment level within the speech production system but that predictions are activated and do interfere at a speech production level. As reported in Chapter 7, we did not find reliable evidence of articulatory interference when predictions conflicted with the picture name at a semantic level but fully overlapped at a phonological level (i.e., in the case of homophones). The fact that we observed articulatory interference only when predictions conflicted with picture names at a speech sound level favours an account in which articulatory interference reflects processing at a speech sound level of representation. However, the failure to observe evidence in articulation of conflict at a semantic level cannot be taken as evidence that conflict at this level does not contribute to articulatory interference. We observed reliable evidence of articulatory interference between 500 and 300 milliseconds prior to the acoustic response. This falls within the time frame typically ascribed to semantic/syntactic processing (see Indefrey & Levelt, 2004; see also Indefrey, 2011) as opposed to phonological or phonetic processing (lexical access begins within 200ms of picture presentation; see Strijkers & Costa, 2011, for a review). We therefore consider this account before returning to the topic of types of speech sound representation.

Two aspects of our data militate against a (purely) semantic account: Firstly, accounts that estimate semantic/syntactic processing at the time point that we observe articulatory interference are serial and do not allow articulatory involvement at this point. Whilst our data contrast with the assumptions of such accounts, they are in keeping with the findings of other articulatory studies: Myogenic activity associated with speech production has been reported to precede the acoustic onset of speech by up to 500ms (see Brooker & Donald, 1980) and to be observed as early as 250 ms post stimulus onset (Riès et al., 2012; 2014). Movement-related cortical potentials can be observed up to 600ms before mouth opening or phonation (Yoshida et al., 1999; Tremoureux et al., 2014; Galgano & Fround, 2008). These timings are consistent with the finding that motor-associated neural activity is commonly observed by 300ms post stimulus presentation (see Laganaro, 2016, for review figure). Our data are also in keeping with questions concerning the sequential architecture of speech production assumed in the studies that informed the estimates cited above (e.g., Strijkers, 2016; Munding, Dubarry, & Alario, 2016).

The second aspect of our data that militates against a purely semantic account of the prediction-related articulatory interference effect concerns the processing of homophones (see Chapters 4 and 7). Predictions facilitated homophone naming despite being incongruent at a semantic level. Homophone naming did not show the articulatory interference effect observed where there was only partial phonological overlap between the picture name and the prediction. As noted previously, these findings suggest that predictions do make contact with the speech production system at a word form level. This conclusion stands whether one assumes that homophones share a single word form representation, or that feedback from a unique word form representation of the prediction leads to lemma level activation of the homophone and subsequent feedforward activation of the picture's unique word form (see Severens, Ratinckx, Ferreira, & Hartsuiker, 2008, for a review): in both cases

word form activation is necessary. Therefore, when relating our findings to the proposal that prediction-by-simulation involves predictive activation of the speech motor system, the question of interest concerns how word form predictions might map to activation within the speech motor system. This is particularly pressing as our data suggest that word form predictions do not appear to be represented at a speech *segment* level within the listener's speech-motor system.

Conventionally, models of both speech production and speech perception have placed considerable emphasis on the phoneme as a representational unit essential to speech processing. Under interactive models of speech production, information contained within phoneme representations is transformed into motor execution via representation at the feature level (e.g., MacKay, 1987; Dell et al., 1993; Kawamoto et al., 1998). Alternatively, under WEAVER++ (referred to above), phoneme level representations allow resyllabification, which then allows activation of the relevant phonetic syllable with its associated motor program. Both modelling approaches allow motor execution to be realised more quickly when segment level representations receive relatively greater activation (e.g., the phonological facilitation effect in PWI). Under both types of model, we would expect to see phonological facilitation effects in response latencies if predictions made contact with the production system at a form level. Phonological facilitation was not observed in any of the seven prediction experiments reported in this thesis.

Under WEAVER++, phonetic syllables are informationally encapsulated, and therefore overlap at the more abstract segment level would not be expected to reduce the articulatory interference effect. This is in keeping with our articulatory findings. However, models that treat phonetic syllables are informationally encapsulated are not compatible with articulatory data acquired in production studies (such as those reviewed above and in Chapter 5), which do show articulatory effects of cascade from segment level planning processes (e.g., McMillan, 2008; McMillan & Corley, 2010; Pouplier, 2007; Pouplier & Goldstein, 2012).

To summarise the discussion so far: The articulatory interference we observed in this thesis cannot be fully explained as arising of conflict between predictions and picture names at a semantic level. It is observed when there is conflict at a syllable level, but does not appear to operate at a speech segment (i.e., phoneme) level. We therefore now consider whether our findings can be understood under an account that specifically posits multiple units of speech sound representation. We choose to refer to the Hierarchical Sate Feedback Control model proposed by Hickok (HSFC, 2012; 2014) due to the fact that it shares key architectural properties with the framework under which Pickering and Garrod make their prediction-as-simulation proposal; both frameworks incorporate forward models and seek to integrate production and comprehension processes. We note however that Hickok himself has queried the likelihood that prediction involves activation of the speech motor system (see Hickok, 2014).

The HSFC is primarily "cast" at a phonological level. A word representation is activated by a conceptual representation. A (neurobiologically) ventral stream links environmental sounds (e.g., heard speech) to conceptual representations, whilst a (neurobiologically) dorsal stream links a motor movement (e.g., speech articulation) to the sound produced. "Auditory phonological representations" are central, in that they act both as targets for speech motor activity and access codes to lexical-conceptual representations (i.e., an interface between dorsal and ventral stream processes). As a word becomes more frequently uttered, motor 'presets' emerge via Hebbian learning. This process allows the formation of a "parallel, direct link" between the lexical-conceptual systems and motor-phonological codes, allowing the lexical-conceptual system to activate the motor and sensory streams of speech production in parallel. In this way, activation of a lexical-conceptual representation invokes parallel input to both motor-phonological and auditory-phonological components of the phonological processing system Under this account we would anticipate that, should predictions be represented within the production system, lexical-conceptual activation of predicted items would result in parallel input to motor- and auditory-phonological components of the phonological processing system.

Phonological (speech sound) representations are distributed over higher and lower level loops and ventral and dorsal processing streams: Input from the lexico-conceptual system enters the phonological system at a high-level loop (auditory- Spt-BA44). From this loop there is parallel projection to a lower-level somatosensory-cerebellum-motor cortex loop. Ventral and dorsal streams interface through the loops, with the auditory and somatosensory elements being components of the ventral stream, and motor elements being components of the dorsal stream. Crucially, with respect to the findings reported in this thesis, the higher level loop involves syllable sized units (an auditory syllable target and a motor syllable program) whereas the lower level loop involves phoneme (or "feature cluster") sized units. Our articulatory data could therefore be explained under an account that saw predictions as being represented within the integrated speech processing system at a higher-loop level but not at a lower loop level. We consider whether such a situation could arise under the HSFC model.

Within the HSFC model, input from the lexical-conceptual level to both higher level phonological components is excitatory, input from the auditory-phonological level to the motor-phonological-level is excitatory, but input from the motor-phonological to the auditory-phonological level is inhibitory. Therefore when input to the motor-phonological level matches input to the auditory-phonological level, all sensory activation is cancelled, meaning that there is no prediction error and excitation proceeds to the lower loop level at which the unit of representation is the segment rather than the syllable. If input to the motor-phonological level does not match that to the auditory-phonological level the resulting prediction error allows for correction at the syllable level. However, because auditory-phonological input to motor-phonological representations is excitatory, activation of an auditory-phonological representation alone is sufficient to elicit activation of a corresponding motor-phonological representation. In the absence of a gating system between higher and lower level dorsal (i.e. motor) representations, this would lead listeners to continually shadow heard speech. Hickok

therefore proposes a basal ganglia based gating system at this point, which is in place during speech comprehension but open during speech production; this has been successfully computationally modelled on a small scale (Hickok, 2012a).

The architecture proposed by Hickok is therefore compatible with our articulatory and response latency findings: Predictions can be understood as being represented within a multimodal phonological system. Within a higher-level loop, which acts on syllable-units, activation from a lexico-conceptual source leads to activation within the dorsal (motor) stream. However, this activation does not proceed to the lower-level loop, which acts on segment-units, due to a gating system that is active during comprehension activities. As higher level motor representations of predictions are active during processing necessary for picture naming these predictions interfere at an articulatory level. Interference operates at a syllable level but not a segment level because the prediction has not been subject to representation at the segment level. We therefore suggest that the HSFC model offers a valuable framework under which to consider the involvement of motor stream representations during comprehension related prediction.

We previously noted that, when taken in their entirety, the findings reported in this thesis are not compatible with the suggestion that prediction-as-simulation involves a production system as modelled under traditional discrete or cascaded approaches. We suggested that the HSFC provides an alternative model under which to consider the proposal that predictions contact the speech production system at a speech sound level. We must now consider whether the HSFC model allows: (i) phonological facilitation at the segment level as observed in perceptual PWI studies; and (ii) segment level articulatory competition as observed in previous articulatory imaging studies. The question in both cases is whether the experimental conditions would lead to the activation of segment-sized representations under the HSFC model. It appears that in both cases the answer is yes: In the perceptual PWI paradigm, phonological information would enter the system via the lower level loop. This loop is associated with segment-units, and we would therefore expect facilitation to operate a segment level. In error elicitation paradigms associated with articulatory evidence of phonological cascade, both target and distractor items are intended for production. Therefore neither item would be expected to be subject to activation gating. This would allow both items to be activated at the lower level loop associated with segment-unit representation.

8.1. Concluding remarks

Adopting an ultrasound imaging approach allowed us to examine speech as action, thereby providing a means to investigate the prediction-as-simulation account proposed by Pickering and Garrod (2004; 2007; 2013). The articulatory evidence presented in this thesis is consistent with the proposal that speech motor representations are activated during comprehension-elicited prediction. However, motor speech representation arising during prediction of another's speech appears to differ from that arising during speech perception and speech production: It appears to involve activation at a syllable level

but not at a segment level. This difference can be understood under the HSFC model (e.g., Hickok, 2012) which, although it was not developed to model a prediction-as-simulation process, incorporates the necessary specificity concerning the relationship between speech sound and speech motor representations.

Appendix A: PWI studies summary

Study	Target	Distractor			Outcome measure		Variables					Finding
							Phonological relatedness		Semantic relatedness		SOA (ms)	
	Picture name	written	acoustic	picture	Voice response latency	other	onset	end	Co-ordinate	associate	0 / Other (ms)	
Alario, Sequi, & Ferrand, 2000	x		x		x				x	x	114 234	RT: PR < UR
Abel, Dressel, Bitzer, Kümmerer, Mader, Weiller, & Huber, 2009	x		x		x	fMRI BOLD	x				x	RT: PRb < UR
Ayora, Peressotti, Alario, Mulatti, Pluchino, Job, & Dell'Acqua, 2011, Exp 1	x		x		x		x	x			x	RT: PRe < UR
Briggs & Underwood, 1982, Exp. 1	x	x			x		x	x			x	RT: label < homophone < control < unrelated
Collins & Ellis, 1992, Exp. 2	x		x									RT: PRb < UR PRe < UR
De Zubicaray, McMahon, Eastburn, & Wilson, 2002	x	x			x	fMRI BOLD			x			RT: Control < PR < UR BOLD: PR < UR and SR in the lSTG, l IFG and rpSTG was
De Zubicaray & McMahon, 2009	x		x		x	fMRI BOLD	x		x			RT: Control < PR < UR < SR BOLD: PR < UR/SR in LH regions associated with lexical-conceptual and phonological processing
Damian & Bowers, 2009	x	x	x		x			x			x	-100 100 200 RT: PRe < UR Effects greatest at 0ms and 100ms SOA
Damian & Martin, 1999	x	x	x		x		x		x		x	x SR > UR at -ive SOAs PRb < UR at +ive SOAs Both types of effect observable with both types of distractor at 0ms SO.
Glaser & Düngelhoff, 1984	x	x			x		x		x		x	RT: Control < PRb < SR

Study	Target	Distractor			Outcome measure		Variables						Finding
							Phonological relatedness		Semantic relatedness		SOA (ms)		
	Picture name	written	acoustic	picture	Voice response latency	other	onset	end	Co-ordinate	associate	0	Other (ms)	
Jescheniak, Schriefers, Garrett, & Friederici, 2002 (delayed naming task)	x		x		.	ERP	x						ERP: Reduced negativity in PR than UR in 250-400ms and 400ms-800ms time windows. SR only differed from UR in later time window.
Lupker, 1982 (Exp 2)	x	x			x		x		x		x		Picture alone < PRe < unrelated
McQueen & Huettig, 2014	LDT	x		x		Manual response latency	x		x				SR < UR < PRe (NB this is a lexical decision task) Implies pictures and written words evoke a phonological form representation, which acts as a competitor to the auditory item and slows LDT
Meyer & Damian, 2007	x			x			x	x			x		Homophones < PRb < UR Pre < UR
Meyer & Schriefers, 1991	x		x				x	x			x	-300 to +150	RT: no distractor < PRb < PRe < UR
Morsella & Miozzo, 2002	x			x			x				x		RT: PRb < UR
Navarette & Costa, 2005	x			x									RT: PRb < UR (language = Spanish)
Roon & Gafos, 2014	Syllable symbol name		x		x		x				x	x	RT: Congruent = distractor and target share place of articulation (e.g., ba-pa). Incongruent = e.g., ba-ta. RTs shortest in tone only condition, shorter in congruent than incongruent condition. Longer SOA leads to longer RT.
Schriefers, Meyer, & Levelt, 1990 (Exp 2)	x		x		x			x	x		x	-150, +150	RT: At SOA 0ms, no distractor < PRb < UR = SR At SOA 150ms, PRb was numerically faster than silence. Semantic effects at negative SOAs, Phonological at positive SOAs (language = Dutch)
Underwood & Briggs, 1984	x	x			x			x	x		x		RT: OR < nonword = PRe = UR < SR No by-condition differences in experiment 2

Study	Target	Distractor			Outcome measure		Variables						Finding
							Phonological relatedness		Semantic relatedness		SOA (ms)		
	Picture name	written	acoustic	picture	Voice response latency	other	onset	end	Co-ordinate	associate	0	Other (ms)	
													(when a word was named instead of a picture)
Yuen, Brysbaert, Davis, & Rastle, 2010	Written nonword syllable		Non word syllable			EPG	x	x			x		Auditory presentation of a rime-overlap syllable produced articulatory location interference when the target and distractor onset differed in place of articulation, but not when they differed in voicing.
Zhao, La Heij, Schiller, 2012	x	x			x		x					x	Language of study = Mandarin PR (words differed only by tone) < UR

Key:

BOLD = blood oxygenation level outcome measure

fMRI = functional magnetic resonance imaging

PRb = Phonologically related, onset overlap

PRe = Phonologically related, end overlap

RT = response time

SR = Semantically related

SOA = stimulus onset asynchrony

UR = unrelated

Appendix B: Context studies

Key:
ERP = event related potential;
G+ = syntactically congruent context;
G- = syntactically incongruent context;
Lex freq = lexical frequency;
Ppt = participant;
P+ = phonologically congruent context;
P- = phonologically incongruent context;

	task				constraint			Sentence stimuli presentation		Response stimuli presentation				manipulation congruency			other	Control / baseline condition		
Study	Picture naming	Word naming	LDT	other	high	med	low	acoustic	orthographic	image	acoustic	orthographic	other	semantic	Syntactic (gender)	phonological		Neutral sentence	acontextual	Findings
Aydelott & Bates, 2004			x		x		x	x			x			x			Signal clarity	x		Low pass filter: S+ < S- but S+ = N = S- Temporal compression: S+ < N = S-
Bentrovato, Devescovi, D'Amico, & Bates, 1999	x				x	x		x		x				x	x			x		S+G+ < S-G+ = S+/G- = N < S-G-
Bentrovato, Devescovi, D'Amico, Wicha, & Bates, 2003		x			x	x		x				x		x	x			x		S+G+ < S-G+ = S+/G- < N = S-G-
Bock & Griffin, 1998	x				x	x	x		RSVP	x				x			Lex freq	x		Frequency effect diminished in high-cloze context
DeLong, Kutas, & Urbach, 2005				ERP	x				RSVP			x		x		x				ERP negativity: P- > P+
Fischler & Bloom, 1979			x		x	x	x		x			x		x					x	For high-cloze: S+ < control < S-. "Cost-free" facilitation (p.17)
Fischler & Bloom, 1985			x		x		x					x		x					x	As above
Foucart, Ruiz-Tada, & Costa, 2015				ERP	x			x			x				x					ERP negativity on article (la/el): G->G+
Forster, 1981		x	x		x				RSVP			x		x					x	High-cloze = S+ = C < S-
Gollan, Slattery, Goldenberg, van Assche, Duyck, & Rayner, 2011	x	x			x		x		SPVP								Lex freq			Frequency effect diminished in high-cloze context
Jacobsen, 1999	x	x						x		x		x			x			x		G+<N<G-
Jordan & Thomas, 2002				x	x	x			x			x							x	
Manenti, Repetto, Bentrovato, Marcone, Bates, & Cappa, 2004		x			x		x							x	x		Ppt age	x		S+G+ < S+G- < N
Piai, Roelofs, & Maris, 2014	x			ERP	x		x		RSVP					x			Lex freq			Frequency effect diminished in high-cloze context. Frequency effect before picture presented (prediction).
Roe, Jahn-Samilo, Juarez, Mickel, Royer, & Bates, 2000	x				x		x	x		x				**x**			Ppt age	x		S+ < N < S- Interference decreases across blocks, facilitation does not
Stanovich & West, 1979		x							x			x		**x**			Target clarity		x	S+ < C in all clarity conditions C < S- in target degraded condition
Tulving & Gold, 1963		x			x									**x**			N words in context			S+ : positive correlation between RT and N words in sentential context S-: negative correlation between RT and N words in sentential context
Wicha, Moreno, & Kutas, 2003b (Spanish)				ERP	x				RSVP		x			**x**	x					ERP N400: S- > S+ ; G- > G+ "readers generate expectations for specific nouns and their articles"
Wicha , Orozco-Figueroa, Reyes, Hernandez, Galvadón de Barreto, & Bates, 2005	x				x			x						**x**	x				x	S+G+ < N < S-G-

Appendix C: Experimental Item Details

sentence-stem (sentence-stems detailed below appeared in Experiment 2 only; see following page for sentence-stems that were also included in further experiments)		target (picture naming agreement)	alternative picture name responses	alternative cloze responses	cloze probability	sentence-stem duration (secs)	sentence-stem no. of syllables	sentence-stem syllables/sec	target word frequency
Homophobia is a fear or dislike of people who are				homosexual	0.7	4.035	16	3.965	
My youngest son's not in to girls. He says he always knew that he's		gay (.5)	rainbow	NA	1	5.181	16	4.178	2.9903
Civil partnerships are like marriage for people who are				homosexual	0.8	4.092	14	3.032	
The green traffic light signals that it's your turn to			traffic light, green	NA	1	3.678	12	3.263	
Should I stay or should I		go (.5)		NA	1	1.436	6	4.178	3.9214
The starter shouted, "Ready, steady,"				NA	1	2.968	9	3.032	
When holding a tennis racquet you should you use a firm				hold	0.8	3.797	13	3.423	
He was hanging on by his fingers and was beginning to lose his		grip (.3)	hand, fist	strength	0.8	3.749	17	4.535	2.6021
When he shook my hand he held on with such a firm				hold	0.7	3.658	12	3.280	
The dentist recommends that you brush your teeth twice a				NA	1	3.391	13	3.834	
It's your lucky		day (.5)	sun, park	chance	0.9	0.972	4	4.115	3.8881
The sun was shining and it was such a lovely warm				morning, feeling	0.8	3.874	13	3.356	
To make bread you have to knead the				mixture	0.9	2.375	10	4.211	
The children made a model dinosaur out of play		dough (.7)	pastry	NA	1	3.255	14	4.301	2.6107
Some pizza places serve those little balls of warm				NA	1	3.859	12	3.110	
He was so dehydrated that they had to hook him up to a				NA	1	3.928	16	4.073	
When the tap was leaking he couldn't bear the incessant		drip (.6)	milk, drop	dribble	0.8	3.506	14	3.993	2.2279
The infection was so bad that she had to be given antibiotics via an intravenous				method	0.9	6.065	15	2.473	
all /g/ or /k/ onset (i.e. velar onset)	**mean**	0.64			0.9	3.215	12	3.8118	2.9580
	SD	0.28	-	-	0.110	0.924	3	0.4757	0.6076
all /d/ or /t/ onset (i.e. alveolar onset)	**mean**	0.60			0.930	3.279	12	3.852	2.978
	SD	0.31	-	-	0.082	1.042	3	0.614	0.573
Experiment 1 overall	**mean**	0.62			0.907	3.247	12	3.8319	2.968
	SD	0.29	-	-	0.099	0.976	3	0.545	0.585
Subsequent experiments overall	**mean**	0.68			0.914	3.098	12	3.924	2.932
(i.e. /k/ and /t/ onset items only)	**SD**	0.34	-	-	0.093	0.847	3	0.521	0.545

Explanatory note: The sentence-stems appearing on the first page of this appendix appeared only in Experiment 2. All other sentence-stems appeared in all the experiments that involved a sentence-stem manipulation (i.e., the experiments reported in Chapters 2, 3, 6, and 7).

sentence-stem		target (naming agreement)	alternative picture name responses	alternative cloze responses	cloze probability	sentence-stem duration (secs)	sentence-stem no. of syllables	sentence-stem syllables/sec	target word frequency
There were five tiers to the wedding				ring	0.9	1.901	8	4.208	
Jenny lit the candles on the birthday		cake (1)	NA	NA	1	2.548	9	3.532	3.0785
Would you like a muffin or would you prefer some lemon				flan, curd	0.8	3.128	14	4.476	
The gardener picked up the watering				NA	1	2.215	10	4.515	
There's no such word as can't. You have to believe that you		can (1)	NA	NA	1	3.94	14	3.553	3.9219
You can drink beer from a glass or straight from the				tap	0.8	3.19	11	3.448	
On his head he wore the school				tie	0.8	2.016	7	3.472	
A soft flat hat is sometimes known as a		cap (.7)	hat	beret	0.9	3.067	11	3.587	2.7924
Your car wheel has lost its hub				NA	1	1.917	7	3.652	
We made a Superman outfit using blue tights and a red sheet to be the				clothes, outfit	0.8	5.029	18	3.579	
You'll know it's Dracula if he's got fangs and is wearing a		cape (.5)	cloak	cloak	0.8	3.55	15	4.225	2.3541
He thinks he can fly when he's wearing his Superhero				mask, outfit	0.8	3.0825	14	4.542	
He loves sailing so they moved to the south			city	port	0.9	2.734	10	3.658	
Because Britain is an island it has a very long		coast (.3)	river	history	0.9	3.324	14	4.212	2.9009
Plymouth is a lovely city on the south			skyscraper	NA	1	2.555	11	4.305	
During the roadworks the central reservation was marked out by				lines	0.9	4.373	16	3.659	
She went to the van and bought an ice-cream		cone (1)	NA	NA	1	2.725	10	3.670	2.0607
Would you like a lolly or would you prefer an ice-cream				NA	1	2.928	13	4.440	
all /k/ onset	**mean**	0.75	-	-	0.91	3.012	12	3.930	2.8514
	SD	0.30	-	-	0.09	0.836	3	0.415	0.6053

sentence-stem		target	alternative picture name responses	alternative responses	cloze probability	sentence-stem duration (secs)	sentence-stem no. of syllables	sentence-stem syllables/sec	target word frequency
The secret to a happy marriage is a bit of give and			video	NA	1	2.839	15	5.284	
We're running out of film. We'll try to film the whole scene in a single		take (.3)	clapperboard	day, frame	0.8	4.701	17	3.616	3.9191
Some people like to give but others always			film	NA	1	3.784	11	2.907	
She thinks that if she doesn't use sunscreen she'll get a better			sunbed	NA	1	3.525	15	4.255	
To me she looked orange but she thought she had a nice		tan (.5)	sunbather	NA	1	3.46	13	3.757	2.4487
Before she goes on holiday she goes for one of those spray			fake tan	NA	1	3.566	15	4.206	
Jimmy managed to fix the drip from the old leaky				boiler	0.8	2.724	12	4.405	
I'd love to have a constant source of beer on		tap (1)		call	0.9	3.294	11	3.339	2.6821
When we want water we just turn on the				hose	0.9	3.56	10	2.809	
The only thing holding it all together was gaffer				NA	1	3.464	13	3.753	
I'm sure you can fix it with a bit of sticky		tape (1)		stuff	0.9	2.565	12	4.678	3.1717
Before discs came in you used to have record TV programs on to video				NA	1	5.294	20	3.778	
The fire alarm's gone off again someone must have burnt the				paper, curtains	0.8	3.155	14	4.437	
She likes butter and jam on her		toast (.9)	plate	scone	0.9	1.977	8	4.047	3.0434
He asked them to raise their glasses in a				NA	1	2.77	10	3.610	
His crass jokes really lower the			note	NA	1	2.065	8	3.874	
Type in the numbers when you hear the dial		tone (0)	tune	NA	1	2.49	11	4.418	2.8069
Her voice has such a lovely			music	timbre, quiver	0.8	2.081	7	3.364	
all /t/ onset	**mean**	0.6	-	-	0.93	3.184	12	3.919	3.0198
	SD	0.4	-	-	0.08	0.874	3	0.622	0.4822

sentence-stem	target	alternative responses	cloze probability	sentence-stem duration (secs)	sentence-stem no. of syllables	sentence-stem syllables/sec	target word frequency
There were five tiers to the wedding		ring	0.9	1.901	8	4.208	
Jenny lit the candles on the birthday	cake	NA	1	2.548	9	3.532	3.0785
Would you like a muffin or would you prefer some lemon		flan, curd	0.8	3.128	14	4.476	
The gardener picked up the watering		NA	1	2.215	10	4.515	
There's no such word as can't. You have to believe that you	can	NA	1	3.94	14	3.553	3.9219
You can drink beer from a glass or straight from the		tap	0.8	3.19	11	3.448	
On his head he wore the school		tie	0.8	2.016	7	3.472	
A soft flat hat is sometimes known as a	cap	beret	0.9	3.067	11	3.587	2.7924
Your car wheel has lost its hub		NA	1	1.917	7	3.652	
We made a Superman outfit using blue tights and a red sheet to be the		clothes, outfit	0.8	5.029	18	3.579	
You'll know it's Dracula if he's got fangs and is wearing a	cape	cloak	0.8	3.55	15	4.225	2.3541
He thinks he can fly when he's wearing his Superhero		mask, outfit	0.8	3.0825	14	4.542	
He loves sailing so they moved to the south		port	0.9	2.734	10	3.658	
Because Britain is an island it has a very long	coast	history	0.9	3.324	14	4.212	2.9009
Plymouth is a lovely city on the south		NA	1	2.555	11	4.305	
During the roadworks the central reservation was marked out by		lines	0.9	4.373	16	3.659	
She went to the van and bought an ice-cream	cone	NA	1	2.725	10	3.670	2.0607
Would you like a lolly or would you prefer an ice-cream		NA	1	2.928	13	4.440	
The secret to a happy marriage is a bit of give and		NA	1	2.839	15	5.284	
We're running out of film. We'll try to film the whole scene in a single	take	day, frame	0.8	4.701	17	3.616	3.9191
Some people like to give but others always		NA	1	3.784	11	2.907	
She thinks that if she doesn't use sunscreen she'll get a better		NA	1	3.525	15	4.255	
To me she looked orange but she thought she had a nice	tan	NA	1	3.46	13	3.757	2.4487
Before she goes on holiday she goes for one of those spray		NA	1	3.566	15	4.206	
Jimmy managed to fix the drip from the old leaky		boiler	0.8	2.724	12	4.405	
I'd love to have a constant source of beer on	tap	call	0.9	3.294	11	3.339	2.6821
When we want water we just turn on the		hose	0.9	3.56	10	2.809	

The only thing holding it all together was gaffer		NA	1	3.464	13	3.753	
I'm sure you can fix it with a bit of sticky	tape	stuff	0.9	2.565	12	4.678	3.1717
Before discs came in you used to have record TV programs on to video		NA	1	5.294	20	3.778	
The fire alarm's gone off again someone must have burnt the		paper, curtains	0.8	3.155	14	4.437	
She likes butter and jam on her	toast	scone	0.9	1.977	8	4.047	3.0434
He asked them to raise their glasses in a		NA	1	2.77	10	3.610	
His crass jokes really lower the		NA	1	2.065	8	3.874	
Type in the numbers when you hear the dial	tone	NA	1	2.49	11	4.418	2.8069
Her voice has such a lovely		timbre, quiver	0.8	2.081	7	3.364	
			0.9194	3.098	12	3.924	2.9317
			0.8	1.901	7	2.809	2.0607
			1	5.294	20	5.284	3.9219

References

Abdel Rahman, R., & Melinger, A. (2008). Enhanced phonological facilitation and traces of concurrent word form activation in speech production: An object-naming study with multiple distractors. *The Quarterly Journal of Experimental Psychology*, *61*(9), 1410-1440

Abel, S., Dressel, K., Bitzer, R., Kümmerer, D., Mader, I., Weiller, C., & Huber, W. (2009). The separation of processing stages in a lexical interference fMRI-paradigm. *Neuroimage*, *44*(3), 1113-1124.

Aglioti, S. M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature neuroscience*, *11*(9), 1109-1116.

Alario, F. X., & Hamamé, C. M. (2013). Evidence for, and predictions from, forward modeling in language production. *Behavioral and Brain Sciences*, *36*(04), 348-349.

Alario, X.-F., Segui, J., & Ferrand, L. (2000). Semantic and associative priming in picture naming. *The Quarterly Journal of Experimental Psychology: Section A*, *53*(3), 741-764.

Alegre, M., de Gurtubay, I. G., Labarga, A., Iriarte, J., Malanda, A., & Artieda, J. (2004). Alpha and beta oscillatory activity during a sequence of two movements. *Clinical neurophysiology*, *115*(1), 124-130.

Allport, F.H. (1924) *Social Psychology*, Houghton Mifflin

Altmann, G., & Kamide, Y. (1999) Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247-264.

Arantes, P., & Barbosa, P. A. (2010). Production–perception entrainment in speech rhythm. In *Proceedings of the 5th International Conference of Speech Prosody, Chicago, USA* (pp. 1-4).

Arnal, L. H., & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends in cognitive sciences*, *16*(7), 390-398.

Articulate assistant advanced user guide: Version 2.14 [Computer software manual]. Edinburgh, UK.

Articulate Instruments Ltd (2012) *Articulate Assistant Advanced User Guide: Version 2.14*. Edinburgh, UK: Articulate Instruments Ltd.

Articulate Instruments Ltd (2008) *Ultrasound Stabilisation Headset Users Manual: Revision 1.4*. Edinburgh, UK: Articulate Instruments Ltd

Atmaca, S., Sebanz, N., & Knoblich, G. (2011). The joint flanker effect: sharing tasks with real and imagined co-actors. *Experimental Brain Research*, *211*(3-4), 371-385.

Audacity (2014). Audacity(R): Free Audio Editor and Recorder [Computer program]. Version 2.0.0 retrieved April 20th 2014 from http://audacity.sourceforge.net/

Avenanti, A., Candidi, M., & Urgesi, C. (2013). Vicarious motor activation during action perception: beyond correlational evidence. *Frontiers in Human Neuroscience7*, *185*(2,906).

Aydelott, J., & Bates, E. (2004). Effects of acoustic distortion and semantic context on lexical access. *Language and Cognitive Processes*, *19*(1), 29-56.

Ayora, P., Peressotti, F., Alario, F. X., Mulatti, C., Pluchino, P., Job, R., & Dell'Acqua, R. (2011). What phonological facilitation tells about semantic interference: a dual-task study. *Frontiers in psychology*, *2*. doi: 10.3389/fpsyg.2011.00057

Baayen, R. H. (2008). *Analyzing linguistic data* (Vol. 505). Cambridge, UK: Cambridge University Press.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, *40*(1), 177-189.

Ball, M. (2016). Speech Imaging. *Equinox Publishing*.

Barbier, G., Perrier, P., Ménard, L., Payan, Y., Tiede, M., & Perkell, J. (2015). Speech planning in 4-year-old children versus adults: Acoustic and articulatory analyses. In *Interspeech 2015*.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278.

Bastiaansen, M. C., Oostenveld, R., Jensen, O., & Hagoort, P. (2008). I see what you mean: theta power increases are involved in the retrieval of lexical semantic information. *Brain and language*, *106*(1), 15-28.

Bates D, Maechler M, Bolker B and Walker S (2014). _lme4: Linear mixed-effects models using Eigen and S4_. R package version 1.1-7, <URL: http://CRAN.R-project.org/package=l

Bates, D., Maechler, M., Dai, B. (2008). Ime4: Linear mixed-effects models using S4 classes (R package version 0.999375-27). Retrieved from: *http://www. r-proiect. org*.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. *The crosslinguistic study of sentence processing*, *3*, 73.

Baus, C., Sebanz, N., de la Fuente, V., Branzi, F. M., Martin, C., & Costa, A. (2014). On predicting others' words: Electrophysiological evidence of prediction in speech production. *Cognition*, *133*(2), 395-407.

Ben-David, B.M., Van Lieshout, P. H., & Nishta, A., 2009, The sound of Stroop: acoustic effects in Stroop interference, *Canadian Acoustics*, *37*(3), 188-189

Bentrovato, S., Devescovi, A., D'Amico, S., & Bates, E. (1999). Effect of grammatical gender and semantic context on lexical access in Italian. *Journal of Psycholinguistic Research*, *28*(6), 677-693.

Bentrovato, S., Devescovi, A., D'Amico, S., Wicha, N., & Bates, E. (2003). The effect of grammatical gender and semantic context on lexical access in Italian using a timed word-naming paradigm. *Journal of psycholinguistic research*, *32*(4), 417-430.

Blakemore, S. J., Frith, C. D., & Wolpert, D. M. (1999). Spatio-temporal prediction modulates the perception of self-produced stimuli. *Journal of cognitive neuroscience*, *11*(5), 551-559.

Block, C. K., & Baldwin, C. L. (2010). Cloze probability and completion norms for 498 sentences: Behavioral and neural validation using event-related potentials. *Behavior research methods*, *42*(3), 665-670.

Bloom, P. A., & Fischler, I. (1980). Completion norms for 329 sentence contexts. *Memory & Cognition*, *8*(6), 631-642.

Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning?. *Journal of Experimental Psychology: General*, *129*(2), 177.

Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive psychology*, *23*(1), 45-93.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International* 5, 341-345.

Boland, J. E. (2005). Visual arguments. *Cognition*, *95*(3), 237-274

Boucher, V. J. (1994). Alphabet-related biases in psycholinguistic enquiries: Considerations for direct theories of speech production and perception. *Journal of Phonetics*.

Boulinguez, P., Jaffard, M., Granjon, L., & Benraiss, A. (2008). Warning signals induce automatic EMG activations and proactive volitional inhibition: evidence from analysis of error distribution in simple RT. *Journal of Neurophysiology*, *99*(3), 1572-1578.

Branigan, H., Pickering, M., McLean, J. & Cleland, A.A. (2007). Syntactic alignment and participant role in dialogue. *Cognition, 104*, 163–197

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482.

Briggs, P., & Underwood, G. (1982). Phonological coding in good and poor readers. *Journal of Experimental Child Psychology*, *34*(1), 93-112.

Brooks, P. J., & MacWhinney, B. (2000). Phonological priming in children's picture naming. *Journal of Child Language*, *27*(02), 335-366.

Brown, E. C., & Brüne, M. (2012). The role of prediction in social neuroscience. *Frontiers in human neuroscience*, *6*, 147.

Brown, E. C., Muzik, O., Rothermel, R., Matsuzaki, N., Juhász, C., Shah, A. K., & Asano, E. (2012). Evaluating reverse speech as a control task with language-related gamma activity on electrocorticography. *NeuroImage*, *60*(4), 2335-2345.

Brysbaert, M., Fias, W., & Reynvoet, B. (2006). The issue of semantic mediation in word and number naming. *Advances in psychology research*, *48*, 181-200.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977-990.

Bürki, A., Cheneval, P. P., & Laganaro, M. (2015). Do speakers have access to a mental syllabary? ERP comparison of high frequency and novel syllable production. *Brain and Language, 150*, 90-102.

Cai, Z. G., Pickering, M. J., & Branigan, H. P. (2012). Mapping concepts to syntax: Evidence from structural priming in Mandarin Chinese. Journal of Memory and Language, 66(4), 833-849.

Camargo, L., Coutinho, P., Madureira, S., & Rusilo, L.C. (2015) Voice quality description from a phonetic perspective: Supralaryngeal and muscular tension settings in ICPHS 2015

Carroll, R. J., & Ruppert, D. (1988). *Transformation and weighting in regression* (Vol. 30). CRC Press. New York, NY: Chapman & Hall.

Chabal, S., & Marian, V. (2015). Speakers of different languages process the visual world differently. *Journal of Experimental Psychology: General*, *144*(3), 539.

Chen, F., Ruiz, N., Choi, E., Epps, J., Khawaja, M. A., Taib, R. & Wang, Y. (2012). Multimodal behavior and interaction as indicators of cognitive load. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *2*(4), 22

Cheyne, D. O. (2013). MEG studies of sensorimotor rhythms: a review. *Experimental neurology*, *245*, 27-39.

Chomsky, N. (1965). *Aspects of the theory of syntax* Cambridge. Multilingual Matters: MIT Press.

Christoffels, I. K., Formisano, E., & Schiller, N. O. (2007). Neural correlates of verbal feedback processing: an fMRI study employing overt speech. *Human brain mapping*, *28*(9), 868-879.

Clark, H. H. (1996) *Using language*. Cambridge, England: Cambridge University Press

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1-39.

Colling, L. J., Thompson, W. F., & Sutton, J. (2014). The effect of movement kinematics on predicting the timing of observed actions. *Experimental Brain Research, 232,* 1193–1206.

Collins, A. F., & Ellis, A. W. (1992). Phonological priming of lexical retrieval in speech production. *British Journal of Psychology*, *83*(3), 375-388.

Coltheart, M., & Rastle, K. (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: human perception and performance*, *20*(6), 1197.

Content, A., Dumay, N., & Frauenfelder, U. (2000). The role of syllable structure in lexical segmentation: Helping listeners avoid mondegreens. In *ISCA Tutorial and Research Workshop (ITRW) on Spoken Word Access Processes*.

Costa, A., Alario, F. X., & Caramazza, A. (2005). On the categorical nature of the semantic interference effect in the picture-word interference paradigm. *Psychonomic Bulletin & Review*, *12*(1), 125-131.

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 498.

Damian, M. F. (2003). Articulatory duration in single-word speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(3), 416.

Damian, M. F., & Bowers, J. S. (2009). Assessing the role of orthography in speech perception and production: Evidence from picture–word interference tasks. *European Journal of Cognitive Psychology*, *21*(4), 581-598.

Damian, M. F., & Dumay, N. (2007). Time pressure and phonological advance planning in spoken production. *Journal of Memory and Language*, *57*(2), 195-209.

Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 345

Dang, J., Wei, J., Suzuki, T., Honda, K., Perrier, P., & Honda, M. (2004). Investigation and modeling of coarticulation in speech production. In *Chinese Spoken Language Processing, 2004 International Symposium on* (pp. 25-28). IEEE.

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatopy of speech perception. *Current Biology, 19*, 381–385.

Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *The Journal of the Acoustical Society of America*, *120*(1), 407-415.

Davidson, L. (2005). Addressing phonological questions with ultrasound. *Clinical Linguistics & Phonetics*, *19*(6-7), 619-633.

Davis, C., Shaw, J., Proctor, M., Derrick, D., Sherwood, S., & Kim, J. (2015). Examining speech production using masked priming. *Proceedings of the 18th ICPhS, Glasgow* (0560).

Delcomyn, F. (1977). Corollary discharge to cockroach giant interneurons. *Nature* 269, 160 – 162.

Dell, G. S. (2013). Cascading and feedback in interactive models of production: A reflection of forward modeling?. *Behavioral and Brain Sciences*, *36*(04), 351-352.

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of memory and language*, *27*(2), 124-142.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review*, *93*(3), 283.

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological review*, *104*(1), 123.

Dell, G. S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science*, *17*(2), 149-195.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological review*, *104*(4), 801.

Dell'Acqua, R., Sessa, P., Peressotti, F., Mulatti, C., Navarrete, E., & Grainger, J. (2010). ERP evidence for ultra-fast semantic processing in the picture–word interference paradigm. *Frontiers in Psychology, 1*. doi: 10.3389/fpsyg.2010.00177

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117-1121.

de Ruiter, J. P., & Cummins, C. (2013). Forward modelling requires intention recognition and non-impoverished predictions. *Behavioral and Brain Sciences*, *36*(04), 351-351.

de Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, 515-535.

de Zubicaray, G. I., & McMahon, K. L. (2009). Auditory context effects in picture naming investigated with event-related fMRI. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(3), 260-269.

de Zubicaray, G. I., McMahon, K. L., Eastburn, M. M., & Wilson, S. J. (2002). Orthographic/phonological facilitation of naming responses in the picture–word task: An event-related fMRI study using overt vocal responding. *Neuroimage*, *16*(4), 1084-1093.

Dhooge, E., & Hartsuiker, R. J. (2010). The distractor frequency effect in picture-word interference: Evidence for response exclusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 878.

Dikker, S., & Pylkkanen, L. (2011). Before the N400: Effects of lexical–semantic violations in visual cortex. *Brain and Language*, *118*(1), 23-28.

Dikker, S., & Pylkkänen, L. (2013). Predicting language: MEG evidence for lexical preactivation. *Brain and Language*, *127*(1), 55-64.

Dolk, T., Hommel, B., Colzato, L. S., Schütz-Bosbach, S., Prinz, W., & Liepelt, R. (2014). The joint Simon effect: a review and theoretical integration. *Frontiers in Psychology*, *5*.

Drake, E., & Corley, M. (2015b). Articulatory imaging implicates prediction during spoken language comprehension. *Memory & Cognition*, *43*(8), 1136-1147.

Drake, E., & Corley, M. (2015a). Effects in production of word pre-activation during listening: Are listener-generated predictions specified at a speech-sound level? *Memory & Cognition*, *43*(1), 111-120.

Düzel, E., Penny, W. D., & Burgess, N. (2010). Brain oscillations and memory. *Current opinion in neurobiology*, *20*(2), 143-149.

Eden, G., & Inbar, G. F. (1978). Physiological model analysis of involuntary human-voice tremor. *Biological cybernetics*, *30*(3), 179-185.

Eiter, B. M., & Inhoff, A. W. (2008). Visual word recognition is accompanied by covert articulation: evidence for a speech-like phonological representation. *Psychological research*, *72*(6), 666-674.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, *15*(2), 399-402.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491-505.

Federmeier, K. D., & Kutas, M. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 202

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*(4), 469-495.

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and language*, *115*(3), 149-161.

Federmeier, K. D., McLennan, D. B., Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*(2), 133-146.

Ferreira, V. S., & Griffin, Z. M. (2003). Phonological influences on lexical (mis) selection. *Psychological Science, 14*(1), 86-90.

Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(6), 1187.

Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: a review of the role of the motor system in language comprehension. *The Quarterly Journal of Experimental Psychology*, *61*(6), 825-850.

Fischler, I., & Bloom, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning and Verbal Behavior*, *18*(1), 1-20.

Flinker, A., Chang, E. F., Kirsch, H. E., Barbaro, N. M., Crone, N. E., & Knight, R. T. (2010). Single-trial speech suppression of auditory cortex activity in humans. *The Journal of Neuroscience*, *30*(49), 16643-16650.

Ford, J. M., Mathalon, D. H., Heinks, T., Kalba, S., Faustman, W. O., & Roth, W. T. (2001). Neurophysiological evidence of corollary discharge dysfunction in schizophrenia. *American Journal of Psychiatry*. *158*, 2069–2071

Forster KI. (1981) Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *Quarterly Journal of Experimental Psychology. 33A*, 465–495

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of verbal learning and verbal behavior*, *12*(6), 627-635.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*(1), 116-124.

Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 1461–1469.

Foucart, A., Ruiz-Tada, E., & Costa, A. (2016). Anticipation processes in L2 speech comprehension: Evidence from ERPs and lexical recognition task. *Bilingualism: Language and Cognition*, *19*(01), 213-219.

Fournier, L., Scheffers, M. K., Coles, M. G., Adamson, A., & Abad, E. V. (1997). The dimensionality of the flanker compatibility effect: A psychophysiological analysis. *Psychological Research*, *60*(3), 144-155.

Fowler, C. A. (2014). Talking as doing: language forms and public language. *New Ideas in Psychology*, *32*, 174-182.

Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language*, *49*(3), 396-413.

Fowler, C. A., & Saltzman, E. (1993). Coordination and coarticulation in speech production. *Language and speech*, *36*(2-3), 171-195.

Frisch, S. A., & Wodzinski, S. M. (2016). Velar–vowel coarticulation in a virtual target model of stop production. *Journal of phonetics*, *56*, 52-65.

Frisch, S. A., & Wright, R. (2002). The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics*, *30*(2), 139-162.

Gafos, A., & Goldstein, L. (2012). Articulatory representation and organization. *Oxford handbook of laboratory phonology*, 220-231.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic bulletin & review*, *13*(3), 361-377.

Galbraith, G. C., Amaya, E. M., de Rivera, J. M. D., Donan, N. M., Duong, M. T., Hsu, J. N., ... & Tsang, L. P. (2004). Brain stem evoked response to forward and reversed speech in humans. *Neuroreport*, *15*(13), 2057-2060.

Gallese, V. (2008). Mirror neurons and the social nature of language: The neural exploitation hypothesis. *Social neuroscience*, *3*(3-4), 317-333.

Gambi, C., & Pickering, M. J. (2016). Predicting and imagining language. Language, Cognition and Neuroscience, 31(1), 60-72.

Gambi, C., & Pickering, M. J. (2013). Talking to each other and talking together: Joint language tasks and degrees of interactivity. *Behavioral and Brain Sciences*, *36*(04), 423-424.

Garnier, M., & Henrich, N. (2013). Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise? *Computer Speech & Language*.

Garnier, M., Lamalle, L., & Sato, M. (2013). Neural correlates of phonetic convergence and speech imitation. *Frontiers in psychology*, *4*.

Garrod, S., Gambi, C., & Pickering, M. J. (2014). Prediction at all levels: forward model predictions can enhance comprehension. *Language, Cognition and Neuroscience*, *29*(1), 46-48.

Gentsch, A., Weber, A., Synofzik, M., Vosgerau, G., & Schütz-Bosbach, S. (2016). Towards a common framework of grounded action cognition: Relating motor control, perception and cognition. *Cognition*, *146*, 81-89.

Giles, H., Coupland, N., & Coupland, J. (1991). 1. Accommodation theory: Communication, context, and. *Contexts of Accommodation: Developments in Applied Sociolinguistics*, *1*, 1-68

Giles, H., Taylor, D. M., & Bourhis, R. (1973). Towards a theory of interpersonal accommodation through language: Some Canadian data. *Language in society*, *2* (2), 177-192.

Glaser, W. R., & Düngelhoff, F. J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 640.

Goffman, L., Smith, A., Heisler, L., and Ho, M. (2008). Speech production units in children and adults: Evidence from coarticulation. *Journal of Speech, Language, and Hearing Research, 51,* 1423-1437

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological review*, *105*(2), 251.

Goldinger S. D. (1997). Words and voices: Perception and production in an episodic lexicon, in *Talker Variability in Speech Processing*, Eds. K. Johnson, J.W. Mullennix. San Diego, CA: Academic Press, pp. 33–66

Goldrick, M., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, *21*(6), 649-683.

Goldrick, M., & Chu, K. (2014). Gradient co-activation and speech error articulation: Comment on Pouplier and Goldstein (2010). *Language, Cognition and Neuroscience*, *29*(4), 452-458.

Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition*, *103*(3), 386-412.

Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: the frequency-lag hypothesis. *Journal of Experimental Psychology: General*, *140*(2), 186.

Goodman, K. S. (1967). Reading: A psycholinguistic guessing game. *Literacy Research and Instruction*, *6*(4), 126-135.

Gower, J. C. (1966). A Q-technique for the calculation of canonical variates. *Biometrika*, 588-590.

Grainger, J., & Holcomb, P. J. (2009). An ERP investigation of orthographic priming with relative-position and absolute-position primes. *Brain research*, *1270*, 45-53.

Greene, J. O. (1988). Cognitive processes: Methods for probing the black box. In C.H. Tardy (Ed.*) A handbook for the study of human communication: Methods and instruments for observing, measuring, and assessing communication processes* (pp. 37-66). Westport, CT.: Ablex Publishing .

Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language*, *38*(3), 313-338.

Grubic, M. (2008) from http://www.grubic.com/analysis.praat

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences, 27*, 377–396.

Grush, R. (1997). The architecture of representation. *Philosophical Psychology*, *10*(1), 5-23.

Guenther, F. H. (2003). Neural control of speech movements. *Phonetics and phonology in language comprehension and production: Differences and similarities*, 209-240.

Gunter, T. C., Friederici, A. D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience*, *12*(4), 556-568.

Hagoort, P., Wassenaar, M., Brown, C.M., (2003a). Syntax-related ERP-effects in Dutch. *Cognition and Brain Research*, *16*, 38-50

Hänzi, S., Banchi, R., Straka, H., & Chagnaud, B.P., (2015). Locomotor corollary activation of trigeminal motoneurons: coupling of discrete motor behaviors. *Journal of Experimental Biology*, *218*, 1748-1758;

Hardcastle, W. J., & Hewlett, N. (Eds.). (1999). *Coarticulation: Theory, data and techniques*. Cambridge University Press.

Hartley, T., & Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language*, *35*(1), 1-31.

Hartsuiker, R.J. (2013). Are forward models enough to explain self-monitoring? Insights from patients and eye movements. *Behavioral and Brain Sciences*, *36*(4),

Haruno, M., Wolpert, D. M., & Kawato, M. (2003, October). Hierarchical MOSAIC for movement generation. In *International congress series* (Vol. 1250, pp. 575-590). Elsevier.

Haueisen, J., & Knösche, T. R. (2001). Involuntary motor activity in pianists evoked by music perception. *Journal of cognitive neuroscience*, *13*(6), 786-792.

Haweo, C. S. (2009). Neural and Behavioral Responses to the Use of Auditory Feedback in Vocal Control. http://scholars.wlu.ca/cgi/viewcontent.cgi?article=1912&context=etd (last accessed 20160416)

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2), 301-307.

Heinks-Maldonado,T.H., Nagarajan, S. S., & Houde, J. F. (2006). Magnetoencephalographic evidence for a precise forward model in speech production. Neuroreport, 17(13), 1375-1379.

Heldner, M., & Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, *38*(4), 555-568

Hickok, G. (2012). Computational neuroanatomy of speech production. Nature Reviews Neuroscience, 13(2), 135-145.

Hickok, G. (2013). Predictive coding? Yes, but from what source?. *Behavioral and Brain Sciences*, *36*(04), 358-358.

Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Language, Cognition and Neuroscience*, *29(1)*, 2-20

Hirschfeld, G., Jansma, B., Bölte, J., & Zwitserlood, P. (2008). Interference and facilitation in overt speech production investigated with event-related potentials. *Neuroreport*, *19*(12), 1227-1230.

Holcomb, P. J. (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain and Language*, *35*(1), 66-85.

Holcomb, P. J., & Grainger, J. (2006). On the time course of visual word recognition: An event-related potential investigation using masked repetition priming. *Cognitive Neuroscience, Journal of*, *18*(10), 1631-1643.

Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in cognitive sciences*, *8*(11), 494-500.

Houghton, G. (1990). The problem of serial order: A neural network model of sequence learning and recall. In *Current research in natural language generation* (pp. 287-319). Academic Press Professional, Inc..

Howes, C., Healey, P. G., Eshghi, A., & Hough, J. (2013). "Well, that's one way": Interactivity in parsing and production. *Behavioral and Brain Sciences*, *36*(04), 359-359.

Huber, S., & Krist, H. (2004). When is the ball going to hit the ground? Duration estimates, eye movements, and mental imagery of object motion. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(3), 431.

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain research*, *1626*, 118-135.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta psychologica, 137*(2), 151-171

Humphreys, K. R., Boyd, C. H., & Watter, S. (2010). Phonological facilitation from pictures in a word association task: Evidence for routine cascaded processing in spoken word production. *The Quarterly Journal of Experimental Psychology, 63*(12), 2289-2296.

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1), 101-144.

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Frontiers in psychology, 2*.

Itier, R. J., & Taylor, M. J. (2004). N170 or N1? Spatiotemporal differences between object and face processing using ERPs. *Cerebral Cortex, 14*(2), 132-142.

Ito, Corley, & Pickering (submitted 2016). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study.

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language, 86*, 157-171.

Jackson, J.H., (1958) Selected Writings of John Hughlings Jackson. J. Tylor (Ed). New York, Basic Books

Jacobsen, T. (1999). Effects of grammatical gender on picture and word naming: Evidence from German. *Journal of Psycholinguistic Research*, *28*(5), 499-514.

Jaeger, T. F., & Ferreira, V. (2013). Seeking predictions from a predictive framework. *Behavioral and Brain Sciences*, *36*(04), 359-360.

Jeannerod, M., Kennedy, H., & Magnin, M. (1979). Corollary discharge: its possible implications in visual and oculomotor interactions. *Neuropsychologia*, *17*(2), 241-258.

Jescheniak, J. D., & Levelt, W. J. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 824.

Mädebach, A., Jescheniak, J. D., Oppermann, F., & Schriefers, H. (2011). Ease of processing constrains the activation flow in the conceptual-lexical system during speech planning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(3), 649.

Jescheniak, J. D., Schriefers, H., Garrett, M. F., & Friederici, A. D. (2002). Exploring the activation of semantic and phonological codes during speech planning with event-related brain potentials. *Journal of Cognitive Neuroscience*, *14*(6), 951-964.

Jonides, J., & Mack, R. (1984). On the cost and benefit of cost and benefit. *Psychological Bulletin*, *96*(1), 29.

Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition, 94*(2), 113-147

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass*, *2*(4), 647-670

Kamide Y., Altmann G.T.M. & Haywood S.L. (2003) The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*, 133–156

Kamide, Y., Scheepers, C., & Altmann, G. T. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research, 32*(1), 37-55.

Kan, I. P., & Thompson-Schill, S. L. (2004). Effect of name agreement on prefrontal activity during overt and covert picture naming. *Cognitive, Affective, & Behavioral Neuroscience, 4*(1), 43-57.

Katz, W. F. (2000). Anticipatory coarticulation and aphasia: implications for phonetic theories. *Journal of Phonetics*, *28*(3), 313-334.

Keating, P. A. (1990). The window model of coarticulation: articulatory evidence. *Papers in laboratory phonology I, 26*, 451-470.

Kello, C. T., Plaut, D. C., & MacWhinney, B. (2000). The task dependence of staged versus cascaded processing: An empirical and computational study of Stroop interference in speech perception. *Journal of Experimental Psychology: General, 129*(3), 340.

Kessler, B., Treiman, R., & Mullennix, J. (2002). Phonetic biases in voice key response time measurements. *Journal of Memory and Language, 47*(1), 145-171.

Kilner, J., Hamilton, A. F. D. C., & Blakemore, S. J. (2007). Interference effect of observed human movement on action is due to velocity profile of biological motion. *Social Neuroscience*, *2*(3-4), 158-166.

Kilner, J. M., Paulignan, Y., & Blakemore, S.-J. (2003). An interference effect of observed biological movement on action. *Current Biology, 13*(6), 522-525

Kilner, J. M., Vargas, C., Duval, S., Blakemore, S. J., & Sirigu, A. (2004). Motor activation prior to observation of a predicted movement. *Nature neuroscience, 7*(12), 1299-1301.

Kim, A., & Lai, V. (2012). Rapid interactions between lexical semantic and word form analysis during word recognition in context: Evidence from ERPs. *Journal of Cognitive Neuroscience*, *24*(5), 1104-1112

Kim, J., Lammert, A. C., Ghosh, P. K., & Narayanan, S. S. (2014). Co-registration of speech production datasets from electromagnetic articulography and real-time magnetic resonance imaging. *The Journal of the Acoustical Society of America*, *135*(2), EL115-EL121.

Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive neuropsychology*, *25*(4), 463-492.

Kiyonaga, K., Grainger, J., Midgley, K., & Holcomb, P. J. (2007). Masked cross-modal repetition priming: An event-related potential investigation. *Language and Cognitive Processes*, *22*(3), 337-376.

Klapp, S. T. (1996). Reaction time analysis of central motor control. In H.N. Zelaznik (Ed.) *Advances in motor learning and control*, pp. 13-35, Champaign, IL: Human Kinetics

Klein, M., Grainger, J., Wheat, K. L., Millman, R. E., Simpson, M. I., Hansen, P. C., & Cornelissen, P. L. (2014). Early Activity in Broca's Area During Reading Reflects Fast Access to Articulatory Codes From Print. *Cerebral Cortex, bht350*.

Knobel, M., Finkbeiner, M., & Caramazza, A. (2008). The many places of frequency: Evidence for a novel locus of the lexical frequency effect in word production. *Cognitive Neuropsychology*, *25*(2), 256-286.

Knoblich, G., Butterfill, S., & Sebanz, N. (2011). 3 Psychological research on joint action: theory and data. *Psychology of Learning and Motivation-Advances in Research and Theory*, *54*, 59.

Knoblich, G., & Flach, R. (2001). Predicting the effects of actions: Interactions of perception and action. *Psychological Science*, *12*(6), 467-472.

Kochetov, A., & Neufeld, C. (2013). Examining the extent of anticipatory coronal coarticulation: A long-term average spectrum analysis. In *Proceedings of Meetings on Acoustics* (Vol. 19, No. 1, p. 060300). Acoustical Society of America.

Koenig, T., Studer, D., Hubl, D., Melie, L., & Strik, W. K. (2005). Brain connectivity at different time-scales measured with EEG. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *360*(1457), 1015-1024.

Kourtis, D., Sebanz, N., & Knoblich, G. (2013). Predictive representation of other people's actions in joint action planning: An EEG study. *Social neuroscience*, *8*(1), 31-42.

Krauss, R. M. & Weinheimer, S. (1964) Changes in reference phrases as a function of frequency of usage in social interaction. *Psychonomic Science,1*, 118-114

Krieger-Redwood, K., Gaskell, M. G., Lindsay, S., & Jefferies, E. (2013). The selective role of premotor cortex in speech perception: a contribution to phoneme judgements but not speech comprehension. *Journal of cognitive neuroscience*, *25*(12), 2179-2188.

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190–207). Oxford University Press.

Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & Cognition*, *11*(5), 539-550.

Lage-Castellanos, A., Martínez-Montes, E., Hernández-Cabrera, J. A., Galán, L. (2010). False discovery rate and permutation test: an evaluation in ERP data analysis. *Statistics in medicine*, *29*(1), 63-74.

Laganaro, M., Python, G., & Toepel, U. (2013). Dynamics of phonological–phonetic encoding in word production: Evidence from diverging ERPs between stroke patients and controls. *Brain and Language*, *126*(2), 123-132.

Lane, H., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech, Language, and Hearing Research*, *14*(4), 677-709.

Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language*, *61*(3), 326-338.

Laver, J. & Mackenzie-Beck J. (2007). Vocal Profile Analysis Scheme-VPAS. Queen Margaret University College-QMUC, Speech Science Research Centre, Edinburgh

Le, P. N. (2012). *The use of spectral information in the development of novel techniques for speech-based cognitive load classification* (Doctoral dissertation, The University of New South Wales).

Levitan, R., & Hirschberg, J. B. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions.

Lelong, A., & Bailly, G. (2011). Study of the phenomenon of phonetic convergence thanks to speech dominoes. In *Analysis of Verbal and Nonverbal Communication and Enactment. The Processing Issues* (pp. 273-286). Springer Berlin Heidelberg.

Levelt, W. J. M. (1989). Speaking: From intention to articulation. 1989. *Bradford, Cambridge, MA.*

Levelt, W. J. M. (1999). Producing spoken language: A blueprint of the speaker. In C. M. Brown, & P. Hagoort (Eds.), *The neurocognition of language* (pp. 83-122). Oxford University Press.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences*, *22*(01), 1-38.

Li, P., & Macwhinney, B. (2013). Competition model. *The Encyclopedia of Applied Linguistics*.

Lidji, P., Palmer, C., Peretz, I., & Morningstar, M. (2011). Listeners feel the beat: Entrainment to English and French speech rhythms. *Psychonomic Bulletin & Review, 18*, 1035–1041.

LimeSurvey Project Team, Carsten Schmitz (2011) LimeSurvey: An Open Source survey tool, LimeSurvey Project Hamburg, Germany. URL http://www.limesurvey.org

Loerts, H., Stowe, L. A., & Schmidt, M. S. (2013). Predictability speeds up the re-analysis process: An ERP investigation of gender agreement and cloze probability. *Journal of Neurolinguistics, 26*(5), 561-580.

Londei, A., D'Ausilio, A., Basso, D., Sestieri, C., Gratta, C. D., Romani, G. L., & Belardinelli, M. O. (2010). Sensory-motor brain network connectivity for speech comprehension. *Human brain mapping, 31*(4), 567-580.

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

Lukatela, G., Eaton, T., Sabadini, L., & Turvey, M. T. (2004). Vowel duration affects visual word identification: evidence that the mediating phonology is phonetically informed. *Journal of Experimental Psychology: Human Perception and Performance, 30(*1), 151.

Lukatela, G., Frost, S. J., & Turvey, M. T. (1998). Phonological priming by masked nonword primes in the lexical decision task. *Journal of Memory and Language*, *39*(4), 666-683.

Lupker, S. J. (1982). The role of phonetic and orthographic similarity in picture–word interference. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *36*(3), 349.

McGarva, A. R., & Warner, R. M. (2003). Attraction and social coordination: Mutual entrainment of vocal activity rhythms. *Journal of Psycholinguistic Research*, *32*(3), 335-354.

MacKay, D. G. (2012). *The organization of perception and action: A theory for language and other cognitive skills*. Springer Science & Business Media.

MacKay, D. G. (1987). The organization of perception and action: a theory for language and other cognitive skills. New York: Springer

McMillan, C. T., & Corley, M. (2010). Cascading influences on the production of speech: Evidence from articulation. *Cognition*, *117*(3), 243-260.

McMillan, C. T., Corley, M., & Lickley, R. J. (2009). Articulatory evidence for feedback and competition in speech production. *Language and Cognitive Processes*, *24*(1), 44-66.

McMillan, C.T. (2008) *Articulatory Evidence for Interactivity in Speech Production*. PhD thesis. University of Edinburgh

McQueen, J. M., & Huettig, F. (2014). Interference of spoken word recognition through phonological priming from visual objects and printed words. *Attention, Perception, & Psychophysics*, *76*(1), 190-200.

MacWhinney, B. (2008). A unified model of language acquisition. *Handbook of Bilingualism: Psycholinguistic Approaches, eds. JF Kroll and AMB De Groot (Oxford, 2005)*, 49-67.

MacWhinney, B. (2001). The competition model: The input, the context, and the brain. *Cognition and second language instruction*, 69-90

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: a reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 503.

Manenti, R., Repetto, C., Bentrovato, S., Marcone, A., Bates, E., & Cappa, S. F. (2004). The effects of ageing and Alzheimer's disease on semantic and gender priming. *Brain*, *127*(10), 2299-2306.

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 843.

Mardia, K. V. (1978). Some properties of clasical multi-dimesional scaling. *Communications in Statistics-Theory and Methods*, *7*(13), 1233-1241.

Marquardt, D. W., & Snee, R. D. (1975). Ridge regression in practice. *The American Statistician, 29*(1), 3-20.

Martin, C. D., Thierry, G., Kuipers, J. R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language, 69(*4), 574-588.

Mathalon, D. H., & Ford, J. M. (2008). Corollary discharge dysfunction in schizophrenia: evidence for an elemental deficit. *Clinical EEG and Neuroscience*, *39*(2), 82-86.

Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language*, *29*(5), 524-545

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language*, *35*, 477-496.

Meyer, A. S., & Damian, M. F. (2007). Activation of distractor names in the picture-picture interference paradigm. *Memory & cognition*, *35*(3), 494-503.

Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(6), 1146.

Meyer, A. S., & van der Meulen, F. F. (2000). Phonological priming effects on speech onset latencies and viewing times in object naming. *Psychonomic Bulletin & Review*, *7*(2), 314-319.

Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural networks*, *9*(8), 1265-1279.

Miall, R. C., & Reckess, G. Z. (2002). The cerebellum and the timing of coordinated eye and hand tracking. *Brain and Cognition*, *48*(1), 212-226.

Mitra, V., Nam, H., Espy-Wilson, C. Y., Saltzman, E., & Goldstein, L. (2011). Articulatory information for noise robust speech recognition. *Audio, Speech, and Language Processing, IEEE Transactions on*, *19*(7), 1913-1924.

Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 555.

Mowrey, R. A., & MacKay, I. R. (1990). Phonological primitives: Electromyographic speech error evidence. *The Journal of the Acoustical Society of America*, *88*(3), 1299-1312.

Mulligan, D., Lohse, K. R., & Hodges, N. J. (2015). An action-incongruent secondary task modulates prediction accuracy in experienced performers: evidence for motor simulation. *Psychological research*, 1-14.

Murray, W. S., & Forster, K. I. (2004). Serial mechanisms in lexical access: the rank hypothesis. *Psychological Review*, *111*(3), 721.

Mylopoulos, M. I., & Pereplyotchik, D. (2013). Is there any evidence for forward modeling in language production?. *Behavioral and Brain Sciences*, *36*(04), 368-369.

Narayanan, S., Nayak, K., Lee, S., Sethy, A., & Byrd, D. (2004). An approach to real-time magnetic resonance imaging for speech production. *The Journal of the Acoustical Society of America*, *115*(4), 1771-1776.

Narayanan, S., Toutios, A., Ramanarayanan, V., Lammert, A., Kim, J., Lee, S.,Nayak, K., Kim, Y-C., Zhu, Y., Bresch, E., Ghosh, P., Katsamanis, A., Proctor, M. .& Byrd, D. (2014). Real-time magnetic resonance imaging and electromagnetic articulography database for speech production research (TC). *The Journal of the Acoustical Society of America*, *136*(3), 1307-1311.

Natale, M. (1975). Social desirability as related to convergence of temporal speech patterns. *Perceptual and Motor Skills*, *40*(3), 827-830.

Navarrete, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. *Journal of Memory and Language*, *53*(3), 359-377.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*(3), 226-254.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. *Basic processes in reading: Visual word recognition*, *11*, 264-336.

Neiberg, D., Ananthakrishnan, G., & Engwall, O. (2008, September). The acoustic to articulation mapping: Non-linear or non-unique?. In *INTERSPEECH* (pp. 1485-1488).

Neppert, J., Petursson, M. Elemente einer akustischen Phonetik. Buske, Hamburg; 1992, cited in Seifert, E., Oswald, M., Bruns, U., Vischer, M., Kompis, M., & Haeusler, R. (2002). Changes of voice and articulation in children with cochlear implants. *International Journal of Pediatric Otorhinolaryngology*, *66*(2), 115-123

Neuper, |C., Wörtz, M., & Pfurtscheller, G. (2006). ERD/ERS patterns reflecting sensorimotor activation and deactivation. *Progress in brain research*, *159*, 211-222.

Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology*, *21*(4), 337-360.

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, *39*(2), 132-142.

Nielsen, K. Y. (2007). Implicit phonetic imitation is constrained by phonemic contrast. In *Proceedings of the 16th International Congress of Phonetic Sciences, Dudweiler, Germany* (pp. 1961-1964).

Nieuwland, M. S., Otten, M., & Van Berkum, J. J. (2007). Who are you talking about? Tracking discourse-level referential processing with event-related brain potentials. *Journal of Cognitive Neuroscience*, *19*(2), 228-236.

Niziolek, C. A., Nagarajan, S. S., Houde, J. F. (2013). What does motor efference copy represent? Evidence from speech production. *The Journal of Neuroscience*, *33*(41), 16110-16116.

Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive psychology*, *63*(1), 1-33.

Nyhus, E., & Curran, T. (2010). Functional role of gamma and theta oscillations in episodic memory. *Neuroscience & Biobehavioral Reviews*, *34*(7), 1023-1035.

Obhi, S. S., & Sebanz, N. (2011). Moving together: toward understanding the mechanisms of joint action. *Experimental brain research*, *211*(3), 329-336.

Oppenheim, G. M. (2013). Inner speech as a forward model?. *Behavioral and Brain Sciences*, *36*(04), 369-370.

Oppermann, F., Jescheniak, J. D., & Görges, F. (2014). Resolving competition when naming an object in a multiple-object display. *Psychonomic bulletin & review*, *21*(1), 78-84.

Osterhout, L., Kim, A., & Kuperberg, G. (2012). The neurobiology of sentence comprehension. In M.J. Spivey, K. McRae, and M.F. Joanisse (Eds.) *The Cambridge handbook of psycholinguistics* (pp. 365-390). New York: Cambridge University Press

Otten, M., Nieuwland, M. S., & Van Berkum, J. J. (2007). Great expectations: Specific lexical anticipation influences the processing of spoken language. *BMC neuroscience*, *8*(1), 1.

Otten, M., & Van Berkum, J. J. (2008). Discourse-based word anticipation during language processing: Prediction or priming?. *Discourse Processes*, *45*(6), 464-496.

Pardo, J. S. (2013). Measuring phonetic convergence in speech production. *Frontiers in psychology*, *4*.

Pardo, J. S., Jay, I. C., & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, Perception, & Psychophysics*, *72*(8), 2254-2264.

Pardo, J. S., Jay, I. C., & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, Perception, & Psychophysics*, *72*(8), 2254-2264.

Pardo, J. S., Jay, I. C., Hoshino, R., Hasbun, S. M., Sowemimo-Coker, C., & Krauss, R. M. (2013). Influence of role-switching on phonetic convergence in conversation. *Discourse Processes*, *50*(4), 276-300.

Perrin, F., & Garcıa-Larrea, L. (2003). Modulation of the N400 potential during auditory phonological/semantic interaction. *Cognitive Brain Research*, *17*(1), 36-47.

Philipp, A. M., & Prinz, W. (2010). Evidence for a role of the responding agent in the joint compatibility effect. *The Quarterly Journal of Experimental Psychology*, *63*(11), 2159-2171.

Piai, V., Roelofs, A., & Maris, E. (2014). Oscillatory brain responses in spoken word production reflect lexical frequency and sentential constraint. *Neuropsychologia*, *53*, 146-156.

Pianesi, F. (2007) Temporal Reference. In M. Everaert and H. van Riemsdijk (Eds.) *The Blackwell Companion to Syntax*. Blackwell Publishing; MA, USA

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(04), 329-347.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension?. *Trends in Cognitive Sciences*, *11*(3), 105-110.

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and cognitive processes*, *12*(5-6), 765-806.

Posner, M. I. (1975). Psychobiology of attention. In M.S. Gazzaniga (Ed.) *Handbook of psychobiology*, (pp. 441-480). London: Academic Press, Inc.

Poulet, J. F., & Hedwig, B. (2006). The cellular basis of a corollary discharge. *Science*, *311*(5760), 518-522.

Pouplier, M. (2008). The role of a coda consonant as error trigger in repetition tasks. *Journal of Phonetics*, *36*(1), 114-140.

Pouplier, M. (2007). Tongue kinematics during utterances elicited with the SLIP technique. *Language and Speech*, *50*(3), 311-341.

Pouplier, M. (2003). The dynamics of error. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2245-2248).

Pouplier, M., & Goldstein, L. (2010). Intention in articulation: Articulatory timing in alternating consonant sequences and its implications for models of speech production. *Language and Cognitive Processes*, *25*(5), 616-649.

Praamstra, P., & Stegeman, D. F. (1993). Phonological effects on the auditory N400 event-related brain potential. *Cognitive Brain Research*, *1*(2), 73-86.

Pulvermüller, F., Huss, M., Kherif, F., del Prado Martin, F. M., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences*, *103*(20), 7865-7870.

Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, *11*(5), 351-360.

Radeau, M., Besson, M., Fonteneau, E., & Castro, S. L. (1998). Semantic, repetition and rime priming between spoken words: behavioral and electrophysiological evidence. *Biological psychology*, *48*(2), 183-204.

Rahman, R. A., & Melinger, A. (2008). Enhanced phonological facilitation and traces of concurrent word form activation in speech production: An object-naming study with multiple distractors. *The Quarterly Journal of Experimental Psychology*, *61*(9), 1410-1440.

Ramnani, N., & Miall, R. C. (2004). A system in the human brain for predicting the actions of others. *Nature neuroscience*, *7*(1), 85-90.

Raney, G. E. (1993). Monitoring changes in cognitive load during reading: an event-related brain potential and reaction time analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition; Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(1), 51.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological review*, *107*(3), 460.

Rastle, K., Croot, K. P., Harrington, J. M., & Coltheart, M. (2005). Characterizing the motor execution stage of speech production: consonantal effects on delayed naming latency and onset duration. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(5), 1083.

Rastle, K., & Davis, M. H. (2002). On the complexities of measuring naming. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(2), 307.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, *124*(3), 372.

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

Reetz, H., & Jongman, A. (2011). Chapter 10: Acoustic characteristics of speech sounds in *Phonetics: Transcription, production, acoustics, and perception* (Vol. 34). John Wiley & Sons.

Riès, S., Legou, T., Burle, B., Alario, F. X., & Malfait, N. (2012). Why does picture naming take longer than word reading? The contribution of articulatory processes. *Psychonomic bulletin & review*, *19*(5), 955-961.

Riès, S., Legou, T., Burle, B., Alario, F. X., & Malfait, N. (2015). Corrigendum to "Why does picture naming take longer than word naming? The contribution of articulatory processes". *Psychonomic bulletin & review*, *22*(1), 309-311.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience.*, *27*, 169-192.

Roe, K., Jahn-Samilo, J., Juarez, L., Mickel, N., Royer, I., & Bates, E. (2000). Contextual effects on word production: A lifespan study. *Memory & Cognition*, *28*(5), 756-765.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, *64*(3), 249-284.

Roelofs, A. (2004). Comprehension-Based Versus Production-Internal Feedback in Planning Spoken Words: A Rejoinder to Rapp and Goldrick (2004).

Roelofs, A. (2015). Modeling of phonological encoding in spoken word production: From Germanic languages to Mandarin Chinese and Japanese. *Japanese Psychological Research*, *57*(1), 22-37.

Rommers, J., Meyer, A. S., & Huettig, F. (2013). Object shape and orientation do not routinely influence performance during language processing. *Psychological science*, 0956797613490746.

Rommers, J., Meyer, A. S., Praamstra, P., & Huettig, F. (2013). The contents of predictions in sentence comprehension: Activation of the shape of objects before they are referred to. *Neuropsychologia*, *51*(3), 437-447.

Rommers, J., Meyer, A. S., Piai, V., & Huettig, F. (2013). Constraining the involvement of language production in comprehension: A comparison of object naming and object viewing in sentence context. Talk presented at the 19th Annual Conference on Architectures and Mechanisms for Language Processing [AMLaP 2013]. Marseille, France. 2013-09-02 - 2013-09-04

Roon, K. D., & Gafos, A. I. (2015). Perceptuo-motor effects of response-distractor compatibility in speech: beyond phonemic identity. *Psychonomic bulletin & review*, *22*(1), 242-250.

Rose, S. B., Spalek, K., & Rahman, R. A. (2015). Listening to puns elicits the co-activation of alternative homophone meanings during language production. *PloS one*, *10*(6), e0130853.

Rothermich, K., & Kotz, S. A. (2013). Predictions in speech comprehension: fMRI evidence on the meter–semantic interface. *NeuroImage*, *70*, 89-100.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*, 696–735.

Sakayori, S., Kitama, T., Chimoto, S., Qin, L., & Sato, Y. (2002). Critical spectral regions for vowel identification. *Neuroscience research*, *43*(2), 155-162

Schiller, N. O.**,** Horemans, I., Ganushchak, L. & Koester, D., (2009). Event-related brain potentials during the monitoring of speech errors. *NeuroImage*, *44*, 520-530.

Schmidt, T., & Schmidt, F. (2009). Processing of natural images is feedforward: A simple behavioral test. *Attention, Perception, & Psychophysics*, *71*(3), 594-606.

Schomers, M. R., Kirilina, E., Weigand, A., Bajbouj, M., & Pulvermüller, F. (2014). Causal influence of articulatory motor cortex on comprehending single spoken words: TMS evidence. *Cerebral Cortex*, bhu274.

Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, *29*(1), 86-102.

Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of memory and language*, *29*(1), 86-102.

Schuhmann, T., Schiller, N. O., Goebel, R., & Sack, A. T. (2012). Speaking of which: dissecting the neurocognitive network of language production in picture naming. *Cerebral Cortex*, *22*(3), 701-709.

Scobbie, J. M., Wrench, A. A., van der Linden, M. (2008). "Head-probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement". In *Proceedings of the 8th International Seminar on Speech Production* (pp. 373-376).

Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action: Candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience, 10* (4), 295–302.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in cognitive sciences*, *10*(2), 70-76.

Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science*, *1*(2), 353-367.

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: just like one's own?. *Cognition*, *88*(3), B11-B21.

Sebanz, N., Knoblich, G., Prinz, W., & Wascher, E. (2006). Twin peaks: An ERP study of action planning and control in coacting individuals. *Journal of cognitive neuroscience*, *18*(5), 859-870.

Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition*, *53*(2), 91-127.

Severens, E., Janssens, I., Kühn, S., Brass, M., & Hartsuiker, R. J. (2011). When the brain tames the tongue: Covert editing of inappropriate language. *Psychophysiology*, *48*(9), 1252-1257.

Severens, E., Ratinckx, E., Ferreira, V. S., & Hartsuiker, R. J. (2008). Are phonological influences on lexical (mis)selection the result of a monitoring bias? *The Quarterly Journal of Experimental Psychology, 61*, 1687–1709.

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics*, *66*(3), 422-429.

Slis, A., & Van Lieshout, P. (2013). The effect of phonetic context on speech movements in repetitive speech. *The Journal of the Acoustical Society of America*, *134*(6), 4496-4507.

Sommer, M. A., & Wurtz, R. H. (2004). What the brain stem tells the frontal cortex. I. Oculomotor signals sent from superior colliculus to frontal eye field via mediodorsal thalamus. *Journal of Neurophysiology*, *91*(3), 1381-1402

Song, J. Y., Demuth, K., Shattuck-Hufnagel, S., & Ménard, L. (2013). The effects of coarticulation and morphological complexity on the production of English coda clusters: Acoustic and articulatory evidence from 2-year-olds and adults using ultrasound. *Journal of Phonetics*, *41*(3), 281-295.

Smolensky, P. (1988). The constituent structure of connectionist mental states: A reply to Fodor and Pylyshyn. *The Southern Journal of Philosophy*, *26*(S1), 137-161.

Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of comparative and physiological psychology*, *43*(6), 482.

Springer, A., Brandstädter, S., Liepelt, R., Birngruber, T., Giese, M., Mechsner, F., & Prinz, W. (2011). Motor execution affects action prediction. *Brain and Cognition*, *76*, 26–36.

Stadler, W., Schubotz, R. I., von Cramon, D. Y., Springer, A., Graf, M., & Prinz, W. (2011). Predicting and memorizing observed action: differential premotor cortex involvement. *Human brain mapping*, *32*(5), 677-687.

Stanley, J., Gowen, E., & Miall, R. C. (2007). Effects of agency on movement interference during observation of a moving dot stimulus. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 915.

Stanovich, K. E., & West, R. F. (1979). Mechanisms of sentence context effects in reading: Automatic activation and conscious attention. *Memory & Cognition*, *7*(2), 77-85.

Starreveld, P. A., & La Heij, W. (1995). Semantic interference, orthographic facilitation, and their interaction in naming tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(3), 686.

Stevens, J. A. (2005). Interference effects demonstrate distinct roles for visual and motor imagery during the mental representation of human action. *Cognition*, *95*(3), 329-350.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J.P., Yoon, K-E., & Levinson, S. C. (2009). Universality and cultural specificity in turn-taking in conversation. In *Proceedings of the National Academy of Science* (Vol. 106, No. 26, pp. 10587-92).

Strijkers, K., & Costa, A. (2011). Riding the lexical speedway: a critical review on the time course of lexical selection in speech production. *Frontiers in psychology*, *2*.

Strijkers, K., Holcomb, P. J., & Costa, A. (2011). Conscious intention to speak proactively facilitates lexical access during overt object naming. *Journal of Memory and Language*, *65*(4), 345-362.

Stone, M. (2005). A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics & Phonetics*, *19*(6-7), 455-501.

Sun, H., Blakely, T. M., Darvas, F., Wander, J. D., Johnson, L. A., Su, D. K., ... & Ojemann, J. G. (2015). Sequential activation of premotor, primary somatosensory and primary motor areas in humans during cued finger movements. *Clinical Neurophysiology*, *126*(11), 2150-2161.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. E. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 632-634

Taylor, W.L. (1953) 'Cloze' procedure: a new tool for measuring readability. *Journalism Quarterly*, *30*, 415

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P. & Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, *17*(2), 273-281.

Tian, X., & Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in Psychology*, *1*(166), 10-3389.

Tilsen, S. (2013). A dynamical model of hierarchical selection and coordination in speech planning. *PloS one*, *8*(4), e62800.

Tilsen, S. (2009). Multitimescale dynamical interactions between speech rhythm and gesture. *Cognitive science*, *33*(5), 839-879.

Tilsen, S. (2007). Vowel-to-vowel coarticulation and dissimilation in phonemic-response priming. *UC Berkeley Phonology Lab 2007 Annual Report*, 416-458.

Tsai, C. C., Kuo, W. J., Hung, D. L., & Tzeng, O. J. (2008). Action co-representation is tuned to other humans. *Journal of Cognitive Neuroscience*, *20*(11), 2015-2024.

Tsai, C. C., Kuo, W. J., Jing, J. T., Hung, D. L., & Tzeng, O. J. L. (2006). A common coding framework in self–other interaction: evidence from joint action task. *Experimental Brain Research*, *175*(2), 353-362.

Tulving, E., & Gold, C. (1963). Stimulus information and contextual information as determinants of tachistoscopic recognition of words. *Journal of Experimental Psychology*, *66*(4), 319.

Underwood, G., & Briggs, P. (1984). The development of word recognition processes. *British Journal of Psychology*, *75*(2), 243-255.

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman,V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 443–467.

Van Rullen, R., & Thorpe, S. J. (2001). The time course of visual processing: from early perception to decision-making. *Journal of cognitive neuroscience*, *13*(4), 454-461.

van Schie, H. T., Mars, R. B., Coles, M. G., & Bekkering, H. (2004). Modulation of activity in medial frontal and motor cortices during error observation. *Nature neuroscience*, *7*(5), 549-554.

Vasilyeva, M., Waterfall, H., Gámez, P. B., Gómez, L. E., Bowers, E., & Shimpi, P. (2010). Cross-linguistic syntactic priming in bilingual children. *Journal of child language*, *37*(5), 1047.

Vatikiotis-Bateson, E., Barbosa, A. V., & Best, C. T. (2014). Articulatory coordination of two vocal tracts. *Journal of Phonetics, 44*, 167-181.

Venezia, J. H., Saberi, K., Chubb, C., & Hickok, G. (2012). Response bias modulates the speech motor system during syllable discrimination. *Frontiers in Psychology, 3*. doi: 10.3389/fpsyg.2012.00157

Verfaillie, K., & Daems, A. (2002). Representing and anticipating human actions in vision. *Visual Cognition*, *9*(1-2), 217-232.

Vissers, C. T. W., Chwilla, D. J., & Kolk, H. H. (2006). Monitoring in language perception: The effect of misspellings of words in highly constrained sentences. *Brain Research, 1106*(1), 150-163.

Vlainic, E., Liepelt, R., Colzato, L. S., Prinz, W., & Hommel, B. (2009). The virtual co-actor: the social Simon effect does not rely on online feedback from the other. *Embodied and grounded cognition*, 102.

Von Holst, E., & Mittelstadt, H. (1950). 1973. *The reafference principle (R. Martin, Trans.) The Behavioral Physiology of Animals and Man. The Collected Papers of Erich von Holst*.

Vousden, J. I., Brown, G. D., & Harley, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive psychology*, *41*(2), 101-175.

Wang, L., Jensen, O., Van den Brink, D., Weder, N., Schoffelen, J. M., Magyari, L., .Hagoort, P., & Bastiaansen, M. (2012). Beta oscillations relate to the N400m during language comprehension. Human brain mapping, 33(12), 2898-2912.

Wang, J., Mathalon, D. H., Roach, B. J., Reilly, J., Keedy, S. K., Sweeney, J. A., & Ford, J. M. (2014). Action planning and predictive coding when speaking. *NeuroImage*, *91*, 91-98.

Watkins, K., & Paus, T. (2004). Modulation of motor excitability during speech perception: the role of Broca's area. *Journal of Cognitive Neuroscience, 16*(6), 978-987.

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, *41*(8), 989-994.

Weber, A., Braun, B., & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language* and Speech, 49(3), 367-392.

Welsh, T. N. (2009). When 1+ 1= 1: The unification of independent actors revealed through joint Simon effects in crossed and uncrossed effector conditions. *Human Movement Science*, *28*(6), 726-737.

Wenke, D., Atmaca, S., Holländer, A., Liepelt, R., Baess, P., & Prinz, W. (2011). What is shared in joint action? Issues of co-representation, response conflict, and agent identification. *Review of Philosophy and Psychology*, *2*(2), 147-172.

Whalen, D. H. (1990). Coarticulation is largely planned 7/3. *Journal of Phonetics*, *18*, 3-35.

Wheeldon, L. (2003). Inhibitory form priming of spoken word production. *Language and Cognitive Processes*, *18*(1), 81-109.

Wicha, N. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not Pope: human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters*, *346*(3), 165-168.

Wicha, N. Y., Moreno, E. M., & Kutas, M. (2003). Expecting gender: An event related brain potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in Spanish. *Cortex*, *39*(3), 483-508.

Wicha, N. Y., Moreno, E., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, *16*(7), 1272-1288.

Wicha, N. Y., Orozco-Figueroa, A., Reyes, I., Hernandez, A., de Barreto, L. G., & Bates, E. A. (2005). When zebras become painted donkeys: Grammatical gender and semantic priming interact during picture integration in a spoken Spanish sentence. *Language and cognitive processes*, *20*(4), 553-587.

Willems, R.M., & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: A review. *Brain Language*, *101*, 278–289.

Wilshire, C., Singh, S., & Tattersall, C. (2015). Serial order in word form retrieval: New insights from the auditory picture–word interference task. *Psychonomic bulletin & review*, 1-7.

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological bulletin*, *131*(3), 460.

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, *7*(7), 701-702.

Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic bulletin & review*, *12*(6), 957-968

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *358*(1431), 593-602.

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current biology*, *11*(18), R729-R732.

Wong, D., Pisoni, D. B., Learn, J., Gandour, J. T., Miyamoto, R. T., & Hutchins, G. D. (2002). PET imaging of differential cortical activation by monaural speech and nonspeech stimuli. *Hearing research*, *166*(1), 9-23.

Woodhead, Z. V. J., Barnes, G. R., Penny, W., Moran, R., Teki, S., Price, C. J., & Leff, A. P. (2014). Reading front to back: MEG evidence for early feedback effects during word recognition. *Cerebral Cortex*, *24*(3), 817-825.

Wrench, A. A., & Scobbie, J. M. (2008). High-speed Cineloop Ultrasound vs. Video Ultrasound Tongue Imaging: Comparison of Front and Back Lingual Gesture Location and Relative Timing. In *Proceedings of the Eighth International Seminar on Speech Production (ISSP)*.

Yuen, I., Davis, M. H., Brysbaert, M., Rastle, K. (2010). Activation of articulatory information in speech perception. *Proceedings of the National Academy of Sciences*, *107*(2), 592-597.

Zhao, H., La Heij, W., & Schiller, N. O. (2012). Orthographic and phonological facilitation in speech production: New evidence from picture naming in Chinese. *Acta psychologica*, *139*(2), 272-2