Speech Perception, Phonological Sensitivity, and Articulation in Early Vocabulary Development

Iris-Corinna Schwarz

MA (Psychology)
MA (Education Science)
Catholic University of Eichstätt-Ingolstadt, Germany

A thesis submitted for the degree of Doctor of Philosophy
University of Western Sydney
May 2007
I hereby declare that this submission is my own work and, to the best of my 
knowledge, it contains no material previously published or written by any other 
person, nor material which has been accepted for the award of any other degree or 
diploma at the University of Western Sydney, or at any other educational 
institution, except where due acknowledgement is made in the thesis.

I also declare that the intellectual content of this thesis is the product of my own 
work, except to the extent that assistance from others in the project’s design and 
conception is acknowledged.
“All things are possible for him who believes.”
Mark 9:23
New King James Version

„Alle Dinge sind möglich dem, der da glaubt.“
Mark 9:23
Lutherische Übersetzung
ACKNOWLEDGEMENTS

In all your ways acknowledge Him...

(Proverbs 3:6, NKJV)

A doctoral thesis is not the work of one single person alone; it grows out of contemplation, conversation, and collaboration with others. I could not have done it without you and thank you all for your unique contribution to the mosaic called “PhD”.

To my principal supervisor Denis Burnham: You truly were my “Doktorvater”. Thank you for your way of guiding, challenging, supporting, advising, and improving my work and me. Your belief in me gave me the confidence that I can actually do this.

To my supervisors Judy Bowey, Christine Kitamura, and Mark Wiggins: Thank you for your professional advice and inspiration. You taught me a lot.

To our lab manager Colin Schoknecht: You are the calm eye in every technical storm. Thank you for coming to the rescue at every emergency - computer-related or not. You are the heart of the lab.

To the MARCS programmers Rua Haszard Morris and Johnson Chen: Thank you for turning ideas into a functioning computer program, and providing technical support at the drop of a hat. I admire your skills.

To Arman Abrahamyian: Thank you for being my personal computer magician even in the most hectic times.

To my fantastic proofreading squad who was literally there for me until the wee hours of the morning: Rikke Bundgård-Nielsen, Doğu Erdener, Renee Glass, Jess Hartcher-O’Brien, Andrew Mathas, Karen Mattock, Nathan Perry, and Damien Smith. You have eyes like a hawk.

To our administrative genies Gail Charlton, Mel Gallagher, Darlene Williams, and Karen McConachie: You make MARCS work.
To all MARCSists, past, current, and future members: You make the lab what it is.

To all the participating families: Thank you for making this study a part in the life of your family for 6 months. It would not exist without your commitment.

To my research assistant and friend Esther Raudonat: Thank you for years of recording thousands of OZI items by typing 0 and 1. Thanks also to Sarah Kilgour, Alexis Collins, and Samia Hawatt for their assistance on the OZI project.

To my Australian “family”, Marcela Bilek, Andrew Mathas, Alexandr Mathas, Erika Bilek, John Spurgin and Kipper: Thank you for our home. You have inspired me to live life.

To my other Australian “family”, Sandra, Stuart, Karen and Alison Mattock and Stan: Thank you for adopting me as a daughter, a sister, and a granddaughter.

To my Australian friends, Arman Abrahamyan (open mind), Mark Antoniou (computer touch), Sharon Atkins, Thomas Britz (omnilocal laptopping), Rikke Bundgård-Nielsen (perfect balance), Karina Buso, Mark Buso (I will not forget), Eric Desjardins (French humour), Doğu Erdener (impossible partner in crime), Renee Glass, Tony Hails (you made me an allrounder), Kara Haley (more than words), Jess Hartcher-O’Brien (inspirational genie), Angela Hennessy (hands on life), Brad Jones, Dasha Konduryina (teen spirit), Nicole Low, Cassidy Lynne, Brett Molesworth (efficiency example), Tamara Morton (integrity), Anna Notley (twisted knot), Rachel Parker (two minds alike), Alison Peachey, Nathan Perry, Amy Priestly (Aims), Kristy Riggs, Damien Smith (I need the pocket version), Glenn Stephenson, Yvette Thornborough, Kylie Tyler, Michael Tyler (someone to look up to), Leighton Wren (friendly punches), Justin Yip (much more than flight booking): I have enjoyed every moment of walking with you all. Thank you for keeping me sane.

To my overseas friends, Albrecht Fleischmann, Carolin Gaßmann, Heike Haberl, Jenny Haide, Jan Ingenhaag, Judith Jabs, Martina Kerscher, Erika Kohlmeier,
Otto Laudien, Heike Leopold, Tanja Meding, Bernhard Müller, Steffi Rihm, Alexandra Roth, Eveline Ruhland, Roswitha Schillinger, Winfried Schillinger, Michael Schölz, Ursula Schölz, Christian Schröder, Thomas Schröder, Sabine Schwarz, Fritz Schwarz, Anja Templer, Sebastian Zips: Thank you for keeping me sane (and up to date) across the oceans. I know what kind of loyalty this takes.

To my parents Dietz-Ulrich and Katrin Schwarz, my sister Mareike Schwarz, my aunt Ursula Sachs, my aunt and uncle Hilke and Friedl Mayr: Thank you for being there for me.

To my friend, fiancé, and future husband Ulf Kalla: Thank you for your unfaltering support and love throughout the years. You kept my eyes on the goal and were always there for me. I know I am lucky.

Iris-Corinna Schwarz

Sydney, 11th May 2007
Abstract

Speech perception, articulation, and word learning are three major tiers of language development in young children, integrating perceptual and productive language abilities. Infant speech perception precedes speech production and is the basis for native language learning. In speech production, children refine their articulation skills beginning with their first vocalic utterances until they reach adult performance level. The third tier describes children’s vocabulary development from their first words to their established receptive and productive lexicon after the vocabulary spurt. Speech perception, articulation, and word learning interact at the level of lexical representations.

By investigating the relationship between the attention to phonological detail in speech and word learning, the degree of phonological detail in the lexical representations can be inferred. This relationship can be described by two models: the vocabulary-driven and phonology-driven model. The vocabulary-driven model proposes that the structure of the lexicon influences attention to phonological detail in speech perception, and this model is consistent with the Lexical Restructuring Model. On the other hand the phonology-driven model proposes that vocabulary increases as a result of increased attention to phonological detail in speech.

To infer the phonological specifications of lexical representations of words in 2½- to 3-year-olds, the variables vocabulary, phonological sensitivity, language-specific speech perception and articulation accuracy were tested in a longitudinal study with 60 participants. For these variables, new measures were developed, adapted, and tested. It was found that phonological sensitivity at 30 months predicted vocabulary at 33 months, but not the opposite. This supports the prediction of the phonology-driven model. However, in an augmented version of the vocabulary-driven model that included all variables, articulation at 30 months was found to predict phonological sensitivity at 33 months. These results are
discussed in the light of the Lexical Restructuring Model, and the interaction of speech perception, articulation skills, and lexical representations, and suggestions for future research are offered.
# TABLE OF CONTENTS

**ACKNOWLEDGEMENTS** ........................................................................................................ IV

**ABSTRACT** .......................................................................................................................... VII

**LIST OF FIGURES** ................................................................................................................ XVI

**LIST OF TABLES** .................................................................................................................. XVII

**PREFACE** ........................................................................................................................... XVIII

**CHAPTER I: SPEECH PERCEPTION** ...................................................................................... I

1.1. Speech perception fundamentals .......................................................................................... 2

1.1.1. The speech perception challenge .................................................................................... 2

1.1.1.1. Speech stream segmentation ....................................................................................... 2

1.1.1.2. Talker variability ......................................................................................................... 3

1.1.1.3. Context effects ............................................................................................................ 3

1.1.2. Units of speech perception .............................................................................................. 4

1.1.2.1. Phones and Phonemes ............................................................................................... 4

1.1.2.2. Tones and Tone-memes ............................................................................................. 4

1.1.2.3. Syllables .................................................................................................................... 5

1.1.2.4. Words ....................................................................................................................... 6

1.1.2.5. Suprasegmental information: Prosody and stress ..................................................... 6

1.1.3. Speech perception through a specialised speech mode? ................................................... 7

1.1.4. A theoretical framework of speech perception ................................................................. 9

1.1.4.1. The Motor Theory ..................................................................................................... 9

1.1.4.2. The Perceptual Assimilation Model ........................................................................... 10

1.2. Speech perception development: Models ........................................................................... 12

1.2.1. A theoretical framework for speech perception development:

    Universal, Attunement, and Perceptual Learning Theories ............................................... 12

1.2.2. The Robust and Fragile Hypothesis ............................................................................. 15

1.2.3. The Native Language Magnet Theory ......................................................................... 16

1.3. Speech perception development: Research ....................................................................... 17

1.3.1. Speech perception in infancy: Prenatal, phonetic, and phonemic periods ...................... 18

1.3.1.1. Prenatal speech perception ......................................................................................... 18

1.3.1.2. Phonetic speech perception from birth to six months ............................................. 19

1.3.1.3. Phonetic speech perception from six to twelve months .......................................... 21

1.3.2. Speech perception at preschool age: The semantic period .......................................... 23

1.3.3. Speech perception in early readers: The orthographic period ...................................... 25

1.3.4. Speech perception in toddlers? .................................................................................... 26

**CHAPTER II: SPEECH PRODUCTION: ARTICULATION** ..................................................... 28

2.1. Speech production fundamentals ....................................................................................... 29

2.1.1. Physiology of the articulatory system ............................................................................ 29

2.1.1.1. The articulators ......................................................................................................... 30

2.1.1.2. Speech sound articulation and classification ............................................................. 31

2.1.2. The system of phonology .............................................................................................. 33

2.1.2.1. Vowels ..................................................................................................................... 34
2.1.2.2. Consonants ................................................................. 34
2.1.2.3. Tones ........................................................................... 35
2.1.3. Models of speech production ............................................... 35
2.1.4. Measurement of speech production accuracy ......................... 37

2.2. Speech production development ............................................... 38
2.2.1. Speech production in the first year of life .............................. 39
2.2.2. Speech production in the second year of life ......................... 41
2.2.3. Speech production in the third year of life ............................ 42
2.2.4. Articulation refinement in childhood .................................... 44

2.3. Articulation acquisition .......................................................... 45
2.3.1. Acquisition of vowels ......................................................... 45
2.3.2. Acquisition of consonants ................................................. 46
2.3.3. Acquisition of tone ............................................................ 47

2.4. Articulation acquisition in context ........................................... 48
2.4.1. Articulation and speech perception ..................................... 48
2.4.2. Articulation and the ambient language ................................. 50
2.4.3. Articulation and word learning .......................................... 51

Chapter III: Speech Production: Word Learning .............................. 53

3.1. Elements of word learning ..................................................... 54
3.1.1. The word as language unit ................................................ 54
3.1.2. Characteristics of lexical entries ......................................... 55
3.1.2.1. Word frequency and phonotactic probability .................... 55
3.1.2.2. Neighbourhood density ............................................... 56
3.1.3. Receptive and productive lexicon ....................................... 57
3.1.4. Mean length of utterance .................................................. 58

3.2. A theoretical framework for word learning ............................... 59
3.2.1. The process of word learning ............................................. 59
3.2.1.1. Effects of input on lexical acquisition ............................ 59
3.2.1.2. Word acquisition strategies and errors ........................... 60
3.2.1.3. Principles of word learning .......................................... 61
3.2.2. Modelling lexical acquisition ............................................ 63
3.2.2.1. Challenges for a Model of Lexical Acquisition ............... 63
3.2.2.2. The Emergentist Coalition Model ................................. 65

3.3. Measuring aspects of word learning ........................................ 67
3.3.1. Statistical word learning .................................................. 67
3.3.2. Word recognition and further segmentation abilities ............. 69
3.3.3. Age of acquisition of words ............................................. 71
3.3.4. Measuring lexicon size .................................................... 72

3.4. Vocabulary development ....................................................... 74
3.4.1. The lexicon in the first year of life .................................... 74
3.4.2. The lexicon in the second year of life: The vocabulary spurt .... 74
3.4.3. The lexicon in the third year of life ................................... 76
3.4.4. Further lexical development ............................................. 77

3.5. Word learning in context ....................................................... 77
CHAPTER IV: INFERRING THE STRUCTURE OF LEXICAL REPRESENTATIONS .......................... 80

4.1. The relationship of the lexicon and phonological detail in speech perception ........................................... 81

4.1.1. The development of the lexicon-phonology relationship .............................................................. 81

4.1.1.1. Phonological specificity of early representations ............................................................... 81

4.1.1.2. Interlude: The Word Recognition and Phonetic Structure Acquisition Model (WRAPSA) .......... 83

4.1.1.3. Lexical representations in the second year .................................................................................. 84

4.1.1.4. Protracted development in childhood ......................................................................................... 86

4.1.2. Lexical restructuring and the vocabulary ......................................................................................... 88

4.2. Predictions for vocabulary-driven and phonology-driven models ......................................................... 91

4.3. Study Aims ........................................................................................................................................... 92

CHAPTER V: METHOD: TASKS, TASK DEVELOPMENT, AND PROCEDURES .................... 94

5.1. Participants ............................................................................................................................................... 95

5.2. Apparatus and Experimental Chamber .............................................................................................. 96

5.2.1. Apparatus ......................................................................................................................................... 96

5.2.2. Experimental Chamber .................................................................................................................. 98

5.3. Experimental Design and Variables ................................................................................................... 99

5.3.1. Vocabulary ....................................................................................................................................... 99

5.3.1.1. The Australian English vocabulary inventory OZI ......................................................... 99

5.3.1.2. Peabody Picture Vocabulary Test III .................................................................................... 103

5.3.2. Phonological Sensitivity ................................................................................................................. 105

5.3.2.1. Mispronunciation Detection ............................................................................................... 106

5.3.2.2. Rhyme Detection ...................................................................................................................... 109

5.3.2.3. Nonword Repetition ................................................................................................................. 111

5.3.3. Language Specific Speech Perception .......................................................................................... 114

5.3.4. Articulation Accuracy ...................................................................................................................... 119

5.3.5. General Cognitive Functioning ..................................................................................................... 121

5.4. Overall Procedure .............................................................................................................................. 122

CHAPTER VI: RESULTS .................................................................................................................. 124

6.1. Results for individual tests .................................................................................................................. 125

6.1.1. Vocabulary ....................................................................................................................................... 126

6.1.1.1. Australian English vocabulary inventory OZI ................................................................. 126

6.1.1.2. Peabody Picture Vocabulary Test III .................................................................................... 127

6.1.2. Phonological sensitivity ................................................................................................................... 127

6.1.2.1. The phonological sensitivity subtests ..................................................................................... 128

6.1.2.2. Phonological sensitivity component scores .......................................................................... 129

6.1.3. Language-specific speech perception LSSP .................................................................................. 131

6.1.4. Articulation accuracy ....................................................................................................................... 133

6.1.5. General cognitive functioning ........................................................................................................ 133

6.1.6. Summary of individual test results ................................................................................................. 135

6.2. Correlations between the variables .................................................................................................... 135

6.3. Regression-based path analyses ......................................................................................................... 139

6.3.1. Data screening ............................................................................................................................... 139
6.3.2. Basic models ................................................................. 140
6.3.3. Augmented models ....................................................... 142
6.4. Summary of results ......................................................... 144

CHAPTER VII: DISCUSSION .......................................................... 145
7.1. Vocabulary-driven or phonology-driven model? ......................... 146
  7.1.1. Discussion of the basic models ...................................... 146
  7.1.2. Discussion of the augmented models ............................... 148
7.2. Critical review of individual measures ................................... 149
  7.2.1. Vocabulary ..................................................................... 149
  7.2.2. Phonological Sensitivity ............................................... 150
  7.2.3. Language-specific speech perception ............................... 150
  7.2.4. Articulation accuracy .................................................. 152
7.3. In a nutshell: The role of articulation .................................... 153
7.4. The complete picture ......................................................... 155
7.5. Implications for future research .......................................... 157

REFERENCES .............................................................................. 160
CONTENTS OF APPENDICES ON SUPPLEMENTARY CD

APPENDIX TO CHAPTER II
2.1.1.2. Speech sound articulation and classification
CD Places of Articulation
2.1.2. The system of phonology
CD IPA Chart (2005)

APPENDIX TO CHAPTER V
5.1. Participants
CD Consent Form
CD Participant Age
CD Thankyou Letter
CD Sponsor Logos
CD T-Shirt
CD Sponsor Letters (A and B)
CD Young Scientist Certificate
CD Ethics Approval
5.3.1.1. Australian English Vocabulary Inventory OZI
CD Permission Request
CD CDI Board Permission
CD Shared Words (OZI, CDI, & Oxford CDI)
CD OZI Questionnaire
CD Comparison of Item Number
CD Update OZI (2006)
5.3.2.1. Mispronunciation detection
CD Scoring Sheet
CD Pictures
CD Picture Pilot Results
5.3.2.2. Rhyme detection
CD Oddity Scoring Sheet
CD Detection Scoring Sheet
5.3.2.3. Nonword repetition
CD Scoring Sheet
CD Nonword List
5.3.3. Language-specific speech perception
CD Thai Stimulus List
CD Thai Stimulus Selection
CD Stimulus Rating Sheet
CD Counterbalancing Conditions
CD LSSP Task Program Script
CD LSSP Task Output Script
CD Button Box
5.3.4. Articulation accuracy
CD Pictures
CD Irfanview Screen
CD Scoring sheet
5.3.5. General cognitive functioning
   CD Scoring Sheet

APPENDIX TO CHAPTER VI

6.1. Results for individual tests
   CD Data Screening
   CD Normality Test Script
6.1.1. Vocabulary
   CD Raw Data
   CD z-Scores
6.1.1.2. Peabody Picture Vocabulary Test III
   CD Development
6.1.2. Phonological sensitivity
   CD Raw Data
   CD Mispronunciation Development
   CD Rhyme Development
   CD Nonword Development
   CD Components 33
   CD Components 36
   CD Components 30
   CD Component Scores
   CD Two Components 33
   CD Two Components 36
   CD Basic Vocabulary-driven Model 33(2)
   CD Basic Vocabulary-driven Model 36 (2)
   CD Basic Phonology-driven Model 36(2)
   CD Augmented Vocabulary-driven Model 33(2)
   CD Augmented Vocabulary-driven Model 36 (2)
   CD Augmented Phonology-driven Model 36(2)
6.1.3. Language-specific speech perception
   CD Raw Data
   CD N Development
   CD NN Development
   CD T Development
   CD N-NN Development
   CD N-T Development
6.1.4. Articulation accuracy
   CD Raw Data
   CD Development
6.1.5. Cognitive functioning
   CD Raw Data
   CD Norm 58
   CD Augmented Vocabulary-driven Model 36(58)
   CD Augmented Phonology-driven Model 36(58)
6.3. Regression-based path analyses
6.3.1. Data screening
   CD Time 1
   CD Time 2
6.3.2. Basic models
   CD Vocabulary-driven Model 33
   CD Vocabulary-driven Model 36
   CD Phonology-driven Model 33
   CD Phonology-driven Model 36

6.3.3. Augmented models
   CD Vocabulary-driven Model 33
   CD Vocabulary-driven Model 36
   CD Phonology-driven Model 33
   CD Phonology-driven Model 36
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Categories of the native phonological space</td>
<td>11</td>
</tr>
<tr>
<td>1.2</td>
<td>Category assimilation types</td>
<td>11</td>
</tr>
<tr>
<td>1.3</td>
<td>Nonnative contrast assimilation</td>
<td>12</td>
</tr>
<tr>
<td>1.4</td>
<td>Models of speech perception development</td>
<td>13</td>
</tr>
<tr>
<td>1.5</td>
<td>The perceptual magnet effect</td>
<td>16</td>
</tr>
<tr>
<td>2.1</td>
<td>The human speech organs</td>
<td>29</td>
</tr>
<tr>
<td>3.1</td>
<td>The development of speech perception, speech production, and lexical acquisition in the first year of life</td>
<td>78</td>
</tr>
<tr>
<td>5.1</td>
<td>Circuit diagram for equipment set-up</td>
<td>97</td>
</tr>
<tr>
<td>5.2</td>
<td>Testroom set-up</td>
<td>98</td>
</tr>
<tr>
<td>5.3a</td>
<td>Comparison of CDI and OZI in 24-month-olds</td>
<td>102</td>
</tr>
<tr>
<td>5.3b</td>
<td>Comparion of CDI and OZI in 30-month-olds</td>
<td>102</td>
</tr>
<tr>
<td>5.4</td>
<td>Preliminary vocabulary development in Australian toddlers</td>
<td>103</td>
</tr>
<tr>
<td>5.5</td>
<td>PPVT administration</td>
<td>104</td>
</tr>
<tr>
<td>5.6</td>
<td>Mispronunciation detection task administration</td>
<td>108</td>
</tr>
<tr>
<td>5.7a</td>
<td>Screen for rhyming word pairs</td>
<td>111</td>
</tr>
<tr>
<td>5.7b</td>
<td>Screen for non-rhyming word pairs</td>
<td>111</td>
</tr>
<tr>
<td>5.8</td>
<td>Rhyme detection task administration</td>
<td>111</td>
</tr>
<tr>
<td>5.9</td>
<td>Pemmie</td>
<td>112</td>
</tr>
<tr>
<td>5.10</td>
<td>Nonword repetition task administration</td>
<td>113</td>
</tr>
<tr>
<td>5.11</td>
<td>LSSP task administration</td>
<td>118</td>
</tr>
<tr>
<td>5.12</td>
<td>Articulation task administration</td>
<td>120</td>
</tr>
<tr>
<td>5.13a</td>
<td>Cognitive Functioning test administration: Object Series</td>
<td>121</td>
</tr>
<tr>
<td>5.13b</td>
<td>Cognitive Functioning test administration: Matrices</td>
<td>121</td>
</tr>
<tr>
<td>6.1</td>
<td>OZI vocabulary scores scattergram</td>
<td>126</td>
</tr>
<tr>
<td>6.2</td>
<td>PPVT vocabulary score development</td>
<td>127</td>
</tr>
<tr>
<td>6.3</td>
<td>Phonological sensitivity development</td>
<td>128</td>
</tr>
<tr>
<td>6.4</td>
<td>LSSP development</td>
<td>132</td>
</tr>
<tr>
<td>6.5</td>
<td>Articulation accuracy development</td>
<td>133</td>
</tr>
<tr>
<td>6.6</td>
<td>General cognitive functioning scattergram</td>
<td>134</td>
</tr>
<tr>
<td>6.7</td>
<td>Correlation scatterplots for vocabulary and phonological sensitivity</td>
<td>138</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1.1.: The five lexical tones in Thai ................................................. 5
Table 5.1.: Switch settings ........................................................................... 97
Table 5.2.: Mispronunciation task items: Difficulty, position, and stress .... 107
Table 5.3.: Mispronunciation discrimination index ..................................... 109
Table 5.4.: Tone classes of Thai sound recordings .................................... 115
Table 5.5.: Duration of the Thai sound recordings ..................................... 115
Table 5.6.: Wordlist of the Queensland Articulation Test ......................... 119
Table 6.1.: Phonological sensitivity development ...................................... 129
Table 6.2.: Correlations between phonological sensitivity measures ........ 130
Table 6.3.: Principal component analyses: Component loadings and coefficients ........................................................................................................... 130
Table 6.4.: LSSP development ................................................................... 132
Table 6.5.: Correlation matrix of all variables .......................................... 137
Table 6.6.: Regression table of the basic vocabulary-driven model .......... 141
Table 6.7.: Regression table of the basic phonology-driven model .......... 141
Table 6.8.: Regression table of the augmented vocabulary-driven model .... 143
Table 6.9.: Regression table to the augmented phonology-driven model ..... 143
Preface

In this thesis, speech perception (Chapter I), articulation (Chapter II), and word learning (Chapter III) are discussed as the three major aspects of language development in young children. The three chapters demonstrate how perceptual and productive language abilities are related in each of these aspects. As a culmination of this, Chapter IV outlines the rationale of the present study by discussing the relationship between the attention to phonological detail in speech and the lexicon in regards to two different models, the vocabulary-driven and phonology-driven model, and deducted hypotheses. On the basis of this, Chapter V describes the tasks that were implemented to ultimately test these models, and Chapter VI presents the results. In Chapter VII, the methods are evaluated and the results are discussed in light of the vocabulary-driven and phonology-driven model and their implications for future research.
Chapter I

Speech Perception
This chapter outlines the general theoretical basis for understanding speech perception in section 1.1., before describing theoretical background and empirical findings on the development of speech perception abilities from infancy to adulthood in the sections 1.2. and 1.3.

1.1. **SPEECH PERCEPTION FUNDAMENTALS**

Speech is a complex stimulus and perceiving speech is much more involved than it appears to be on the surface. Issues in speech perception are set out here, followed by levels of analysis and models of speech perception.

1.1.1. **The speech perception challenge**

Perceiving spoken language is a challenging task. For example, the speech signal is continuous; and yet we clearly perceive pauses between words (Cole & Jakimik, 1978). Even infants learn quickly to segment correctly at word boundaries - on average with 18 months (P.W. Juszczyk, Houston, & Goodman, 1998). This segmentation problem plus other issues in speech perception, speaker variability, and context effects are set out below.

1.1.1.1. **Speech stream segmentation**

Listening to a foreign language that one does not understand gives some measure of the difficulty of the task infants face when trying to make sense of the continuous stream of speech surrounding them. Different cues have been suspected to assist infants in mastering speech stream segmentation, such as prosodic pauses and computation of adjacent syllable probabilities: 6-month-olds prefer speech in which pauses are inserted correctly as markers for clausal units rather than pauses inserted at other locations (Hirsh-Pasek et al., 1987), and 9-month-olds prefer correct marking of smaller phrasal units such as subject-predicate in order to extract words from the speech stream (P.W. Juszczyk et al., 1992). Detection of reoccurring sound sequences is also required in order to predict which sequences of sounds form words. Nine-month-olds have acquired this sensitivity to legal sound
clusters in their native language (P.W. Jusczyk et al., 1998). Thus, there appear to be two segmentation strategies infants use for segmentation - ‘bracketing’ and ‘clustering’ (Goodsitt, Morgan, & Kuhl, 1993). Bracketing involves the use of cues provided by pauses, pitch contour discontinuity, and vowel length, while clustering uses the higher probability of certain syllable combinations within one’s language to identify speech units. Calculating probabilities (e.g., Aslin, Saffran, & Newport, 1998; Newport & Aslin, 2004) as a computational process of word acquisition is discussed in more detail in Chapter III.

1.1.1.2. Talker variability

Spoken language differs greatly across speakers. To preserve understanding, variability across talkers should not be interpreted as a difference in meaning, but merely as an acoustic deviation of the same lexical items. Differences in speaking rate are one aspect of talker variability, inter-speaker pitch contour changes in syllables and words are another. Infants as young as two months have been found to cope with these aspects of talker variability in discrimination tasks (P.W. Jusczyk et al., 1998), and 5-month-olds can identify their names in a setting with several speakers (Newman, 2005).

1.1.1.3. Context effects

Beyond the variability between talkers, it is also highly unlikely that the same individual produces the same speech sound sequence twice in exactly the same fashion. This within-speaker variability means that speech sound sequences - although containing the same sounds - are never pronounced in the same way, irrespective of order (Miller & Eimas, 1995). If speech sound sequences contain the same sounds, but in a different order, the acoustic qualities of the utterance will also differ because of the effect of coarticulation. Coarticulation refers to the fact that speech sounds are not independent of surrounding speech sounds. Finally, listening conditions are also not always ideal to perceive speech as environments can be noisy.
1.1.2. Units of speech perception

The issue of coarticulation naturally leads to consideration of the units of speech, the individual articulatory units that are being co-articulated and speech perception (see also Smits, 2001).

1.1.2.1. Phones and Phonemes

The term phone, and its corresponding adjective phonetic, represents the acoustic realisation of the smallest possible unit of the speech stream at a language-general level. Phones are represented in square brackets, for example [d], [t], and [tʰ]. Phoneme is the label used for a minimal speech segment within a particular language; that is “the shortest segment that makes a significant difference between utterances” (A. M. Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967, p. 431). Phonemes are represented with slashes, for example /d/ and /t/. Phonetic and phonemic segments do not necessarily correspond to the written letters of a word. For example, the English word ‘sit’ is composed of three phonemes, i.e. /s/, /i/, and /t/. The phonemes /t/ and /d/ are distinct in English and can form minimal word pairs such as ‘tip’ and ‘dip’. However, the phones [t] and [tʰ] are simply different phonetic realisations, allophones, of the phoneme /t/, and could even be articulatory particularities of different speakers. For example, the [tʰ] in “țăr” and the [t] in “ștăr” are both allophones of the English phoneme class /t/.

1.1.2.2. Tones and Tonemes

In tonal languages, fundamental frequency at the lexical level equals meaning. Thus, for example [ma] with a low tone and [ma] with a high tone are different words in a tone language such as Thai or Mandarin. As for phones and phonemes with consonants and vowels, tones are minimal spoken language units across different tone languages, whereas tonemes describe classes of tones between which there are semantic differences within any given tonal language (Pardo Garcia, 1993). Thus, tone is lexical in tonal languages, that is, it conveys meaning
(Cruttenden, 1986; Pike, 1948). The tonal language of concern in this thesis is Central Thai, which distinguishes five tones in two categories. Low [], mid [], and high [] tones are classified as static tones, whereas rising [] and falling [] tones belong to the dynamic class of tones (Luksaneeyanawin, 1984).

Table 1.1: Central Thai’s five lexical tones and their five-way difference in meaning, demonstrated on the phonetic string /kʰā/, a one-syllable word.

<table>
<thead>
<tr>
<th>Tone class description</th>
<th>Word example</th>
<th>Lexical meaning</th>
<th>Numeral coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mid []</td>
<td>kʰā</td>
<td>to be stuck</td>
<td>0</td>
</tr>
<tr>
<td>Low []</td>
<td>kʰā</td>
<td>galangal (a spice)</td>
<td>1</td>
</tr>
<tr>
<td>Falling []</td>
<td>kʰā</td>
<td>to kill</td>
<td>2</td>
</tr>
<tr>
<td>High []</td>
<td>kʰā</td>
<td>to trade</td>
<td>3</td>
</tr>
<tr>
<td>Rising []</td>
<td>kʰā</td>
<td>leg</td>
<td>4</td>
</tr>
</tbody>
</table>

1.1.2.3. Syllables

Syllables are fundamental units of language (Treiman, Bowey, & Bourassa, 2002), and are even considered more basic than phonemes by some (Blevins, 1995). There is neuropsychological evidence for the early sensitivity to syllabic stimuli in humans: specific left-hemispheric processing is associated with CV-syllable stimuli in newborns as recorded with auditory evoked potentials (Molfese & Molfese, 1979). This mechanism is especially sensitive to second resonance frequency transitions, and its lateralisation indicates an early specialisation in the detection of acoustic cues in the speech signal (Molfese & Molfese, 1979). Four-day-old infants, tested with the number of four to six phonemes in bisyllabic utterances, discriminate on the basis of syllabic rather than phonemic information (Bijeljac-Babic, Bertoncini, & Mehler, 1993) and even at four years, children can perform the task of word syllabification by tapping while not yet being able to segment into phonemes (I. Y. Liberman, Shankweiler, Fischer, & Carter, 1974). At six years of age, still only 70% of a sample of 49 children succeeded in phoneme segmentation, compared to 90% who master syllabic segmentation (I. Y. Liberman et al., 1974). In more advanced readers (sixth graders and adults), knowledge of the spelling was then found to affect syllabification in ambivalent words such as “mammal” and “camel” (Treiman et al., 2002).
This propensity to attend to syllables is useful in word segmentation because syllable boundaries often correspond with word boundaries, therefore language-specific syllable-initial and syllable-final rules also define word boundaries (Blevins, 1995). For example, English permits at least ten basic syllable structures - V, CV, CVC, VC, CCV, CCVC, CVCC, VCC, CCVCC, CVCCC (Blevins, 1995).

1.1.2.4. Words

Words are distinct sound patterns forming meaningful speech units (P.W. Jusczyk & Aslin, 1995); and are made up of one or more morphemes (Cole & Jakimik, 1978). A morpheme is a meaningful unit of sounds, for example, “careful” has two morphemes, ‘care’ and ‘quality’, as does “cats”: ‘feline animal’ and ‘plural’. To build a lexicon, recognition of words is the most fundamental task in early speech perception. Prosodic, phonemic, lexical, syntactic, semantic, thematic, and situational cues are employed as information sources to identify words (Cole & Jakimik, 1978). For example, 7½-month-olds are able to recognise common target words in sentential contexts (P.W. Jusczyk & Aslin, 1995), 10½-month-olds can even locate boundaries of words that do not follow the typical trochaic (strong-weak) word stress pattern in English, but start with weak syllables (P.W. Jusczyk, 1996). This ability equips them to create lexical representations of words and to develop their vocabulary. However, it is not until mid-childhood that children demonstrate an awareness for the word as a unit of spoken language: 4- to 5-year-olds predominantly still use stress to detect words, but 7-year-olds have acquired a concept of words (Tunmer, Bowey, & Grieve, 1983).

1.1.2.5. Suprasegmental information: Prosody and stress

Spoken language also carries meaning on other dimensions than words and sentences. Prosody indicates whether a sentence is a declarative statement or a question, also providing information about the emotional state of the speaker. In English, rising intonation toward the end of an utterance is applied to questions. Pitch variations are physically represented in fundamental frequency F0 changes
(see section 2.1.1.2.). Two- to 3-month-olds clearly discriminate rising or falling pitch patterns on syllables and are therefore able to process detailed prosodic information (Karzon & Nicholas, 1989).

In a stress-timed language like English, stress assists in conveying lexical meaning. Stressed words in a sentence convey speakers’ intention and attitude: ‘Do the dishes now’ as opposed to ‘Do the dishes NOW’. Stress also distinguishes word classes, for example ‘CONduct’, the noun, from the verb ‘to conDuct’. In English, there are two types of syllables in multi-syllabic words: stressed strong syllables containing full vowels, and unstressed weak syllables with reduced vowels. The most common word type in English is the two-syllable word with a stressed strong initial syllable followed by a weak second syllable and only 25% of English words start with a weak syllable (Cutler & Carter, 1987). Due to its typicality, word-initial stress serves as a cue for speech segmentation in language acquisition (Cutler, Dahan, & van Donselaar, 1997) and infants show early sensitivity to stress: 2-month-olds discriminate two-syllable utterances on the basis of stress (P.W. Jusczyk & Thompson, 1978).

1.1.3. Speech perception through a specialised speech mode?

The question of whether speech is a special form of stimulation for humans has been the focus of many scientific debates. Liberman and colleagues (A. M. Liberman et al., 1967) believe speech to be processed differently from nonspeech signals as the sheer number of phonemic segments would overload the auditory system resulting in an unanalysable, merged buzzing sound at normal speaking rates of 400 words per minute. Comparison of discrimination performance for speech and equivalent nonspeech sounds, shows that humans’ speech discrimination is far better than their nonspeech discrimination (A.M. Liberman, Harris, Kinney, & Lane, 1961). Furthermore, unlike nonspeech, speech discrimination performance peaks at phonetic category boundaries, at just those points where identification functions change, a phenomenon known as categorical speech perception (e.g., MacKain, Best, & Strange, 1981). The fact that categorical discrimination functions are not found in nonspeech analogues of speech signals is
used to argue for a special “speech mode” (Diehl, Lotto, & Holt, 2004); a specialised speech decoding system mapping speech signal segments onto internal representations, most probably at the syllable or even word level rather than at the phoneme level (A. M. Liberman et al., 1967).

Eimas, Siqueland, Jusczyk, and Vigorito (1971) found that 1-month-old infants performed better in between-category than within-category speech discrimination, and inferred that speech is innate and restricted to humans. However, subsequent research also showed categorical speech discrimination for non-human animals such as chinchillas and rhesus macaques monkeys (see Kuhl, 1993 for an overview). Kuhl argues that the tendency to partition sounds into gross categories is deeply rooted in the nature of the auditory system shared by humans, chinchillas, and macaques, and that human speech has evolved to take advantage of this (Kuhl, 1978). Further evidence for a common speech processing system in primates and humans was found in cotton-top tamarin monkeys (Ramus, Hauser, Miller, Morris, & Mehler, 2000). This contrasts with the notion that categorical speech perception is unique to humans and that speech is special for humans.

Discrimination differences between speech and nonspeech depend on the nature of the nonspeech stimuli (Mattingly, Liberman, Syrdal, & Halwes, 1971), possibly due to different means of processing equivalent information in speech and nonspeech, as Liberman and colleagues admit (1961). Contrasting 1-month-old infants’ discrimination of speech and nonspeech, Morse (1972) found no support for a specialised speech mode, as the infants discriminated both in a linguistically relevant manner according to place of articulation and intonation. In a review of studies of infants’ speech and nonspeech discrimination, Jusczyk (1981) concludes that there are no discernible differences in infants’ response to speech and nonspeech and concludes for infants there is categorical partitioning of general acoustic characteristics. It is possible that speech is special in adulthood, but that speech perception still occurs in infancy via general auditory mechanisms (Werker & Lalonde, 1988). However, more recently, infants as young as two months showed preference for speech over structurally similar nonspeech sounds,
indicating again the unique status of speech in comparison to other sounds (Vouloumanos & Werker, 2004).

1.1.4. A theoretical framework of speech perception

This issue of categorical speech perception and a special speech mode begs the question of theoretical frameworks for speech perception. Here, two accounts of speech perception are reviewed, motor theory and the perceptual assimilation model.

1.1.4.1. The Motor Theory

Proponents of the Motor Theory of Speech Perception argue that it would be nonparsimonious to assume two separate systems for encoding and decoding language. Thus the Motor Theory claims that speech perception and production are combined in the one system (A. M. Liberman et al., 1967). The objects of speech perception are the speaker’s intended phonetic gestures, cerebrally represented as motor commands to the articulators, the human speech organs. These gestural commands function as phonetic categories for speech perception as well as for speech production (A.M. Liberman & Mattingly, 1985). They are realised in the same biologically distinct module, an innate vocal tract synthesizer in which the acoustic speech signal is translated into the gestural representation via synthesis (Kerzel & Bekkering, 2000). Perceiving speech means perceiving a specific pattern of intended gestures (A.M. Liberman & Mattingly, 1985). Thus, Motor Theory defines speech perception as extracting information about intended vocal tract activity from acoustic signals (Kerzel & Bekkering, 2000). An important feature of the vocal tract activity is the fact that it does not produce phoneme level speech, as demonstrated by the fact that phoneme production changes in different speech contexts (see section 1.1.1.3). In their classic example, Liberman and colleagues (1967) demonstrated that production of the phoneme /d/ differs according to which vowel it is followed by. The acoustic patterns for /di/ and /du/ do at no point overlap in a common /d/ phoneme. The coarticulation of consecutive phonemes has implications for the gestural representations as there are as many gestural
commands as there are legal phoneme combinations in the smallest segments within any spoken language.

As the speech module is proposed to be of a biological nature and innate, the Motor Theory implies that prelinguistic infants possess a general sensitivity to the acoustic manifestations of linguistic gestures relevant to any language (A.M. Liberman & Mattingly, 1985) and so perceive phonetic distinctions on adult level. The developmental loss of perceptual ability (see section 1.2.2) is explained by degradation of that sensitivity to unused linguistic gestures. In order to recognise acoustic sounds as speech in the first place, the infant needs to make the link between perception and production by mapping the acoustic segments onto gestures; phonetic categories can only be correctly identified when acoustic patterns are linked to the underlying linguistic gestures. This, however, does not imply that the motor control of the articulators must be executed perfectly in order to perceive speech and therefore does not contradict the delay between early speech perception abilities and articulation accuracy in children.

1.1.4.2. The Perceptual Assimilation Model

The Perceptual Assimilation Model (PAM) is founded on a direct-realist viewpoint of speech perception. According to this perspective, the listener apprehends directly what is perceived. Incorporating premises of the ecological approach (Fowler, 1986; Gibson & Pick, 2000), the PAM is based on the principles of perceptual learning (Best, 1995). Mature listeners judge whether nonnative speech sounds are speech sounds, whether they fall into the native phonological space, and whether they assimilate to native phoneme categories (see Figure 1.1). If they are not classed as speech (Best, 1993), the sounds are non-assimilable (see Figure 1.2). For example, American English speaking adults assign Zulu click contrasts to the nonspeech category and perceptual abilities which such contrasts are akin to abilities with nonspeech rather than speech stimuli (Best, McRoberts, & Sithole, 1988). If two sounds in a minimal speech contrast fall into the native phonological space, but are outside native phoneme categories, an uncategorised-uncategorised
assimilation takes place. If one of the sounds in a speech contrast is assigned to a native phoneme category and the other simply into the phonological space, uncategorised-categorised assimilation occurs. According to the sound’s proximity to native gestures, nonnative speech sounds are assigned in two ways within the native phonological space (see Figure 1.3.): to a native phoneme category, either as a good, an acceptable, or a deviant exemplar, or to the general speech sound category. Otherwise they are classed as uncategorisable UNC (Best, 1995). When discriminating nonnative speech contrasts, assimilation effects occur in different patterns. Two-category discrimination (TC) describes the assignment of two speech sounds in a speech contrast to two different native phoneme categories. In this case, discrimination performance is excellent (Best, 1993). If discrimination happens on the basis category-goodness CG, two sounds are assigned to the same native category, but differ in the degree to which they represent it. If the contrasts both are assigned to the same native phoneme category and do not differ in their degree of representation - single-category SC -, discrimination is comparatively poor.

![Figure 1.1.: Native phonological space with the categories native speech, nonnative speech, and nonspeech (reproduced with permission from Best, 2006)](image1)

![Figure 1.2.: Three types of assimilation to the categories within and outside the native phonological space (reproduced with permission from Best, 2006)](image2)

The PAM entails developmental aspects of perceptual learning. Young infants discriminate contrasts not on the basis of a well-defined native phonological space, but rather perceive speech nonlinguistically (Best, 1995). Through perceptual attunement, the native phonological space becomes more defined and filled with native gestural invariants, the phonemic categories. By ten to twelve months,
infants have developed a sense for the gestural patterns of their native language, yet their articulatory gestures are still underspecified. By preschool age, the linguistic focus continues towards higher-order gestural invariants and greater language-specificity, but still lacks phoneme level organisation (Best, 1994). At some stage before adulthood, maturity in detection of phoneme categories in the native phonological space has been achieved. However, recognising refined linguistic structures and assigning them on the basis of their articulatory-gestural similarities and discrepancies come at a price: mature listeners actually obtain less information from the speech environment than young learners (Best & Strange, 1992).

1.2. Speech perception development: Models

As can be seen in the general models of speech perception considered above some indication of how speech perception develops appears necessary for a coherent and comprehensive model. Here, general principles of models of speech perception development are considered ahead of two specific models – the Robust and Fragile Model and the Native Language Magnet Model. These are followed in the next section by consideration of empirical studies of speech perception development divided into studies in infancy, preschool, and school age children.

1.2.1. A theoretical framework for speech perception development: Universal, Attunement, and Perceptual Learning Theories

Aslin and Pisoni (1980) describe the influence of pre- and postnatal experience on perceptual development in terms of three theories. Universal, Attunement, and Perceptual Learning Theory each begin with different premises about the
developmental status of the infant’s discrimination abilities at birth (see Figure 1.4.).

For *Universal Theory*, speech discrimination capabilities are fully developed at birth and extend to all possible phonetic contrasts in the world’s languages. As they already function at an optimal level, all that exposure and experience with the native language can accomplish during infancy is to maintain discrimination abilities. However, if there is a lack of exposure to specific contrasts, discrimination abilities will be selectively lost due to either neural, or attentional mechanisms or a combination of both (Pisoni, Lively, & Logan, 1994). Depending on the type of mechanism such *loss* could be irreversible (in the case of neural atrophy) or simply *attenuated* (in the case of attentional mechanisms). Therefore, according to Universal Theory, infants should perceive all possible speech contrasts in the world’s languages while adults should perceive only those that are phonemic in their language environment (Burnham, Earnshaw, & Quinn, 1987).

*Attunement Theory* assumes a less than optimally developed level of speech discrimination in newborns (Aslin & Pisoni, 1980). With increasing exposure, discrimination abilities of speech sounds of their native language are facilitated,
while performance on phonetically irrelevant contrasts either remains unrefined or declines to the extent of total loss (Pisoni et al., 1994). Specific linguistic experience enhances, realigns and/or sharpens the initially partially developed category boundaries. This theory implies that discrimination performance of infants and adults should be equal on non-phonemic contrasts, but better for adults in at least some of the phonemic contrasts (Burnham et al., 1987).

According to *Perceptual Learning Theory*, newborn infants possess at birth limited speech discrimination abilities and hence depend on early exposure and experience with specific contrasts to improve speech perception abilities (Aslin & Pisoni, 1980). The rate of improvement is in turn dependent on the frequency with which phonetic contrasts occur in the language learning environment, their acoustic distinctiveness, and the infants’ attention. Therefore, superior discrimination for a phonetically irrelevant contrast unrepresented in the ambient phonetic array as compared to a phonetically relevant contrast is impossible.

In contrast to the strong dependence on perceptual learning, *Maturational Theory* limits the influence of the surrounding language environment to a minimum (Pisoni et al., 1994). Discriminability of any contrasts is independent of early experience and unfolds following an internal maturation process (Aslin & Pisoni, 1980). This means that discrimination of all possible phonetic contrasts should initially be equal with only maturational age, reflecting the underlying internal process, affecting improvement, decline, or maintenance of discrimination performance (Pisoni et al., 1994).

These three theories plus maturational theory clearly define the possibilities and limits of perceptual development. It is likely, however, that no single theory class can account for all empirical findings. Indeed a combination of these models in an integrative approach might ultimately provide the best for a description of the development of speech perception (Pisoni et al., 1994). Two models, the Robust and Fragile Model, and the Native Language Magnet Theory, are now considered.
1.2.2. The Robust and Fragile Hypothesis

With respect to modification of speech perception abilities, Burnham (1986) defines two periods of loss and attenuation; one in the second half of the first year of life, the second approximately between four and eight years of life, especially around early school age (Burnham, 2003). The term “loss” is not meant to indicate an absolute sensory loss, as lost nonnative discrimination abilities can be retrained. It must be understood in the light of a reorganisation of perceptual biases rather than an actual loss (Werker, 1994). The point in time when perceptual sensitivity to a specific nonnative contrast is reduced is positively related to its degree of psychoacoustic robustness (Burnham, 1986). Early reorganisation corresponds to a lack of phonetic exposure to a non-robust (fragile) nonnative contrast, later reorganisation to a lack of phonological experience with a robust nonnative contrast. Fragile contrasts are psycho-acoustically weak and this lack of psychoacoustic robustness makes them both rare across the world’s languages and vulnerable to developmental reorganisation due to lack of exposure between six and twelve months. Robust contrasts are psycho-acoustically strong and easily discriminable, and so they are common across the world’s languages (Burnham, 1986). Therefore, discrimination of nonnative robust contrasts can be retrained more easily as they remain available phonetically in many of the world’s languages despite lack of phonemic relevance. The distinction between robust and fragile contrasts, and indeed between the two periods of perceptual attenuation, is more likely to occur along a continuum rather than as a distinction between categories (Burnham, 1986). Specific linguistic experience is not required for the maintenance of robust contrasts, at least in the first period of reorganisation; their perception mainly develops on the basis of attunement (realignment and sharpening) whereas perceptual development of fragile contrasts requires ongoing experience (Burnham et al., 1987; see also Stevens & Keyser, 1989, regarding the notion of perceptual salience).
1.2.3. The Native Language Magnet Theory

Vowels can be represented in a two-dimensional vowel space in terms of their first and second resonance frequencies (areas of spectral prominence) values (see sections 2.1.1.2. and 2.1.2.1.). Within this space exemplars of particular vowels cluster around particular areas. Within these areas there is a location at which exemplars are classed as best representatives of the vowel and therefore are vowel prototypes (Kuhl, 1991). The more peripheral particular exemplars are to a vowel prototype in the vowel space, the worse the goodness of representation rating for such vowel exemplars. The similarity of exemplars to prototypes has an effect on speech perception: with a prototype of a category as first referent in a speech perception task, generalisation to other exemplars occurs (Kuhl, 1991) and discrimination is poor. The prototypes are thus said to function as perceptual magnets (see Figure 1.5.). This perceptual magnet effect of speech perception occurs in both adults and six-month-old infants (Grieser & Kuhl, 1989). Perceptual magnet effects have also been shown for consonants (Kuhl, 2001). The perceptual magnet effect does not occur in nonhuman primates (Kuhl, 1991) which supports a version of the ‘speech is special’ argument.

Figure 1.5.: The perceptual magnet effect: Peripheral exemplars of a phonetic category are perceived closer to the prototype than their actual acoustic distance to the best representative of the category. The prototype acts like a magnet (reproduced with the first author’s permission from Kuhl & Iverson, 1995)
The discovery of the perceptual magnet effect gave rise to the *Native Language Magnet* (NLM) theory (Kuhl & Iverson, 1995). NLM theory proposes three phases of perceptual reorganisation in which the influence of native language prototypes are evident. In the first phase, infants segment the speech stream into categories via general auditory processing mechanisms, solving the issue of acoustic differences by assigning them to different phonetic categories in a language-general manner (Kuhl, 1993). Initially, the natural category boundaries are fairly unspecified (Kuhl, 2001), but categorical representations are constantly aligned towards the category-typical phoneme, so by six months of age infants enter the second phase and display language-specific magnet effects due to experience with the ambient language environment (Kuhl & Iverson, 1995; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992). Infants from very different language backgrounds - English, Swedish, and Japanese - show this developmental alignment (Kuhl, 2001). As a consequence of the emergence of these native language magnets, some of the natural phonetic boundaries cease to be perceived in the third phase, such that infants older than six months do not discriminate those language-general differences as well as before. According to NLM theory, the perceptual reorganisation around prototypical magnets is based on neural restructuring, creating mental speech perception maps of the acoustic distances between sounds (Kuhl, 2001), and presents a mechanism to explain the empirically-observed perceptual reorganisation during the first year of life (Kuhl & Iverson, 1995).

1.3. **Speech perception development: Research**

Research on speech perception development will now be considered in the light of the above developmental models. As this thesis concerns speech perception and language development in the toddlerhood period, research in the preceding infancy period, the preschool period, and the following school period will be of interest. These age groups correspond roughly to a four-stage model of language development which differentiates a phonetic, phonemic, semantic, and orthographic stage (Burnham, Tyler, & Horlyck, 2002). At the phonetic stage,
infants up to six months exhibit language-general speech perception abilities. In fact, infants in this age group are able to discriminate most, if not all, phonetic contrasts they have been tested on (Werker & Polka, 1993a). At the phonemic stage, infants’ nonnative speech discrimination performance starts to decline (Werker & Lalonde, 1988). When children start to develop a lexicon, they enter the semantic stage, and with the acquisition of reading and writing, the orthographic stage (Burnham et al., 2002).

1.3.1. Speech perception in infancy: Prenatal, phonetic, and phonemic periods

In infancy, there are three distinct periods of speech perception development: before birth, younger than six months, and older than six months.

1.3.1.1. Prenatal speech perception

Speech perception bootstraps language acquisition even prior to birth. By the sixth gestational month, the foetus’ cochlea has reached its full size and final shape (Rubel, 1985), is enclosed in its bony capsule, and has developed mature elongated outer hair cells (Bredberg, 1985). These developments permit the beginning of auditory perception even in the womb, as determined by studies measuring foetal responses such as heart rate changes or motor responses (Lécanuet, 1998). First, cochlear responses are shown to midfrequency ranges; sensitivity to higher and lower frequencies increases later on (Hecox & Deegan, 1985). Sounds that travel to the unborn’s ear are predominantly the mother’s voice, amplified through bone conduction, and voices and noise from the environment, muffled by maternal tissue and fluid1. In utero recordings reveal that words and differences even on phonemic level are intelligible (Lécanuet, 1998; Querleu, Renard, Versyp, Paris-Delrue, & Crépin, 1988) and foetuses discriminate speech-relevant features such as speaker gender and phonetic contrasts towards gestation termination (Slater, 1998). By birth, there has been native language input of the mother tongue for at

---

1 This speech perception environment is recreated in experimental conditions using low-pass filtered speech (e.g. P.W. Jusczyk, Cutler, & Redanz, 1993).
least three months, and as a consequence newborns display preferences for their
mother’s voice (DeCasper & Fifer, 1980), their native language (Moon, Panneton
Cooper, & Fifer, 1993), and even stories they were exposed to prenatally
(DeCasper, Lécanuet, Busnel, Granier-Deferre, & Maugeais, 1994; DeCasper &
Spence, 1986).

1.3.1.2. Phonetic speech perception from birth to six months

Since the milestone study in which 1- to 4-month-olds were found to perceive the
subtle between-category contrasts /ba/-/pa/ and /da/-/ta/, differing only in voicing
(Eimas et al., 1971), infants have been tested on numerous vowel and consonantal
contrasts, both native and nonnative to the language of the infants tested.

Firstly, with regard to vowel discrimination, Trehub (1973) demonstrated the
discrimination of the native vowel contrasts /a/-/i/ and /i/-/u/, either presented
alone or following a common consonant, in Canadian English learning 1- to 4-
month-olds. The /a/-/i/ contrast discrimination is even possible under irrelevant
pitch interference (Kuhl & Miller, 1982) and talker variability (Marean, Werner, &
Kuhl, 1992). More subtly, American English learning 2-month-olds perceive the
vowel differences of the /i/-/t/ contrast (Swoboda, Morse, & Leavitt, 1976) and 5- to
6-month-olds discriminate /a/-/ø/, both vowels synthesised (Kuhl, 1983), showing
robust category assignment even for irrelevant within-category differences. For
Swedish learning 6-month-olds, Lacerda (1992a; 1992b) found discrimination of
the native vowel contrast /a/-/ʌ/. Thus for native vowel contrasts, young infants
display a remarkably robust adult-like discrimination, with room for further
attunement (e.g., the Swedish infants could not yet discriminate a second resonance
frequency difference in the /a/-/ʊ/ contrast).

Of greater impact for the notion of developmental reorganisation is infant
discrimination of nonnative vowel contrasts. Kuhl and colleagues (1992) found
also that Swedish vowels act as perceptual magnets for Swedish but not American
infants, and that American English vowels have the same effect for American but not for Swedish infants. Six-month-old infants therefore exhibit language-specific magnet effects. English learning 4-month-olds discriminated the German contrast pairs [Y]-[U] and [y:]-[u:]. There was a perceptual asymmetry in the direction of the change in favour of [y:] and [Y]: [y:] and [Y] were rated by adults in a preliminary study to be more native-like vowels compared to [U] and [u:] and acted - when presented first - as perceptual magnets (Polka & Werker, 1994; Werker & Polka, 1993a). In a cross-language design, 6- to 8-month-old English and German infants were tested both on a German, but non-English contrast [dut]-[dyt] and a English, but non-German contrast [dɛt]-[dæt] (Polka & Bohn, 1996). There were neither discrimination nor age differences found between the groups, the only significant effects being those of contrast presentation order. This order effect shows that, if the vowel presented first has more extreme articulatory postures in the vowel space, it will operate like a perceptual magnet irrespective of whether the contrast is native or nonnative (Polka & Bohn, 1996).

We now turn to native and nonnative consonantal contrasts. Native consonantal contrasts differing in place of articulation such as /ba/-/ga/ were discriminated by English learning 2-month-olds (Morse, 1972) and 5- to 6-months-olds (Moffitt, 1971). Discrimination of fricatives is more difficult: 6-month-olds discriminate the contrasts /fa/-/θa/ (Holmberg, Morgan, & Kuhl, 1977), and 2-month-olds the contrasts /va/-/ða/ (Levitt, Jusczuk, Murray, & Carden, 1988). French newborns discriminate several native consonantal contrasts in CV syllables ranging from /ba/-/da/ to /bi/-/gi/ (Bertoncini, Bijeljac-Babic, Blumstein, & Mehler, 1987). Manner of articulation native contrasts like /ba/-/ma/ are also successfully discriminated by 2- to 4-month-olds (Eimas & Miller, 1980).

---

2 Fricatives are consonants created by the airflow through a narrow channel, formed by two articulators. In the case of the English /θ/, the lower lip touches the upper row of front teeth to create frication, i.e. the airflow. For a complete descriptions of this and similar terms, see Chapter II.
Again, cross-language consonant contrast discrimination provides support for the language-general discrimination abilities in young infants. 1- to 4-month-olds from an English speaking environment discriminate the French and Polish oral-nasal contrast [pa]-[pă] as well as the Czech contrast [řa]-[za] (Trehub, 1976); and 2-month-olds learning Kikuyu, a Kenyan Bantu language, discriminate the English contrast [ba]-[pa] despite their lack of experience with voicing differences (Streeter, 1976). Thus it can be concluded that newborn infants discriminate both vowel and consonant contrasts whether they are native or nonnative to their language.

1.3.1.3. Phonemic speech perception from six to twelve months

Infant speech discrimination is global, however, only in the first six postnatal months. Discrimination performance then undergoes a dramatic developmental change at different times for different types of contrasts (Werker & Polka, 1993a).

Vowel discrimination is the first to be affected by reorganisation based on ambient language experience. While English learning 4-month-olds discriminate the German contrast pairs [Y]-[U] and [yː]-[uː], their 6-month-old counterparts cannot (Polka & Werker, 1994). However, 6- to 8-month-olds still perform slightly better on this task than do 10- to 12-month-olds, which indicates a continuing reorganisation. In the above mentioned cross-language study with English- and German-learning infants on the vowel contrasts [dut]-[dyt] and [dɛt]-[dæt] (Polka & Bohn, 1996), there was no difference in discrimination performance between the 6- to 8-month-olds and a group of 10- to 12-month-olds. This indicates variation in the onset of the decline in nonnative speech perception as a function of the particular contrast involved. However, for most vowel contrasts that have been studied thus far, discrimination performance declines between four and eight months of age (Werker & Polka, 1993b).

For consonants, the shift to enhanced language-specific perception occurs later, around six to twelve months (Werker & Polka, 1993a). Canadian English learning
6- to 8-month-olds and 10- to 12-month-olds were compared in their ability to discriminate the Hindi contrast [ᵶ]-[ᵣᵣ] (Werker, Gilbert, Humphrey, & Tees, 1981) and the Nthlakamp’x³ contrast [ki]-[qi] (Werker & Tees, 1984a). The younger age group performed well, whereas the older infants failed to even reach criterion in a speech discrimination task. Another cross-language study used the American English contrast [r]-[l] (Kuhl, 2001) which is nonnative in Japanese: Japanese adult listeners were found to have considerable difficulty discriminating the two consonants as were 11-month-olds, whereas 7-month-olds Japanese infants perform just like their 7-month-old American English learning counterparts. In addition, the performance of the American English learning infants at eleven months improved further compared to the younger age group suggesting that attunement for this contrast in sympathetic language environments (Kuhl, 2001).

Development of tone perception also shows reorganisation: 6- and 9-month-old Chinese- and Australian English-learning infants were tested on their discrimination of speech and nonspeech tones (Thai and synthetic violin). It was found that the Australian infants showed a decline in their discrimination of tone speech from six to nine months, but not for nonspeech violin tone, whereas Chinese infants’ discrimination performance remained stable (Mattock & Burnham, 2006). This shows that perceptual reorganisation for tone is a function of the native language environment and that tone but not music is linguistically perceived.

Compared to the phonetic stage of speech development, the phonemic period between six and twelve months of age is characterised by increased attunement towards the native language: infants specialise more and more in their mother tongue. This shift towards phonemic perception (Burnham et al., 2002) is consistent with the emergence of perceptual magnets⁴ (Kuhl, 1993). By their first birthday, infants appear to have developed linguistically significant speech sound

³ Nthlakamp’x - Inglekepm’x in its other native term - or Thompson, is an Interior Salish (Native Indian) language spoken in British Columbia, Canada. The chosen CV-syllables differ in Nthlakamp’x in their place of articulation.

⁴ At least in vowel perception, replicable language-specific magnet effects have been shown (Werker, 1995).
categories (Werker & Lalonde, 1988), perhaps corresponding to the emergence of a phonological system between ten to twelve months (Best et al., 1988).

1.3.2. Speech perception at preschool age: The semantic period

Only a few studies have investigated nonnative speech contrast perception in children (Werker & Polka, 1993a). It has been found English language 4-year-olds discriminate the Hindi voicing contrast [tʰa]-[dʰa] (Werker & Tees, 1983) and the Zulu lateral fricative contrast [ʃ]-[h] (Insabella & Best, 1990) better than do adults, and at a similar level to 10- to 12-month-olds. However, one other Zulu contrast showed improved discrimination at four years compared to ten months which leaves any description of nonnative speech perception at preschool age inconclusive at this point (Werker & Polka, 1993a).

Burnham (2003) introduced the concept of language-specific speech perception (LSSP). LSSP is measured by subtracting discrimination scores for nonnative speech (NN) from native speech (N). Thus LSSP (= N - NN) shows the degree of perceptual attention paid towards native versus nonnative features of speech (Burnham, 2003). In the first six months, the LSSP discrimination index is low because young infants perform equally well on native and nonnative discrimination tasks (see 5.3.3. for index formula). From six to twelve months, as discrimination performance on many nonnative speech contrasts declines, the LSSP index increases, indicating a first shift towards the native language. This continues into the preschool period: for English-speaking preschoolers, while the categorical perception scores are above .80 for native speech contrasts and asymptote to the maximum score of 1.0 in 6- and 8-year-olds, nonnative scores hover around .40, with a distinct reduction for 6-year-olds, eventually improving again for 8-year-olds (Burnham, 2003). Earlier, categorical perception of a native and a nonnative consonant contrast continuum emerges as a function of age in infants, 2- and 6-year-olds (Burnham, Earnshaw, & Clark, 1991): perception of the native contrast improves and perception of the nonnative deteriorates between
two and six years. This also indicates specialisation in the native language in the preschool period.

While there is a paucity of studies of speech perception in the preschool period, there are a number of studies of phonological awareness. Although a conceptually different skill than speech discrimination (Nesdale, Herriman, & Tunmer, 1984), phonological awareness nevertheless relies upon speech perception ability. *Phonological awareness* (phoneme awareness in the following) is defined as awareness of the most basic language units, the phonemes (Nesdale et al., 1984). Phoneme awareness tasks differ greatly in nature, procedure, degree of difficulty, and therefore yield widely-ranging findings. Reports of high levels of phonemic segmentation ability in 4-year-olds (Fox & Routh, 1975) oppose the early, now outdated, notion of no presence of such ability in children younger than seven years (Bruce, 1964). Commonly used test methods of phoneme awareness for 4-year-old preschoolers are tasks of phoneme recognition, phoneme blending, phoneme counting, phoneme deletion, phoneme segmentation, nonword repetition, onset rime identification (for detailed task descriptions, see Van Bon & Van Leeuwe, 2003), phoneme substitution (Bialystok, Majumder, & Martin, 2003), and invented spellings (Mann, Tobin, & Wilson, 1987).

Recent research shows that phoneme awareness depends on children’s oral language skills in semantics, syntax, and morphology (Cooper, Roth, & Speece, 2002), on their home literacy environment (exposure to reading-related media and parents’ reading initiatives) (Foy & Mann, 2003), on their productive phonology skills (Webster & Plante, 1995), and on the strength of their phonological representations (Foy & Mann, 2001). As in speech perception tasks, the analysis of preschoolers’ error patterns in phoneme recognition tasks allows inferences about the organisation of the developing phonological system with its representations to be drawn (Brady, 1997). For example, voicing differences tend to cause confusions in preschoolers’ performance on phoneme identity tasks, indicating their closeness in phonological space, whereas place of articulation differences are easier (Treiman, Broderick, Tincoff, & Rodriguez, 1998).
There is evidence that links phoneme awareness skills to speech perception abilities. Speech perception predicts together with verbal memory and cognitive functioning increase in phoneme awareness scores in kindergarteners and to an even greater extent phoneme awareness abilities in first graders (McBride-Chang, Wagner, & Chang, 1997). Phoneme-specific relations were found between phoneme awareness, speech perception, and articulation in preschoolers for the target phoneme /r/ (Sénéchal, Ouellette, & Young, 2004). Also, speech perception and phoneme awareness implement similar strategies, for example the weighing of acoustic speech cues: phoneme awareness predicts the emergence of analytical weighing strategies in 5-year-olds (Mayo, Scobie, Hewlett, & Waters, 2003). Preschoolers’ weighing of two cues, fricative spectrum and vowel resonance frequency transition, was tested longitudinally on the /so/-/ʃo/ contrast in the words ‘sew’ and ‘show’. As speech perception bootstraps first language acquisition, phoneme awareness skills can affect the development of second language acquisition: poor phonological awareness skills cause slower phonological representation of new vocabulary, as evident in Chinese third graders learning English words (Hu & Schuele, 2005), but not when learning familiar native Chinese names or native referents to new visual objects. Most commonly noted, however, are the links between phoneme awareness and reading ability.

The term phonological sensitivity has been previously used to describe early phoneme awareness or a precursor of it (Bowey, 2001; Thomas & Sénéchal, 2004), but not for children younger than three years. The term phonological sensitivity is used in this thesis to describe a basic, pre-phonemic sensitivity to detail in speech.

1.3.3. Speech perception in early readers: The orthographic period

There is a wealth of studies linking phoneme awareness to reading ability. Phoneme awareness tasks, phoneme deletion, oddity, and detection predict early reading skills in 5- and 6-year-old school children (Hulme et al., 2002), training in phoneme segmentation ability improves reading ability (Vellutino & Scanlon, 1987), and more generally, a phoneme awareness training program has been found
to have both short- and long-term effects on reading and spelling skills in German first and second graders (Schneider, Küspert, Roth, & Visè, 1997). Phoneme awareness training effects are mediated by general verbal ability (Torgesen & Davis, 1996), but not by working memory span (Oakhill & Kyle, 2000). Poor phonological coding is related to reading difficulty (Vellutino & Scanlon, 1987), and to developmental dyslexia as measured via phoneme awareness skills (Swan & Goswami, 1997).

As set out earlier, there is a reduction in nonnative speech perception around six years of age (Burnham et al., 1991). Further research shows that school children with good reading ability for their age are also those children with high LSSP scores – they score relatively well on native (N) and relatively poorly on nonnative (NN) speech contrasts (Burnham, 2003). Once reading ability and phoneme segmentation skills improve, grade 1 students typically show a peak in their LSSP index. Thus, the speech perception concept of LSSP is closely linked to the onset of reading (Burnham, 2003).

1.3.4. Speech perception in toddlers?

Due to a lack of research in that area, not much is known about speech perception skills in toddlers. As noted above, Burnham and colleagues (1991) compared categorical perception of the native [ba]-[pʰa], and the nonnative [ba]-[pa] contrast continuum in infants, 2- and 6-year-olds, and adults. From infancy to two years of age, identification of both speech contrasts improves, while between two and six years a steep increase in native speech perception\(^5\) comes with a drop in nonnative speech perception down to chance level (Burnham et al., 1991). This effect indicates that beyond two years of age, perceptual reorganisation is still ongoing and worth a closer look.

There is an effect of word learning on speech perception in the semantic period (Stager & Werker, 1997; Werker, Fennell, Corcoran, & Stager, 2002; see section

\(^5\) as measured by sharpness of identification curves and categorical scores -
4.1.1.3.), showing the influence of speech perception on language development. The acquisition of orthographic representations of spoken language through literacy training is an example of a native language skill’s influence on speech perception. The question of interest here is whether other skills have a similar influence in the semantic, toddler period. The rapid acquisition of word meaning in the vocabulary spurt during the semantic stage of speech perception development (Burnham et al., 2002) is a possible candidate. The LSSP index is a suitable instrument to index the relationship between native and nonnative speech perception and lexical acquisition. This thesis pioneers the LSSP task adapted to toddlers and tests predictions concerning the relationship between speech perception and the lexicon (see Chapter IV).
Chapter II

Speech Production: Articulation
2.1. **SPEECH PRODUCTION FUNDAMENTALS**

Speech perception precedes speech production but the relationship between these skills is complex. Ahead of discussing how speech perception relates to speech production (section 2.4.), separate consideration of the nature of spoken language (section 2.1.), speech production development (section 2.2.) and articulation (section 2.3.) is necessary.

Articulation is the motor skill component of speech, and as with any skill, practice improves performance (Menn & Stoel-Gammon, 1996). In order to produce understandable speech, the child must learn to coordinate the sophisticated articulatory system and all its components. Before describing the characteristics and qualities of speech sounds, the physiology of the mature components involved in speech production shall be briefly introduced.

2.1.1. **Physiology of the articulatory system**

A schematic representation of the articulatory system is shown in Figure 2.1. Speech sounds are created through airflow from the lungs in the majority of the

![Articulatory System Diagram](image)

*Figure 2.1.* The human speech organs: Constriction of the airflow through the three cavities and combinations of different articulatory organs determine the production of vowels and consonants. The glottis divides the speech production system into the supraglottal and subglottal part as indicated by the dotted line (adapted from Colson, 1979).
world’s languages (J. Clark & Yallop, 1995), and so the respiratory tract with trachea and lungs can be said to be part of the speech production system. The larynx, or glottis, located at the upper end of the trachea, divides the speech production system into two parts, the sub- and the supraglottal vocal tract. Thus the respiratory tract including the trachea and lungs can be said to be the subglottal part of the speech production system. The components of the speech production system in the supraglottal part, the articulators, modify the airflow through the pharyngeal, oral, and nasal cavities, thereby yielding a variety of speech sounds (J. Clark & Yallop, 1995).

2.1.1.1. The articulators

The articulatory system comprises of mobile and immobile articulators which act upon airflow from the lungs to produce speech sounds (Drumright, King, & Seikel, 2000). Through the mobile articulators, vowels are produced through shape manipulation of the pharyngeal and oral cavities, while consonantal sounds are elicited by more radical constriction. Mobile articulators are the tongue, lips, the jaw, and the soft extension of the palate, the velum. Immobile articulators are the teeth, the palate and the cavities. In the following, a selection of articulators is described in more detail: the larynx, velum, tongue, and lips.

The larynx regulates the airflow to and from the lungs through the vocal cords. There are two types, the false (ventricular) vocal folds and the true vocal folds. Only the true vocal folds are used for voicing\(^6\) (Zemlin, 1998). While also assisting in swallowing, coughing, and vomiting as well as protecting the respiratory system from foreign substances entering, the main articulatory task of the larynx is to periodically interrupt the airflow to produce speech (Laver, 2002). This periodic vocal fold vibration is known as phonation and is the most important function of the larynx in the speech production process (J. Clark & Yallop, 1995; Drumright et al., 2000).

\(^6\) See section 2.1.1.2 for an explanation of voicing.
The soft palate, the velum, consists of muscular tissue that, when raised, seals off
the nasal cavity removing nasal quality from vowel production; and when lowered,
adds nasality to the produced vowel (J. Clark & Yallop, 1995). A lowered velum
can also block the oral cavity completely, so that air flows through the pharyngeal
and nasal cavity, producing nasal consonants such as [m] in the English word
“more”\(^7\).

The tongue is the articulator that predominantly modifies the shape and size of the
oral cavity. Its mobility and plasticity is an essential element in speech
production. In reference to their use in the process of articulation, the tongue
consists of three areas: the tip, the blade, and the tongue body (J. Clark & Yallop,
1995). Two groups of muscles govern tongue movement: the extrinsic muscles are
responsible for its highly versatile positioning, and the intrinsic muscles regulate
tongue shape in collaboration with the extrinsic muscles. Slow positional
adjustments performed by the extrinsic muscles are mainly involved in vowel
production, whereas fast localised movements directed by the intrinsic muscles are
implicated of consonant production (Perkell, 1969, as cited in J. Clark & Yallop,
1995).

The lips form the end of the vocal tract. Like the tongue muscles, manifold
combinations of muscular movement yield a wide variety of lip configurations (J.
Clark & Yallop, 1995). For example, for labial stop consonants such as [p] and [b],
the lips rapidly close and release, for liquids such as [w], they round, and for
fricatives such as [f], they spread allowing the lower lip to contact the upper teeth.
Vowel production also requires a range of different articulatory lip configurations
from spread to rounded (Zemlin, 1998).

2.1.1.2. Speech sound articulation and classification

The speech sound examples provided above demonstrate that the relationship
between muscular activity in articulators and articulatory output is complex and

\(^7\) To articulate stop consonants the velum must be fully raised.
can only be understood in its overall context. Analysing individual muscle activity in the vocal tract has no meaning, unless it is considered in relation to the overall articulation (Abbs, 1986). The source of speech production is phonation, the vocal fold vibration (Laver, 2002). On the basis of phonation, voiced and voiceless sounds are differentiated across the world’s languages (J. Clark & Yallop, 1995). Voiced sounds commonly occur in every language. The degree of voicing is acoustically measured by voice onset time, VOT for short. VOT denotes the interval between the release of a consonant and the start periodic vocal fold vibration in a following vowel (Ladefoged, 2005). The VOT value is positive if the phonation follows the obstruction and release of the airflow (in the case of voiceless sounds), and negative, if the phonation occurs prior to the airflow obstruction (in the case of prevoiced sounds) (Borden, Harris, & Raphael, 2003).

Voiceless sounds are created without phonation with airflow similar to normal breathing. In English, stop consonants and fricatives can be voiced or voiceless. All English vowels are voiced, as are nasal and lateral consonants unless preceded by a voiceless consonant in which case they are assimilated and devoiced⁸ (J. Clark & Yallop, 1995).

Further classification is achieved by determining the location in the vocal tract in which the constriction for a particular sound takes place. This distinction is called place of articulation. Eight different locations along the continuum of the vocal tract for consonants can be specified: labial, dental, alveolar, postalveolar, palatal, velar, uvular, pharyngeal, and glottal (J. Clark & Yallop, 1995). In combination with other articulators, in particular tongue positions, possible places of articulation for consonants can be sub-specified (see CD Places of articulation). Places of articulation distinctions occur early in speech perception development.

Vowels can be classified as back or front according to their place of articulation (Sussman, 1990), and also on an open-close continuum specified by the degree to which the mouth is opened. The acoustic properties of vowels can be described

---

⁸ A voiceless stop such as [p] in “play” devoices the following [l] that otherwise would have been voiced.
through resonance frequencies which reflect the resonances of the vocal tract, depending on its varying size and shape (see also the following section 2.1.2.1. on vowels). The first resonance frequency \( (F1) \) is determined by changes in tongue position, the second resonance frequency \( (F2) \) by front-backness (Ladefoged, 2005). Together, \( F1 \) and \( F2 \) are the main determinants for vowel quality\(^9\). The third resonance frequency \( (F3) \) does not play such an important in distinguishing vowels and can be predicted by the first two resonance frequencies\(^{10} \) (Ladefoged, 2005).

Additional to the place of articulation, the manner of articulation, the degree of constriction and the way the constriction is formed, is used to classify consonants (J. Clark & Yallop, 1995). Dynamic and stable manners of articulation can be differentiated. Stop consonants for example are dynamic as their production requires a complete obstruction of the airflow, and its subsequent release cannot be prolonged. On the other hand, fricatives for example are stable, as their constriction can be prolonged.

### 2.1.2. The system of phonology

History has seen several attempts of speech sound classification (Jensen, 2004). A major landmark on this journey toward phonetic organisation was Chomsky and Halle’s book about the sound patterns of English (1968). It gave rise to the generative approach to phonology (J. Clark & Yallop, 1995). For scientific purposes, the sounds of the world’s languages are classified by the International Phonetics Association (see CD IPA Chart). In the IPA chart, symbols represent

---

\(^9\) The acoustic vowel space of the first two resonance frequencies tends to be symmetrical and triangular across vowel systems of the world’s languages. The five Spanish vowels \( /a e i o u/ \) are a good example of how vowels evolve to a fairly even distribution, with \( /a \ a u/ \) as the most distinct vowels marking the corners and \( /e o/ \) taking the intermediate space (Ladefoged, 2005).

\(^{10}\) Front- or backness of a vowel can also be specified by the second and third resonance frequency dimension \( (F2 \) and \( F3) \) (Syrdal & Gopal, 1986). Further, in analysing the speech signal of vowels, the place of articulation distinction was found to be encoded in the transition onset of \( F2 \) and \( F3 \) of the study’s CVC syllables (Sussman, 1990), confirming a previous finding that despite deleting up to 90% of the vowel nucleus this information in the signal still prevails (Strange, Jenkins, & Johnson, 1983). Vowel height perception is in general represented by judging the distance of any adjacent spectral peaks, such as \( F0 \) and \( F1 \), \( F1 \) and \( F2 \), and \( F2 \) and \( F3 \) (Fahey, Diehl, & Traummüller, 1996).
the sounds of speech, conveying their articulatory and acoustic properties at the same time.

2.1.2.1. Vowels

As introduced above, vowels are classified according to several parameters: tongue height, tongue position, and also by lip rounding (Jensen, 2004; Ladefoged, 2005). On the basis of sub-specifications of these parameters, primary and secondary cardinal vowels are systematised.

*Australian English vowels*

The standard Australian English vowel transcription used to be based on British English pronunciation, but there are distinct differences between the two dialects (Laver, 2002). Embedded in the CVC syllable /hVd/, the vowels in the words “hid”, “head”, and “had” are raised (*i.e.* higher tongue and more fronted position) compared to British English, whereas they are more fronted in the words “hard” and “hud”. The diphthongs (see below) in “hade”, “hide”, and “hoyd” are all shifted such that “hade” in Australian English is similar to “hide” in British English, and Australian English “hide” similar to British English “hoyd” (Cox & Palethorpe, 2006). The Australian English vowel system distinguishes monophthongs (single vowel targets such as /ɪ ɛ æ ɑ θ d õ ɔ ʒ/) and diphthongs (two points in the timecourse of vowels such as /ei ai ɔι ɔu əu/) (Cox, in press; Harrington, Cox, & Evans, 1997). There is an ongoing historical shift in Australian English vowels as in other dialects of English (Cox & Palethorpe, 2001) which requires regular updated analyses of current Australian English vowels of young speakers on a national scale (Cox, 2006).

2.1.2.2. Consonants

Consonants are classified into obstruents and sonorants according to their manner of articulation (Jensen, 2004). Obstruents are produced by different degrees of obstruction to the airflow in the vocal tract, whereas sonorants require a relatively
free airflow without pressure building up and are often voiced (Ladefoged, 2005). In English, stops (plosives) such as [tʰ] in “ten”, fricatives such as [z] in “zip”, and affricatives such as [ʃ] at the beginning and end in “church” are obstruents. Sonorants in English include nasals such as [n] in “no” and approximants are laterals such as [l] in “leap” and glides such as [w] in “win” (Jensen, 2004).

**Australian English consonants**

There is variation in the articulation of some consonants in Australian English compared to their pronunciation in other English language dialects. Generally, the plosive [t] is voiceless and created by the tongue tip obstructing at the upper alveolar ridge and side teeth. In Australian English, /t/ is typically reduced (Tollfree, 2001). There are three types of consonantal reduction in plosives realised in Australian English: the plosive can become a fricative as in /tʰ/, turn into a glide or liquid through vocalisation, or lose articulation in the oral cavity through glottalisation such as in /ɾ/, /ɾ̝/, and /ɾ̝ʰ/, with the latter being the most prevalent reduction found in speakers of Australian English (Tollfree, 2001).

### 2.1.2.3. Tones

Lexical tone (see section 1.1.2.2.) is classified into seven categories (Wang, 1967). The IPA chart codes five pitch levels, such as “extra high”, “high”, “mid”, “low”, and “extra low”. These can be combined with various contours (Jensen, 2004). The tonal language used in one perceptual task in the present study is Thai and the Thai tonal system was explained in Chapter I.

### 2.1.3. Models of speech production

Models describing the process of speech production can be divided into two groups: models of language processing and production, and models of speech motor control (A. Smith, 2006)\(^\text{11}\). The theoretical difference between the two types lies within the emphasis the models place upon certain aspects of the model.

---

\(^{11}\) Vihman gives a more differentiated classification of speech production models (1996).
Levelt’s model of speech production (1989), an example of a model of language processing and production, focuses on modules such as speech comprehension system, conceptualiser, formulator, and the lexicon, paying little attention to the articulators. Models of this category propose a linguistic unit and look then for its manifestation in the acoustic or physiological speech output (A. Smith, 2006). Other examples are the psycholinguistic model of speech processing (Vance, Stackhouse, & Wells, 2005) and the connectionist framework for phonological development (Plaut & Kello, 1999). The problem that these models face is that there is no simple one-to-one mapping between the linguistic unit and speech output, a fact that becomes clear in the differences between child and adult speech production (Hewlett, Gibbon, & Cohen-McKenzie, 1998; Menn, 1980; Shibamoto & Olmsted, 1978; Straight, 1980; Vance et al., 2005) and in the phenomenon of coarticulation (Guenther, 1995; A. Smith, 2006; Sussman, Duder, Dalston, & Cacciatore, 1999). For example the acoustic output for the initial phoneme /p/ differs according to whether we intend to say the word “pat”, “pet”, “pit”, “pot”, or “put”, thus the existence of coarticulation renders ineffective the notion of phoneme-specific phonological representations as blueprints for the articulatory movements. In fact, in the case of the vowel /u/ the anterior lip rounding movement starts several segments prior to the vowel and continues to show acoustic influence several segments after (Daniloff & Moll, 1968; as cited in A. Smith, 2006). Research findings like this indicate a complex mapping structure between linguistic units and articulatory movements that cannot be described in single units alone, neither phonemes nor syllables (A. Smith, 2006).

The role of perception in a model of speech production is not clear12. Do children learn to imitate speech sound properties or do they copy articulator movements to produce speech? The relationship of perception and production is not straightforward (Menyuk, Menn, & Silber, 1986; see also section 2.4.1.); perceptual discrimination difficulties are not unambiguously linked to production errors (Eilers & Oller, 1976; Locke, 1980). Some contrasts are more difficult to

---

12 For a review of two models (those of Vihman and Jusczyk) integrating speech perception and production, see Werker (1993).
discriminate than others, but these are not necessarily the phonemes commonly mispronounced or substituted in the speech of 2-year-olds (Eilers & Oller, 1976).

On the other hand, models of speech motor control such as Barlow and Farley’s model (1989) differentiate within the articulator system many specific modules such as cortical fine motor control, general motor control, brainstem vocalisation system, anatomical motor pathways, and articulating muscles (A. Smith, 2006; Vihman, 1993). Research using this perspective compares acoustic output from humans with barn owls and bats to discover underlying constraints and form neural models (Sussman, 2002; Sussman, Fruchter, Hilbert, & Siros, 1998). It focuses on the motor execution stage of speech production and the acoustic output (Rastle, Croft, Harrington, & Coltheart, 2005).

Each model type has its own advantages. An integrated view specifying both the modules of the higher-level psycholinguistic system and the underlying articulation system is yet to be proposed.

2.1.4. Measurement of speech production accuracy

There are two distinct ways of measuring speech sounds: on the basis of the acoustic output (i.e. analysis of recorded speech), or of the physiological production pattern (i.e. analysis of articulator movements). The latter can be conducted via a variety of techniques such as x-ray films (Daniloff & Moll, 1968; as cited in A. Smith, 2006), video analysis, and optical movement tracking systems such as Optotak (Goffman & Smith, 1999; Green, Moore, & Reilly, 2003; Walsh & Smith, 2002). Acoustic output is measured using recordings of spontaneous speech, in child speech production recordings normally combined with the presentation of a standard set of toys (e.g. Tyler & Edwards, 1993) or with elicited imitation (e.g., Corrigan & Di Paul, 1982).

To check the accuracy of recorded speech productions, for example with VOT comparisons to adult speech, the productions must be transcribed according to the appropriate phonological classification system. This poses classification problems
in speech production development research because infants and young children often produce sounds that are not native in the adult system (Kent & Miolo, 1996). Analysing young infants’ utterances also involves the decision of which sounds are to be included; generally only non-cry productions are analysed (Kent & Miolo, 1996; Stark, 1986; Vihman, 1996). This is also the case if standardised measures are used, such as the Stark Assessment of Early Vocal Development Revised SAEVD-R (Nathani, Ertmer, & Stark, 2006; Stark, Bernstein, & Demorest, 1993). The approximation of child speech to adult production does not necessarily have to be undertaken on a phonemic level, it can also be in word units for infants older than one year (Ingram, 2002). Articulation accuracy in the present study was judged on diphones of the initial consonant and the following vowel as a compromise between single phoneme and whole word analysis (see section 5.3.4.).

2.2. SPEECH PRODUCTION DEVELOPMENT

The development of speech production involves refinement of phonological knowledge in two areas: perceptual knowledge about acoustic detail in sounds, and articulatory knowledge about motor plans of sound production. Perceptual feedback of children’s own productions assists in approximating adult speech (Munson, Edwards, & Beckman, 2005a). Another major factor in speech production development is physiological maturation of the vocal tract and articulator control gain which will not be considered in detail here, especially as the infant articulatory apparatus already resembles the adult system by four months (for further reading, see Dworkin, Meleca, & Stachler, 2003; Green, Moore, Higashakawa, & Steeve, 2000; Green, Moore, & Reilly, 2002; Kent & Miolo, 1996; McCune, Vihman, Roug-Hellichius, Bordenave Delery, & Gogate, 1996; Ménard, Schwartz, & Boë, 2004; Vihman, 1996). In the following sections, phonological development is described at the articulatory level of phonological features and utterance, ahead of consideration of the acquisition and production of words in Chapter III.
2.2.1. Speech production in the first year of life

Sounds produced in the first year of life are categorised into pre-canonical (zero to six months), canonical (six to ten months) and post-canonical (ten months and older) (Ejiri, 1998; Oller & Eilers, 1988; Oller, Eilers, Neal, & Cobo-Lewis, 1998).

The first sounds an infant makes are reflexive, high-pitched and vowel-like in their nature (DeVito, 1970). For the first month of life, the infant is in the phonation stage (Oller, 1980), producing reflexive and vegetative sounds (Stark, 1986). Not taking distress signals into consideration (Vihman, 1996), the infant forms sounds with an open vocal tract with little movement of tongue and jaw. Due to their speechlike phonation they are called quasi-vowels (Oller & Eilers, 1988). They turn into cooing and goooing between two and three months by adding articulated sounds in the back of the vocal tract, similar to the velar [k] and [g], to the quasi-vowels (Locke, 1993; Oller, 1980; Oller & Eilers, 1988; Stark, 1986). This is the appearance of the first marginally syllable-like structures, mostly consisting of a variation of the velar stop [g]. Syllabicity is a general indicator for speechlikeness of utterances (Kent & Miolo, 1996). Gooing sounds are only marginal syllables as both the consonant-like and the vowel-like parts are underdeveloped compared to adult articulation (Kent & Miolo, 1996; Oller & Eilers, 1988).

For the first four months of life, the sounds produced by the infant bear no apparent relationship to the linguistic environment; they are undifferentiated to a large extent, and appear simply to be the result of vocal cord exercising. Volume and pitch variations reflect the infant’s internal state (DeVito, 1970). These limitations in production come from the fact that the young infant’s vocal tract resembles more that of nonhuman primates than adult speakers (Kent & Miolo, 1996; Stark, 1986; Vihman, 1996). It is not until four months that ontogenetic remodelling of the vocal tract permits more speechlike sounds although developmental change in the articulators is not completed (Kent & Miolo, 1996). At four months, the infant reaches the expansion stage in which vocal behaviour diversifies into vowel-like sounds, growling, and bilabial trills (raspberries) that
can signify marginal babbling despite its incomplete and imprecise vocal tract obstruction (Oller, 1980).

At around six months, the infant enters the babbling stage as indicated by the reduplication of sounds (Locke, 1993). Vowel-like sounds become less frequent - although they remain the predominant feature - and consonant-like sounds become more frequent such that a 2:1 ratio (V:C) is reached\(^\text{13}\) (DeVito, 1970; Mowrer, 1980, as cited in Kent & Miolo, 1996). Sound reduplication marks the onset of syllabicity, including vocal tract closures that imitate CV-syllables, often repeated as in [mamama], [gagaga], or [dadada] (Oller & Eilers, 1988; Stark, 1986). Babbling is characterised by syllabic frames filled with vowel-like content (Davis & MacNeilage, 1995; Gildersleeve-Neumann, Davis, & MacNeilage, 2000). The frames are jaw-articulated and interact with their content: front vowels are produced together with front (alveolar) consonants and back vowels with back (velar) consonants, as determined on the basis of a corpus of 6700 utterances of weekly recordings of six normally developing infants (Davis & MacNeilage, 1995). The underlying unit of babbling is the canonical syllable, representing a similar relationship as that between syllables and adult speech (Kent & Miolo, 1996; Locke, 1993). The canonical syllable contains a consonant (frame) and a vowel (content) (Vihman, 1996). The significance of the onset of the canonical syllable lies in the infant’s ability to produce mature phonetic sequences which are the phonetic building blocks of words (Oller & Eilers, 1988).

Towards the end of the first year of life, the prior reduplicative babbling changes into multisyllabic strings which encompass different places of articulation, for example, [daba] and [baga], in the infants’ productions (Locke, 1993). This variegated form of babbling is another milestone towards adult-like speech production, but it is debated whether it co-occurs with the onset of reduplicated babbling, as occurrence frequencies of 57% for reduplicated and 43% for variegated babbling are reported for 6- to 9-month-olds (B. L. Smith, Brown-Sweeney, & Stoel-Gammon, 1989).

\(^{13}\) In the first two months, the vowel:consonant ratio is 4.5:1 (Kent & Miolo, 1996).
2.2.2. Speech production in the second year of life

Naturally, a growing proportion of utterances in the second year of life are emerging words (see Chapter III regarding lexical acquisition). There is continuity between babbling and language development from the first year of life and beyond (Locke, 1993). For example, 13-month-olds repeat reduplicated multisyllabic strings within utterances every 200 ms, which is the syllable repetition rate of adult speech (Kent & Bauer, 1985). The rhythm of babbling and early single words reflects the overall rhythmic qualities of the ambient language (Vihman, DePaolis, & Davis, 1998). Also, babbling loses its randomness and becomes more and more context-specific. For example, 14% to 40% of babbling utterances of five children re-occurred in specific contexts more frequently than predicted by chance, indicating a sound-meaning relationship between babbling and its context that is not be modelled by adult speech (Blake & Fink, 1987). Babbling also persists even after the onset of word production (Davis & MacNeilage, 1990) and passes on its characteristics to the entries of the early productive lexicon. Thus, the first 50 words children typically utter contain consonants featured in infant babbling (Locke, 1993; Stoel-Gammon, 1983). However, the co-occurrence of same place of articulation of vowels and consonants within one utterance as found in babbling cannot be demonstrated in the early words of 18-, 21-, and 24-month-olds, apart from a weak remaining interaction of back vowels and velar consonants (Tyler & Langsdale, 1996). Thus, the interaction of consonant and vowel production seems to be limited to a phase at the very onset of phonological acquisition, and also appears to be specific to the place of articulation (Tyler & Langsdale, 1996).

A typical English speaking 2-year-old has acquired the stop consonant system and its place of articulation distinctions, although a few assimilations of place of articulation such as [gog] for “dog” can still occur in spontaneous speech (Moskowitz, 1973). They produce nasal consonants correctly, but still struggle with fricatives. The voiced fricatives pose a great challenge to the child: /s/ and /θ/, however, are generally produced earlier than /θ/ (Moskowitz, 1973). Moskowitz (1973) studying two monolingual American English speaking 2-year-olds found
they had acquired 19 of the 24 consonants of the English alphabet, and were in the process of acquiring /v z ẓ/, but did not yet show evidence of /z ɬ/.

The acquisition of consonant clusters is one of the most challenging tasks for every child growing up to speak English. The English language permits a large number of consonant clusters in syllable initial, medial, and final position, even in monosyllabic words, for example “strength” (McLeod, van Doorn, & Reed, 2001b). Therefore, the acquisition of consonant clusters is one of the most pervasive processes in phonological development and protracts into middle childhood (Gierut & O’Connor, 2002; McLeod et al., 2001b). At age two, children are generally able to produce a range of consonant clusters in word-initial and word-final position\(^{11}\), but not necessarily in adult-like form. Simplifications include consonant substitution and reduction. They also produce non-English consonant clusters, mostly containing /w/, for example /fw/ (McLeod et al., 2001a). The onset of cluster production towards the end of the second year coincides with the vocabulary spurt, a period of rapid word gain in the lexicon (Ingram, 1989; see also Chapter III). Due to considerable individual variation and the extended period of acquisition, consonant clusters were not included in the articulation task of the present study.

2.2.3. Speech production in the third year of life

Children at the beginning of their third year still choose to produce consonants most frequently featured in babbling (Stoel-Gammon, 1985). With only one exception, all consonants produced in meaningful speech by more than half of Stoel-Gammon’s sample of 34 normally developing children are listed among the most frequently babbled sounds (Stoel-Gammon, 1985).

However, children’s speech production is not only influenced by babbling, but also by the phonotactic structure of the words in the ambient language. Young English

\(^{11}\) Word-final clusters are generally acquired before clusters at word initial position; two-element clusters are earlier part of the children’s articulatory inventory than three-element clusters (McLeod, van Doorn, & Reed, 2001a).
speaking children’s speech productions reflect the phonotactic probabilities of the English language lexicon (Zamuner, Gerken, & Hammond, 2004). Phonotactic probability or phonological pattern frequency is the frequency with which a sequence of phonemes occurs in words of a particular language (Munson, 2001). This frequency can be word- and syllable-position specific, as /ŋ/ in English can only occur in syllable-final position, whereas /h/ can never occur syllable-final. For example, /y/ has a very low frequency in English in word-final position, and in fact occurs there in only three words: “beige”, “rouge”, and “loge” (Zamuner et al., 2004). On the other hand, /t/ is a very common word coda\(^{15}\) phoneme and therefore has a high phonotactic probability in word-final position. Children over the age of two are influenced in their speech productions by phonotactic probabilities (see section 3.1.2.1), as shown in studies in which CVC-nonwords are used as labels for imaginary animals. When asked to name the animals, children were more inclined to say coda consonants in high phonotactic probability word environments than in low phonotactic probability nonwords (Zamuner et al., 2004).

The word-initial phonetic inventory for children between two and three years contains all stop consonants, all nasals, glides, and the liquid /l/ as well as the voiceless fricatives /ʃ s h/, while the voiceless palatal fricative /ʃ/ is still emerging (Tanner Dyson, 1988). Tanner Dyson (1988) also showed that production of the liquid /r/ shows the most obvious development from transitional to established articulation between test sessions at 24 and 39 months of her quasi-longitudinal study. All consonants that were produced by each of the ten children studied in at least two different adult lexical items were included in the inventories. It was found that the word-final phonetic inventory of 39-month-olds includes all voiceless stops with the voiced stops /g/ and /d/ being transitional, all fricatives, but still some transitional occurrences in affricatives, all nasals, and the liquid /r/ developing from transitional to established. However, the liquid /l/ was not yet present at word-final position (Tanner Dyson, 1988). In general, consonants at

\(^{15}\) A word consists of three parts: onset, nucleus, and coda. It is possible that they can correspond to one phoneme only in short words (Demuth, Culbertson, & Alter, 2006). For example, the word “tip” has /t/ as onset, the vowel as nucleus and /p/ as its coda.
word-initial position have a greater likelihood to be produced accurately than in the medial or final word position.

2.2.4. Articulation refinement in childhood

The development of articulation accuracy does not conclude with the third birthday; speech sound production continues to become further refined. Some difficult aspects of speech production such as consonant clusters are still being polished at eight and nine years (McLeod et al., 2001b). From a sample of 1756 Australian English speaking children a speech sound inventory was derived for the ages three to nine years (Kilminster & Laird, 1978). The articulation of a phoneme was considered as acquired if produced correctly by 75% of the children in an age group. The results show that the consonants /h η p m w b n d t j k g z/ are acquired by the age of three, /l ñ tʃ/ by the age of four, /dʒ s z r v/ by the age of six, and /ð θ/ by the age of eight (Kilminster & Laird, 1978). Thus, even single consonants are protracted in their articulatory refinement, not only clusters.

A different method to investigate articulation accuracy that does not involve assigning recorded speech productions to adult phonemes is the measurement of movement patterns during speech production (Goffman & Smith, 1999). The five CVC words “man”, “pan”, “ban”, “fan”, and “van” all differ in the labial place of articulation. A kinematic Optotrack measurement of the lower lip movement revealed that 4- and 7-year-olds, just like adults, possess clearly defined individual movement patterns for each segment instead of a global template which is differentiated with later maturation as previously suggested (Goffman & Smith, 1999). This does not contradict the findings that articulation continues to improve well into middle childhood; it only means that such development cannot be measured only by lip movement recording as many later articulatory refinements include voicing.

There is an effect of lexical status on articulation accuracy in childhood: a sample of 100 3- to 7-year-old children performed three speech production tasks: picture
naming, word repetition, and nonword repetition (Vance et al., 2005). Significant improvement of the children’s performance was found across the age groups, confirming the tasks’ sensitivity to developmental changes. This was especially the case in the nonword repetition task in which increasing word length correlated with higher error rate. In the 3-year-olds, the scores of both repetition tasks were better than the naming task indicating that access to the lexical representations of the pictures and execution of the corresponding motor programme presents higher demands (Vance et al., 2005). However, the word repetition scores of the 4-year-olds show an advantage over nonword repetition scores, indicating that lexical access results in greater accuracy. This again suggests that a well-established lexicon with detailed phonological representations of the entries enables faster and more accurate articulation.

2.3. ARTICULATION ACQUISITION

In contrast to the previous section which described general phonological development, this section focuses on the development of production of specific speech sounds.

2.3.1. Acquisition of vowels

Vowel-like sounds are the productions of infants younger than four months (see section 2.2.1.) and account for the majority of sounds at that age. In the young infant, they can be divided into comfort and discomfort sounds. Discomfort cries in young infants are defined by a narrow, contracted vocal tract. Comfort sounds are produced with an open, relaxed vocal tract (DeVito, 1970). The acquisition of control over vowel production is a slow process (Lieberman, 1980), as it involves mastery of the tongue as articulator which can only be achieved with prolonged production experience (Vilhman, 1996). The American English vowel productions of five 13-month-olds contained most frequently the central vowels /ʌ/ and /ə/, accounting together for 46% of the total vowel production (Kent & Bauer, 1985).
The acoustic vowel space becomes slowly more elaborate with age, as analyses of resonance frequency structure show (Kent & Miolo, 1996). The fundamental frequency $F0$ and the resonance frequencies of $F1$ and $F2$ are higher in younger children and fall as they grow (Lieberman, 1980; Stark, 1986). Instead of attempting to mimic adult resonance frequencies, children scale their vowel productions to preserve vowel space relations that match the adult vowel space relations (Lieberman, 1980). Adult-like back vowels emerge no earlier than three years (Lieberman, 1980). As the vocal tract of children is still shorter than that of adults, their vowel productions are judged to be adult-equivalent on the basis of spectral patterns rather than absolute values of resonance frequencies which cannot match because of the physiological difference on vocal tract properties (Lieberman, 1980).

2.3.2. Acquisition of consonants

The phonemes of the English language can be described by a series of contrasts: for example, /b/ differs from /p/ only in the feature of voicing; /d/ from /θ/ only in place of articulation; but /d/ differs from /p/ in voicing and place of articulation (DeVito, 1970). Yet, one-feature contrasts are not an evident feature of infant speech development. When children develop the sound system of their language, they first seem to acquire grossly defined phonemes that do not exist in adult speech. For example, they acquire a velar phoneme that includes all phonemes produced by interaction of the tongue with the velum, for example [g], [k], and [h]. With increasingly refined motor control, the production of the gross velar phoneme becomes more differentiated and includes additional features such as voicing (DeVito, 1970). This development has three phases from 18 to 30 months, as demonstrated with the VOT for the word-initial bilabial stops /b/ and /p/ (Macken & Barton, 1980). First, there is no measurable acoustic distinction in infants’ production of these voiced and voiceless stops. When children produce measurable contrasts, they still fall within the adult VOT phoneme categories, therefore listeners do not perceive them to belong to different sound classes (both stops are heard as /b/). Only in the third phase do children produce adult-like distinctions of /b/ and /p/ and the voicing contrasts /b/-/p/ and /d/-/t/ are finally acquired (Macken
& Barton, 1980; see also Velten, 1943). CV-syllables including these contrasts should therefore be discriminable for children between two and three years\(^\text{16}\).

The first discernible consonants in American English are /h/ and /r/, accounting together for 87% of consonant production in the first two months, followed by /k/ and /g/ (12%) (Irwin, 1947, as cited in Kent & Miolo, 1996). From two to four months, /h/ is predominant among all consonants produced (60%), followed by /r/ (16%) and /g/ (12%). Towards the end of the first year, the consonant production distribution changes dramatically: the frequency of /h/ is reduced to 37%; /d/ increased its frequency 45-fold to 18%; /m/ and /b/ doubled their frequency to 16%; and the glottal stop /ʔ/ now only occurs infrequently (5%). Besides /h/, the most frequently occurring consonants are together with /g/ the ones predominantly used in babbling. From nine months onwards, infants display language-specific characteristics in their consonant production (de Boysson-Bardies, Sagart, & Durand, 1984; de Boysson-Bardies et al., 1992).

### 2.3.3. Acquisition of tone

Since tone is a significant feature in the majority of the world’s languages (Yip, 1995, 2002), a summary of the development of tone production in Mandarin, the language spoken by the largest number of people in the world, shall be given here to complete the picture of articulation acquisition. Mandarin has four lexical tones for stressed syllables: tone 1 has a level fundamental frequency contour; tone 2 has a rising contour with a slight dip at 20% into the production; tone 3 has a falling and rising contour that changes at the 50% mark; and tone 4 has a falling contour (Wong, Schwartz, & Jenkins, 2005). Most studies report that by the time children produce all tones correctly, they still struggle with the production of some consonants such as affricatives, fricatives and the liquid /l/ (Clumeck, 1980). This finding that tones are acquired before some consonants is repeated across a range of studies and languages, including Cantonese, Lao, and Gã, a West African tonal

\(^{16}\) The contrasts /h/-/p/ and /d/-/t/ were used in this study as part of the LSSP task (see section 5.3.3.).
language (Clumec, 1980). Tone acquisition begins together with consonant and vowel production and is overall phonetically more accurate (Clumec, 1980). Still, mastery of the acquisition of the tonal system takes time. Differences in methodology and the reliability of judges in the studies make it difficult to pinpoint a definite acquisition age for tones. In a recent study, ten Mandarin speakers judged the quality of productions of tones (low-pass filtered so that decisions were based on fundamental frequency contours without lexical or semantic influence). They found that the tones in filtered monosyllabic words of 3-year-old Mandarin children are not yet adult-like (Wong et al., 2005). All children produced the level, rising, and falling tones with high accuracy, but not the dipping tone (Wong et al., 2005). However, the dipping tone was also the most difficult for the judges to perceive in the adult control productions (also low-pass filtered). This confirms earlier evidence of tone acquisition for Mandarin in which the level and falling tones are the first to be produced correctly, followed by the rising and finally the dipping tone (Li & Thompson, 1977).

2.4. Articulation Acquisition in Context

The development of articulation does not occur in isolation. Cognitive and social influences are reflected in speech production development (Munson, Edwards et al., 2005a; Stark, 1986), but will not be discussed here. There are, however, other areas that influence or are influenced by the development of articulation, such as speech perception (as discussed in Chapter I), phoneme awareness, and word learning (see Chapter III). The interaction between articulation and these areas is briefly discussed below.

2.4.1. Articulation and Speech Perception

In adults, articulation improves via perceptual feedback and is therefore dependent on practice, as demonstrated in adults’ improvement in the pronunciation of nonwords that are phonetically illegal in English (Schulz, Stein, & Micallef, 2001). This relationship between articulation and speech perception also has relevance developmentally. Infants at 4½ to five months respond with
speechlike productions only when presented with speech, and not nonspeech (Kuhl & Meltzoff, 1982). Their vocalisations imitate the speech sounds they perceive: when listening to /a/, the productions are /a/-like; when listening to /i/, the productions are /i/-like (Kuhl & Meltzoff, 1982). Infants older than six months use perceptual feedback to develop and improve their babbling. Effects of a lack of auditory input in infant vocalisations can be detected as early as seven to eight months (Menn & Stoel-Gammon, 1996).

Systematic error patterns in children, significant for a protracted developmental period, also appear to establish a connection between speech perception, detailed articulatory representations, and articulation (Munson, Edwards et al., 2005a). Common systematic substitution errors are [t] for /k/ and [θ] for /s/ for example, but at least some substitutions should be labelled covert contrasts. Covert contrasts are significant acoustic and articulatory differences that are perceived reliably as the same sound because adults class them into one perceptual category (Munson, Edwards et al., 2005a). The protractedness of children’s articulatory knowledge and its refinement on the basis of perceptual knowledge is supported by the fact that preschool children’s and even adolescents’ correct pronunciations still differ systematically from adult pronunciations (Munson, Edwards et al., 2005a; Walsh & Smith, 2002).

Articulation has been found to have a relationship not only with speech perception in general, but with language-specific speech perception (LSSP) (see Chapter I). Articulation ability correlates positively with native and negatively with nonnative speech perception in preschooolers (Burnham, 2003). This has an impact on the relationship with LSSP in preschoolers: those who articulate well for their age show the highest LSSP scores. As it makes sense that children who produce speech well display a greater perceptual bias towards speech sounds of their native language (Burnham, 2003), a measure of articulation was included in this study to explore the onset of this relationship in younger children.
Articulation accuracy has also been found to depend upon phoneme awareness (Thomas & Sénéchal, 1998). Using the phoneme /r/ as an index, two groups of 80 3-year-olds were contrasted: one group who could pronounce /r/ accurately in all targets, and another group who could in none. The groups differed in their awareness for the phoneme /r/ purely on the basis of their articulation accuracy, as they were matched for other variables such as cognitive abilities (Thomas & Sénéchal, 1998). Longitudinal follow-up at age eight showed that articulation quality at three years predicted phoneme sensitivity at eight years with vocabulary, letter knowledge, and phoneme sensitivity for a control phoneme all controlled (Thomas & Sénéchal, 2004). Also, perceptual phoneme training affects phoneme production in 2-year-olds, as demonstrated with fricatives (S. K. Griffiths & Johnson, 1995); with repeated exposure to fricative contrasts, production of the distinct phonemes improved. Clinically, articulation difficulties often co-occur with phonological discrimination problems which also supports the articulation-phoneme awareness link (Rescorla & Bernstein Ratner, 1996; Rescorla & Schwartz, 1990; Stoel-Gammon, 1991; Tallal & Stark, 1980).

2.4.2. Articulation and the ambient language

In a monolingual language environment, the speech input that infants perceive is language-specific. Just as there is a native language bias in speech perception (see Chapter I), so there is in speech production. Contrary to early claims of universality of babbling, the infant’s tendency to approximate adult speech in babbling has been termed “babbling drift” (Locke, 1983) and is now widely accepted and supported by empirical studies. French and Arabic infants’ vocalisations were discriminated above chance level from six months onwards by phoneticians and from eight months onwards by phonetically untrained individuals (Boysson-Bardies, Sagart, & Durand, 1984). In a similar vein, babbling from Swedish and American English language background 12- and 18-month-olds could be distinguished by native speakers of the two languages more reliably at the later age (Engstrand, Williams, & Lacerda, 1998). Language-specific effects are also apparent in acoustic analyses of early vowel productions of 10-month-old French, Hong Kong Cantonese, and Algiers Arabic infants (de Boysson-Bardies,
Halle, Sagart, & Durand, 1989). The $F1-F2$ plots\textsuperscript{17} differed significantly as a function of the different linguistic backgrounds, paralleling adult speech differences between the languages. Therefore, babbling continues and deepens the native language bias already present in the infants’ speech perception, possibly bridging the gap to the production of meaningful words.

2.4.3. Articulation and word learning

As indicated in section 2.2.4., the lexical status of items has an effect on articulation in 4-year-olds (Vance et al., 2005). However, the interaction of articulation and word learning has its origins even earlier in development. The auditory and kinaesthetic feedback loop plays an important role in the elaboration of babbling to word learning in the first and second year of life (Menn & Stoel-Gammon, 1996). It enables children to recognise words in adult speech similar to their own productions. The first words of infants who frequently produce [baba] in babbling are likely to be “ball” and “bottle”, whereas frequent productions of [dada] often result in “daddy” and “dog”. It is possible that this is due to the established feedback loop (Menn & Stoel-Gammon, 1996).

On the other hand, articulation influences phonological representations in the early lexicon. An examination of the lexical items of 1-year-olds showed that once the voiced initial stop consonants /p t k/ were acquired, their correct articulation spread immediately throughout the lexicon and was applied to all older lexical items that were previously pronounced incorrectly (Tyler & Edwards, 1993). For example, in the same recording session at which one participant realised the correct voiceless aspiration of /k/, the VOT values of 72% of his initial /k/ items of the lexicon became adult-like.

Articulation accuracy is affected by the semantic information of word productions. Adults produce words with new information more slowly and more accurately, a finding that was replicated with 2-year-olds (Goffman, Schwartz, & Marton, 1996).

\textsuperscript{17} The two resonance frequencies $F1$ and $F2$ are typically plotted together to describe a vowel space, with $F1$ on the abscissa and $F2$ on the ordinate of the plot.
The children took longer to reproduce words representing new information in simple noun-verb-noun sequences as well as within single words and also showed greater accuracy in their production than with words representing old information (Goffman et al., 1996). It seems then that a simultaneous process runs during production using perceptual and articulatory feedback to encode the new information. How these factors combine in the early production of words is considered in the next chapter.
Chapter III

Speech Production: Word Learning
In this chapter, elements and characteristics of the lexicon are defined in section 3.1., ahead of outlining a theoretical framework for the word learning process in section 3.2. In section 3.3., ways to measure word learning are discussed and followed by a description of early lexical development in section 3.4. This is put into a contextual perspective at the end of the chapter, ahead of a discussion of lexical representations in the next.

3.1. **Elements of Word Learning**

In this section, words are introduced in their role as language units and lexical representations, before the nature of the lexicon is described.

3.1.1. **The word as language unit**

In Chapter I, the word was defined as a distinct sound pattern that is a meaningful speech unit at the same time (P.W. Jusczyk & Aslin, 1995). From a perceptual perspective, acoustic word patterns must be segmented in the speech stream; a problem taken on by the statistical language learning approach (see section 3.3.1.). Only the assignment of meaning to those patterns, however, makes words true units of language. This process involves other factors such as cognitive maturation, directional attention, and categorisation, as discussed later in this chapter. In a developmental perspective, whether a word is acquired or not depends on the definition of word acquisition (Ingram, 1989). For example, the nature of the definition can stretch from the child’s purely receptive understanding of adult words, either with variance in meaning or true to the adult meaning, to any consistently used child vocalisations, or to adult-like use and pronunciation of an adult language word (McCarthy, 1954, as cited in Ingram, 1989). A generally agreed definition of a word as being acquired in a child’s production inventory does not include adult-like pronunciation, but rather context-consistent use of an identifiable utterance (Ingram, 1989).
3.1.2. Characteristics of lexical entries

When mapping the meaning of a word onto its sound pattern, children create a representation in their mental lexicon (e.g., Fisher, Church, & Chambers, 2004; Johnson-Laird, 1987; Locke, 1988), obtaining meaning information by multifarious different means (Capone & McGregor, 2005). In adults, this lexicon reflects vocabulary and properties of the ambient language closely. The properties of the developing lexicon in young children are discussed in the following.

3.1.2.1. Word frequency and phonotactic probability

Word frequency refers to the number of times a given word occurs in a language corpus - spoken or written (Field, 2004). Generally, word recognition in spoken and written language is positively correlated with word frequency: the higher the word frequency, the faster the recognition. In adults, word frequency affects picture-naming speed over and above the influence of age of acquisition of the target words (Brysbaert, 1996; Carroll & White, 1973b). In a gating task\(^\text{18}\) with 7-, 9-, and 11-year-old children, older children require less phonetic information to recognise low-frequency words in the gradually increasing display of segments of monosyllabic words (Metsala, 1997a).

Phonotactics refers to the rules that govern the legality of phonetic combinations within a particular language (Vitevich & Luce, 1999). Phonotactics occur with a relative frequency in segments and segment sequences, which determines their phonotactic probability (Vitevich, Luce, Pisoni, & Auer, 1999). Infants are sensitive to the phonotactic rules of their native language, at nine, but not at six months of age, as a comparison of Dutch and American infants showed (P.W. Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993), but when the phonotactic patterns are more distinctly different, as in English versus Norwegian, even 6-month-olds show sensitivity to phonotactic violations (P.W. Jusczyk,

\(^{18}\) Gating is a paradigm in spoken word recognition tasks in which the participant listens to segments of a word presented with increasing duration over a series of trials. The listener’s task is to identify the word (Grosjean, 1980, 1996).
Friederici et al., 1993). In nonwords, phonotactic probabilities can be expressed as wordlikeness. Nonwords with high phonotactic probability are repeated faster than nonwords with low phonotactic probability (Vitevich et al., 1999).

3.1.2.2. Neighbourhood density

Neighbourhood density refers to the number of lexical neighbours a specific word has. Lexical neighbours are defined as words that differ only in one grapheme, for example, “pat” and “cut” are lexical neighbours to “cat”, whereas “cup” is not (Vitevich & Luce, 1998). Neighbourhood density is positively correlated with phonotactic probability: high probability phonotactics occur in dense neighbourhoods. They do, however, vary with lexicality: phonotactic probability acts on a sublexical and neighbourhood density on a lexical level (Vitevich et al., 1999). In adults, words and even nonwords with sparse neighbourhoods are named more slowly and produced with more errors (Vitevich, 2002; Vitevich, Armbrüster, & Chu, 2004; Vitevich & Luce, 2005). Neighbourhood density interacts with word frequency such that word recognition in school children is facilitated for high-frequency words with few lexical neighbours, and for low-frequency words with many lexical neighbours (Metsala, 1997a; see 3.1.2.1). Nonwords that contain high phonotactic probability segments are responded to more quickly than if they contain low probability phonotactics. This may appear counter-intuitive, but can be explained on the basis of lexical decision making: the more nonwords resemble high-frequency words, the faster they are identified as nonwords and will be processed on a sublexical level, whereas lexical decision between nonwords and phonotactically similar infrequent words is more difficult (Vitevich et al., 1999).

Preschool and school children’s lexical neighbourhood for receptive vocabulary is more sparsely populated than the adult equivalent (Charles-Luce & Luce, 1990, 1995; Storkel, 2002); as children’s vocabulary is smaller, lexical entries are more distinct from each other. If, given this less dense lexical neighbourhood, it is the existence of minimal word pairs within the lexicon that drives the level of detail in the lexical representations, then children’s lexical representations should be less
specified (Fisher et al., 2004). Neighbourhoods in the productive lexicon appear to be denser, which could be an effect of phonotactic probabilities and limited articulation ability (Coady & Aslin, 2003).

Infants at 15 months are capable of detecting neighbourhood similarity among words. They show a novelty preference for a lexical neighbour prototype after they have been familiarised with a list of lexical neighbours (Hollich, Jusczyk, & Luce, 2000). Lexical neighbourhood has also an effect on word learning in 17-month-olds. Hollich, Jusczyk, and Luce (2002) found that infants exposed to stimuli embedded in a high density neighbourhood learned the target words better than if stimuli were embedded in a low density neighbourhood, which is likely to be an effect of phonotactic probability. However, when the exposure was prolonged, they found an inhibitory effect on word learning in high density neighbourhoods presumably due to competition from lexical neighbours (Hollich et al., 2002). This sensitivity to probabilistic phonotactics and individual phonetic segments increases between 2½ and 3½ years of age, as demonstrated in a nonword repetition task that controlled for articulation ability (Coady & Aslin, 2004). Moreover, the effect of lexicality increases over age. Leigh and Charles-Luce (2002) found that both 5- to 6- and 8- to 9-year-olds are affected in their token production by phonotactic probabilities, but the younger children are not as influenced by the fact whether they are asked to repeat a word or a nonword as the older children are19.

3.1.3. Receptive and productive lexicon

Vocabulary can be measured via the receptive or productive lexicon of the child. The receptive or comprehensive lexicon contains all words a child understands but may not yet be able to produce. Receptive understanding of some words commonly emerges around 9 months (Benedict, 1979). Word comprehension is acquired quickly. For example, it took 20- to 31-month-old children only five trials

19 For a further discussion of the interaction between phonotactic probability, neighbourhood density, and levels of lexicality in adults, see Luce and Large (2001), and Vitevich, Armbrüster, and Chu (2004).
to learn new word-object combinations in a two-choice design out of a set of 100 pairings of which one half of the combinations had to be inferred by ruling out a known distracter (Vincent-Smith, Bricker, & Bricker, 1974). The receptive lexicon is related to phonological memory in 5-year-olds (Bowey, 1996), more specifically to nonword repetition, reflecting both phonological processing capacity and phonological sensitivity (Bowey, 2001).

The *productive* or expressive lexicon contains all words a child can produce, even if the productions are not yet accurate. The first words are typically produced around twelve months (P. Griffiths, 1986).

The receptive vocabulary is larger and more difficult to measure than the productive lexicon (Hamilton, Plunkett, & Schafer, 2000), especially beyond the first birthday (P. Griffiths, 1986). There is an average gap of five months between the time at which the 50 word stage is reached in the receptive lexicon compared to the productive lexicon (Benedict, 1979). Maekawa and Storkel (2006) found that phonotactic probability, neighbourhood density, word frequency, and word length all predict age of first production of words. The authors suggest that individual variation on the predictors’ phonotactic probability, neighbourhood density, and word frequency for expressive vocabulary may be explained within the framework of the emergentist coalition model (see section 3.1.2.5.).

3.1.4. **Mean length of utterance**

Mean Length of Utterance (MLU) is a measure of children’s production ability in terms of utterances longer than just single words and allows the assignment of developmental stage to language acquisition (Crain & Lillo-Martin, 1999). Thus via MLU stages such as of single-word, first word combinations, and simple sentences can be differentiated (Ingram, 1989) or more simply the one-, two-, and multi-word stages (E. V. Clark, 1996). Originally, MLU was intended as a measure of grammatical development (Brown, 1973), but it is now commonly viewed as a measure of global language production ability (DeThorne, Johnson, & Loeb, 2005). Indeed, when MLU was predicted from a composite measure of tense accuracy and
expressive vocabulary size the number of words in the productive lexicon explained more than half of the variance in regression analyses, suggesting that MLU is more strongly correlated with semantic than with syntactic measures (DeThorne et al., 2005). For the above study, MLU on a morphemic rather than a word basis was used. However, morpheme-based MLU is almost perfectly correlated with the word-based MLU used in vocabulary checklists like the MacArthur CDI (Fenson et al., 1993, see sections 3.3.4. and 5.3.1.1.), and can be used just as effectively (Parker & Brorson, 2005). A variation of the MacArthur CDI including word-based MLU was used in the study reported here.

3.2. A THEORETICAL FRAMEWORK FOR WORD LEARNING

3.2.1. The process of word learning

To establish a comprehensive description of language acquisition, the various aspects of the word learning process must be considered. Errors that children make in word learning allow the deduction of what strategies children use and the principles underlying the process of first language acquisition.

3.2.1.1. Effects of input on lexical acquisition

Parental speech characteristics have a strong influence on the process of lexical acquisition. However, there are different aspects of parental language input. In a sample of 14- to 26-month-olds, individual variability in vocabulary growth rate correlated more highly with the degree of children’s exposure to parental speech than with their learning capacity (Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991). In addition, the clarity of parental pronunciation and the extent to which parents use words in an informative context were also important. In particular, the frequency with which parents use certain words is highly correlated with children’s order of word acquisition (Huttenlocher et al., 1991), especially for the first ten words (Barrett, Harris, & Chasin, 1991). Parents’ labelling frequency also has a direct effect on children’s vocabulary size (Tan & Schafer, 2005); the higher
the frequency of labelling, the greater the vocabulary size in 16- to 20-month-olds (Tan & Schafer, 2005).

This also corresponds to the finding that parental labelling strategies are one predictor for the composition of the child’s lexicon (Poulin-Dubois, Graham, & Sippola, 1995). In a longitudinal study with monthly sessions over the second year of life, parental labels given in a picture-book naming task changed from basic to subordinate after the vocabulary spurt had started, and thereafter parental labelling levels predicted the level of abstractness in children’s vocabulary (Poulin-Dubois et al., 1995). Parental speech input continues to influence children’s language beyond the first two years. When instructed to teach adjectives rather than names in a storybook reading task, parental input affected the productive vocabulary of 2- to 4-year-old children (Hall, Burns, & Pawluski, 2003). At three years, children are still highly responsive to their mother’s speech rate. When mothers were trained to implement a certain type of slower speech when interacting with their children in an experimental situation, their toddlers were found to instantly decrease their speech rate in return (Guitar & Marchinkoski, 2001).

3.2.1.2. Word acquisition strategies and errors

There are a number of mechanisms at work in the process of language acquisition, but no single mechanism accounts for all aspects of that process (Crain & Lillo-Martin, 1999). First language acquisition is relatively fast and effortless, and relatively robust to differences in environmental input. Explanatory mechanisms such as trial and error learning, corrective parental feedback, imitation learning, parental expansions on child speech, or parental language simplifications cannot fully explain language acquisition because children make far too few errors and learn far too rapidly, especially during the vocabulary spurt (Crain & Lillo-Martin, 1999).
As an example for the limited explanatory power of single aspects of the language acquisition process, four common error types are briefly described: underextension, overextension, overlap, and mismatch (Barrett, 1996). If a word is underextended in a child’s vocabulary, it is not used for the full category of objects; for example, the label “cat” is only applied to the family’s cat. If on the other hand the word “cat” is overextended, it is applied to all small, four-legged, furry animals. An overlap describes the coexistence of under- and overextension: in our example, “cat” is not only used for all cats, but also for other pets. Finally, a mismatch is an error in which the applied word use does not match the appropriate use at all (Barrett, 1996). These error types show a part of children’s language acquisition strategies. In addition, there are principles that describe multiple aspects of the complex word learning process and these are set out in the next section.

3.2.1.3. Principles of word learning

Word learning is guided by a range of constraints or biases (Cattell, 2000). Each of these adds a tessera of understanding to the mosaic of the complex process of language acquisition. The following principles were discovered in studies with children of English-speaking background.

Infants have a large body conceptual knowledge which facilitates word acquisition (Diesendruck, Hall, & Graham, 2006; Diesendruck, Markson, & Bloom, 2003), and in a complementary process word learning facilitates the acquisition of conceptual knowledge (Booth & Waxman, 2002). For example, children have a *shape bias*; they extend labels for objects predominantly on the basis of shape rather than size, colour, material, or texture (Diesendruck & Bloom, 2003). In adulthood, function is the dominant component that drives the lexical extension of a label (Graham, Williams, & Huber, 1999). Moreover, 2-year-olds extend a novel word differently depending on whether it is introduced as an artifact or an animate object; they show a pure shape bias only for artifacts while for animates they extend both on the basis of shape and texture (Booth, Waxman, & Huang, 2005). These child versus adult differences show that expectations based on conceptual knowledge
influence lexical acquisition in general and that not only perceptual information, but also conceptual knowledge drives word learning from an early age onwards (Booth et al., 2005). Children’s shape bias is linked to noun learning: in a longitudinal study, 17-month-olds showed shape bias while learning artificial nouns in a novel word generalisation task (Gershkoff-Stowe & Smith, 2004), and this accelerated their noun production outside the laboratory training setting.

English-learning children also display a *noun bias*, also known as a referential bias in their early lexicon (Nelson, 1973; Nelson, Hampson, & Kessler Shaw, 1993). More than half of the first 50 words are nouns, compared to one-fifth verbs (Goldfield, 1993). Interestingly, an analysis of the distribution of nouns and verbs in maternal speech towards their 12-month-old children has revealed a noun bias in language input (Goldfield, 1993). Noun frequency in maternal speech correlates positively with noun proportion in children’s first 50 words (Goldfield, 1993). This noun bias is even reflected in articulation: the object words of 18- to 25-month-olds’ spontaneous productions are more accurate than their action words when matched for consonant and syllable structure in the speech samples (Camarata & Leonard, 1986).

The *taxonomic bias* refers to the tendency to extend a label acquired for one member of a category to all other members believed to be part of that category (Cattell, 2000). L.B. Smith (2000) presents the impressive example of a 22-month-old who saw a tractor for the first time. It was big, green, of a particular brand, and at work in a field. Days later, the child generalised the newly acquired word correctly for another tractor that was not big, not green, had a different brand and was not in a field.

Additionally, the process of word learning is guided by three bootstrapping principles: the object kind principle, the mutual-exclusivity assumption, and the whole-object assumption. The *object kind principle* denotes the tendency to interpret a novel label as a referent to an object (E. V. Clark, 1979; Merriman, Evey-Burkey, Marazita, & Jarvis, 1996). For example, a group of 2-year-olds
exposed to six novel nouns named more objects than a group exposed to six novel verbs named actions (Childers & Tomasello, 2002). The mutual-exclusivity assumption or mutual-exclusivity bias specifies that objects usually do not share referents, resulting in the tendency not to map novel labels onto familiar objects (Markman & Wachtel, 1988; Merriman et al., 1996). The whole-object assumption captures the fact that children map novel labels onto whole objects rather than only object parts when the object is unfamiliar (Woodward, 2000). For example, 3-year-olds considered a novel label like “pewter” and “chrome” an attribute or part of an object only for familiar objects (Markman & Wachtel, 1988).

3.2.2. Modelling lexical acquisition

Any model or theory of language acquisition must be able to integrate all these observed principles of word learning in order to do justice to the complexity of the process. Ahead of the description of an integrative model, earlier models that do not include all aspects of the word learning process, such as the competition model of lexical acquisition (MacWhinney, 1987) and the competition, attention, and learned lexical descriptions model (Merriman, 1999), shall be briefly presented here. As the discussion of the following list of challenges for a model of lexical acquisition demonstrates, neither the attentional learning account (L. B. Smith & Samuelson, 2006) nor the associative learning account (L. B. Smith, 2000) can explain the complexity of the word learning process alone, whereas the emergentist coalition model addresses exactly that (Hirsh-Pasek & Golinkoff, 2000).

3.2.2.1. Challenges for a Model of Lexical Acquisition

An all-encompassing model or theory of language acquisition must answer three main challenges associated with the word learning process: timing, constraints, and social context. The first main challenge is the question of the timing of language acquisition onset (Tomasello & Akhtar, 2000). If word learning can simply be explained by associative learning (L. B. Smith, 2000), then language acquisition should start earlier, as 6-month-olds reliably retain associations across contexts (Amabile & Rovee-Collier, 1991), 3-month-olds retrieve trained learned
motor patterns in invariant contexts (Hayne, Greco-Vigorito, & Rovee-Collier, 1993), and even 2-month-olds exhibit motor memory retrieval under certain training conditions (Vander Linde, Morrongiello, & Rovee-Collier, 1985).

As we know, word learning does not emerge that early, and in fact the development of word comprehension at around nine months (Benedict, 1979) coincides with the emergence of another fundamental skill: joint attention (Tomasello & Akhtar, 2000). Indeed, joint attention enables the child to interact with a proficient language speaker - a prerequisite for language acquisition (Tomasello & Akhtar, 2000). There are high correlations between early word comprehension and production, and joint attentional engagement of child and mother (Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998), and the amount of time children spend in joint attentional interaction with their mothers within an observation session explains about 50% of the variance in their receptive and productive lexicon (Carpenter et al., 1998). Silvén (2001) found the onset of joint attention even in 3-month-olds, with 84% of the infants engaging in joint attention at six months of age. This predicted vocabulary: The more 6-month-olds engaged in joint attention, the larger was their receptive lexicon at the onset of word learning (Silvén, 2001). Attention develops from casual attention to focused attention across the ages 10, 26 and 42 months. Ruff and Capozzoli (2003) found that the degree of increase in focused attention from 26 to 42 months corresponded to vocabulary acquisition rate.

The second main challenge is that of constraints to the process of language acquisition. As noted above the whole-object assumption, the taxonomic assumption, noun-category bias, and mutual-exclusivity assumption, conceptually derived from children’s errors, work as constraints that must be taken into account for a representative model of language acquisition (Woodward, 2000). However, rather than focusing on their constraining aspect, they should, perhaps, be seen as principles that guide the word learning process (Tomasello & Akhtar, 2000).
This leads to the third main challenge, that of the role of the social context (Snow, 1999; Tomasello & Akhtar, 2000). Across cultures and across individuals the language-learning context varies considerably. In normal human social, and linguistic environments, children acquire language independent of how much they are directly spoken to, whether they are mainly addressed by adults or other children, and whether they are explicitly instructed or presented with learning conditions sensitive to their communicative needs (Tomasello & Akhtar, 2000). There are two basic requirements in regard to the social context of language acquisition: the environment must present a learning opportunity, and the child must be able to take advantage of that opportunity (Tomasello, 1992). This does not mean that social factors do not impact on how well language is acquired. Indeed low-income toddlers have lower scores in productive vocabulary size and utterance complexity, as well as a later onset age of word combinations (Arriaga, Fenson, Cronan, & Pethick, 1998); and family literacy involvement is one of the most powerful predictors of oral language development in 3-year-olds, which in turn predicts literacy (Chaney, 1992). However, these studies serve here only an exemplary purpose; the main role of social context in language acquisition for a model of word learning is to account for the overall robustness of language acquisition towards input variation across different social contexts. The social-pragmatic approach to word learning addresses this issue (Akhtar & Tomasello, 2000).

3.2.2.2. The Emergentist Coalition Model

The emergentist coalition model of lexical acquisition integrates social, attentional, and linguistic cues for language acquisition, and is therefore able to answer the three challenges posed above. For example, explanations of the verb-learning process can be predicated on the basis of event cognition, and performative learning can be based upon research of children’s communicative intent (Tomasello & Akhtar, 2000). The emergentist coalition model developed out of a preliminary model of the principles of word learning (Hirsh-Pasek, Golinkoff, Hennon, & Maguire, 2004; Hollich, Hirsh-Pasek et al., 2000). In a first tier of the
model, three main principles appear around twelve months of age: the principle of reference (words map onto objects, action, and attributes), the principle of extendibility (words label more than the original object), and the principle of object scope (words map onto objects as a whole). Out of the principles of the first tier and maturation, the second tier develops including the principle of conventionality (general agreement on labels), the principle of categorical scope (extension of labels based on category rather than perceptual similarity), and the principle of the novel name-nameless category\(^{20}\) (novel labels map onto unnamed categories). Emergence of the second tier principles coincides with the vocabulary spurt (Hollich, Hirsh-Pasek et al., 2000).

A further elaboration of the model incorporating social pragmatics and attentional mechanisms (Hirsh-Pasek et al., 2004) is based upon three tenets, maintaining the idea of tiers: firstly, children use more than one cue - a coalition of attentional, social, and linguistic cues - to acquire word meaning; secondly, these cues change their weighting during the course of development; and thirdly, not all cues are present from the start, but emerge during development (Hirsh-Pasek, Golinkoff, & Hollich, 2000). Cues for word learning such as perceptual salience of the object, temporal contiguity between labelling and object presentation, and prosody\(^{21}\), emerge early. They decrease in weight when grammatical cues, social context, and social eye gaze emerge as cues for lexical acquisition. For example, in a study with 12- to 25-month-olds, all children used both perceptual salience and social eye gaze when learning to attach a label to an object. However, in the case of competing cues, the 24- to 25-month-olds, but not the 12- to 13-month-olds, and the 19- to 20-month-olds only gradually overcame the influence of perceptual salience (Hirsh-Pasek et al., 2000). Similarly, 18- and 24-month-olds were compared on

\(^{20}\) The principle of the novel name-nameless category (or N3C principle) correlates with the onset of the vocabulary spurt. In a longitudinal study 16- to 20-month-olds who had acquired the N3C principle and displayed fast-mapping had larger vocabularies. The proportion of children who could not fast-map was followed until they mastered the N3C principle which co-occurred with their individual onset of the vocabulary spurt (Mervis & Bertrand, 1994).

\(^{21}\) The model is inconclusive in the case of the prosody cue: in 2000 (Hirsh-Pasek et al.), prosody was presented as early emerging, in 2004 (Hirsh-Pasek et al.), prosody is counted as a second phase cue. Infants are sensitive to prosody in speech at an early age (e.g. P.W. Jusczyk, Cutler et al., 1993); however, they do not employ prosody as a cue to acquire word meaning until later in development.
their use of the referential cues of adult gaze direction and target salience, operationalised as remote-controlled target activation (Moore, Angelopoulos, & Bennett, 1999). At 24 months, the children always responded to social gaze, independent of interfering target activation, whereas at 18 months, good comprehension was only found when the cues were not conflicting (Moore et al., 1999).

The emergentist coalition framework suggests that traditionally separate fields of study such as speech perception, speech segmentation, word learning and grammatical understanding must unite their methods and integrate their results in order to capture and explain the complex process of language acquisition (Hollich, 2006).

3.3. **Measuring Aspects of Word Learning**

There are different stages of the word learning process: words must be identified in the speech stream, their phonetic patterns must be stored to be recognised, meaning must be associated with those patterns, and finally, words must be produced with sufficient accuracy to be understood. This section presents empirical approaches to measure each of these stages except the last, articulation accuracy, which was discussed in Chapter II.

3.3.1. **Statistical word learning**

Statistical language learning is an approach that integrates Chomsky’s idea of innate language acquisition strategies with connectionist models of language acquisition (Aslin, Saffran, & Newport, 1999; for an overview on the connectionist approach, see Plunkett, 1996; Plunkett, Karmiloff-Smith, Bates, Elman, & Johnson, 1997). The basic underlying notion is that human language learners compute language patterns rapidly according to statistical principles (Charniak, 1993).
The key to recognising the sound patterns of words is detection of word boundaries. This is not an easy task, as only very few relevant cues are available in the continuous speech stream (Cutler, 1996). However, 3-day-old French infants discriminated the bisyllabic pattern “mati” in ambient speech from a different word boundary condition which assigned the syllable “ma” to the previous and “ti” to the following word (Christophe, Dupoux, Bertoncini, & Mehler, 1994). To do this, infants seem to use cues other than prosody: 11-month-olds, but not 9-month-olds, preferred a story in which a pause of one second length was inserted between words over pause insertion within words - a preference that disappeared in a low-pass filtered condition in which phonetic detail is not evident (Myers et al., 1996). Also, familiarity with the more common English strong-weak stress pattern was not pivotal for word boundary detection, as the 11-month-olds were just as sensitive to pause markers for word boundaries for words with a strong-weak stress pattern as for words with weak-strong stress (Myers et al., 1996).

Given that there are seldom natural pauses between words in spoken language (see section 1.1.1.1.), learning of word boundaries appears to be almost “incidental”. Indeed, adults and 6- and 7-year-old children who heard an unsegmented artificial language stream, as an implicit background task during a foreground computer task, learned six words that they were later able to discriminate from six novel words (Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). The only cues to word boundaries in this language were the transitional probabilities between syllables, the likelihood of one segment following another. For example, in the two-word sequence “prettybaby” the transitional probability of the syllable [by] following [ba] is higher than [ba] following [ty], because many other words can be combined with “pretty” (Saffran, Newport, & Aslin, 1996). Word-internally, the probability of one syllable following another will be higher: they co-occur together every time the word is used. Across word boundaries, the transitional probability of two consecutive segments will be low (Saffran et al., 1997).

The statistical computation of transitional probability used in language learning has its onset in infancy: 8-month-olds can segment the nonsense speech stream
“bidakupadotigolabubidaku” into four three-syllable components after only two minutes of exposure in which the probability for within-word transitions was set at 1.00, and for between-word transitions at 0.33 (Saffran, Aslin, & Newport, 1996). After three minutes of exposure 8-month-olds discriminated words from word parts that spanned across word boundaries (Aslin et al., 1998). Even with a corpus of 60 words, 9-month-olds quickly extracted the new words, using statistical information from syllable structure, consonant voicing position, and segmental position (Saffran & Thiessen, 2003). Even if the probabilities are not consistent across all transitions, children are able to extract regularities: for example, children can abstract despite the errors their nonnative parents make and learn language probabilities correctly (Singleton & Newport, 2004).

However, probabilistic cues are not the only sources of boundary cues infants use to define word boundaries. Boundary cues such as aspiration of /k/ in word-initial position could assist infants in discriminating words from across-boundary segments. For example infants correctly distinguish #CVC# words (e.g. “cash”) from C#VC# segments (“dark ash”) (Mattys & Jusczyk, 2001). However, segmenting vowel-initial rather than consonant-initial words from the speech stream is more difficult - infants do not master VC-word segmentation before 16 months (Mattys & Jusczyk, 2001).

3.3.2. Word recognition and further segmentation abilities

Infants of 4½ months of age begin to recognise the sound pattern of their own names (Mandel, Jusczyk, & Pisoni, 1995), and by six months, they can retain meaning as shown by their assignation of especially recognisable patterns such as “Mummy” and “Daddy” to the corresponding person (Tincoff & Jusczyk, 1999). Further evidence for an early onset of word recognition abilities comes from Jusczyk and Aslin (1995) who found that 7½-month-olds, but not 6-month-olds can detect familiar monosyllabic word patterns in fluent speech, as shown by listening preference. When familiarised with nonwords through repetition, 7½-month-olds show sensitivity to one-feature phonetic changes in the target, as they did not listen longer for passages containing the derivative. This suggests detailed
phonetic encoding during familiarisation with a potential lexical entry (P.W. Jusczyk & Aslin, 1995).

However, word recognition skills take a long time to mature to adult level, as preschool and school children still produce many errors in spoken word recognition tasks (Gerken, Murphy, & Aslin, 1995; Swingley, Pinto, & Fernald, 1999). According to Fisher and colleagues (2004), these errors do not necessarily reflect underspecification in the early lexicon, because the word recognition process entails several steps - from analysing the acoustic word pattern in the speech stream, generating a phonological representation of this pattern, comparing the pattern with existing lexical entries, identifying the best match, and retrieving knowledge about its meaning from the selected item, and every one of these steps offers room for error (P.W. Jusczyk, 1997). Similar to adults' incremental word processing, 24-month-olds also show evidence of continuous processing of acoustic-phonetic information. Their response time in a word recognition task was faster if the labels of target and distracter picture did not overlap phonetically as in “ball” and “doll” as opposed to when they did, for example in “dog” and “doll” (Swingley et al., 1999). Similarly, 18- and 21-month-old infants recognised words and part words of which only the first 300ms were presented (Fernald, Swingley, & Pinto, 2001). This again indicates a level of lexical representation with greater phonetic detail than whole-word representation. The issue of lexical representation levels is discussed further in Chapter IV.

Speech segmentation is not only important at the word level, but also at the sentence level. Clauses are perceptual units for 7- to 10-month-old infants, as shown by tests using pause-insertion either at clause boundaries or within clauses (Hirsh-Pasek et al., 1987). Infants prefer a natural pause at clause boundary condition, thus showing sensitivity to the acoustic correlates of grammatical sentence structures. This was confirmed with 9-month-olds, but not with 6-month-old infants (P.W. Jusczyk et al., 1992). Speech segmentation ability is an essential prerequisite for vocabulary acquisition in later language development, as a recent retrospective analysis of previous data shows (Newman, Ratner, Jusczyk,
Jusczyk, & Dow, 2006). In 12-month-olds, performance on speech segmentation
tasks, but not on speech discrimination and prosodic preference tasks, predicted
expressive vocabulary scores at 24 months, and well-developed segmentation
abilities in infancy predicted better language skills even in 4- to 6-year-old children
(Hulme, Muter, & Snowling, 1998; Muter, Hulme, Snowling, & Taylor, 1998;
Newman et al., 2006).

3.3.3. Age of acquisition of words

For the design of many developmental studies it is important to know at what age
children acquire certain words. To obtain age of acquisition data for words, it was
previously common practice to use adult ratings of the age at which items from a
wordlist are acquired (Carroll & White, 1973a, 1973b; K. J. Gilhooly & Logie,
1980). Age of acquisition ratings predict word retrieval from the lexicon in tasks
such as object- and picture-naming, anagram solving, word completion and lexical
decision making. However, such ratings do not provide specific age of acquisition
data for early lexical acquisition, because the scales do not sub-specify within
years, and collapse the first two years of language development. Validity studies
only compared age of acquisition ratings with the order of standardised
vocabulary test items or with vocabulary data from participants older than 5
years (K.J. Gilhooly & Gilhooly, 1980).

To overcome these problems, researchers have focused on improving the rating
procedure by asking teachers to estimate the age of acquisition for a set of words in
6- to 8-year-old school children on the basis of their experience (Brysbaert, 1996),
by asking 5-year-old children when they had learned certain words (Walley &
Metsala, 1992), by teasing out interfering effects of word frequency (Morrison &
Ellis, 1995), and by specifying effects of word concreteness, root frequency and
context availability in Italian (Colombo & Burani, 2002). The first real age of
acquisition norms were obtained for children from $2\frac{1}{2}$ to eight years of age
(Morrison, Chappell, & Ellis, 1997) and used to confirm the effect on word naming
speed (Ellis & Morrison, 1998) and the relation to word frequency (Morrison,
Hirsh, Chappell, & Ellis, 2002). However, none of the objective age of acquisition
norms includes very early data on children under two years of age (Chalard, Bonin, Méot, Boyer, & Fayol, 2003; Morrison et al., 1997).

3.3.4. Measuring lexicon size

Traditionally, studies measuring the productive lexicon have relied on three different approaches: Parent diaries, experimental observation, and vocabulary checklists. Parent diaries are best suited for lexicon sizes up to 50 words (Reznick & Goldfield, 1994), and thus only apply to early development before the vocabulary spurt (Reznick & Goldfield, 1992). While parental reports on total vocabulary size have proved to be reliable (Harris & Chasin, 1999; Reznick & Goldfield, 1994; Robinson & Mervis, 1999), obtaining more specific vocabulary data can be problematic. For example, parents tend to overestimate the proportion of nouns in their children’s lexicons (Goldfield, 2000; Pine, 1992), and most of the early diary studies investigate language acquisition in children of linguists, posing the question of controlling experimenter bias (for an overview, see Ingram, 1989).

Experimental observation sessions are used both to test speech comprehension and speech production in young children. Speech comprehension sessions predominantly employ two paradigms: a picture-pointing task in which the child is asked to recognise the target word from a choice of several pictures, and a toy-moving task, in which the child is asked to act out what was said. The toy-moving task involves a lower risk of correct responses by chance (Foster, 1990). Speech production sessions try to elicit oral responses in picture-naming tasks (e.g., Carroll & White, 1973b; Johnson & Clark, 1988; Melnick, Conture, & Ohde, 2003) or in imitation tasks which are designed to overload the child’s short term memory to avoid simple repetitions of the target structures (Corrigan & Di Paul, 1982; Ingram, Christensen, Veach, & Webster, 1980), or to simply record spontaneous speech samples (Masur & Eichorst, 2002; Salerni, Assanelli, D’Odorico, & Rossi, 2007). In summary, experimental observation sessions are not an exhaustive measure of vocabulary, as they only test a part of the lexicon.
*Vocabulary checklists* have proved to be the most efficient and most representative measure of early vocabulary development, especially beyond the vocabulary spurt. Among the checklists, the most widely used questionnaire is the MacArthur Communicative Development Inventory (CDI) developed in the USA for US American English (Fenson et al., 1993), adapted to British (Hamilton et al., 2000) and New Zealand English (Reese & Read, 2000) and translated into many of the world’s languages. It is an exhaustive, parental measure of vocabulary size for which short forms were also developed (Corkum & Dunham, 1996; Fenson, Pethick et al., 2000; Reznick & Goldsmith, 1989). There are three versions of the CDI: Words and Gestures (WG) for 8- to 16-month-old infants, Words and Sentences (WS) for 16- to 30-month-olds (Fenson et al., 1993), and CDI III, an unpublished extension for 30- to 37-month-olds. The CDIs WG and WS have been found to chart progressive language development between ten and 27 months of age appropriately (Feldman et al., 2000). Correlations of CDI scores at two and three years with measures of cognition, receptive language skills, and parent-child conversational measures have revealed concurrent and predictive validity (Feldman et al., 2005). Normal vocabulary scores on the CDI at two years predict normal language skills at age three, although there is some variability (Feldman et al., 2005).

The New Zealand English CDI version shows good predictive validity across 21 months (Reese & Read, 2000), when predicting scores on the Peabody Picture Vocabulary Test III (Dunn & Dunn, 1997) and the Expressive Vocabulary Test (Williams, 1997). Comparing CDI scores and data from a systematic diary study of one child’s expressive vocabulary showed that the CDI underestimated expressive lexicon size (Robinson & Mervis, 1999), while the Italian CDI version showed good correlation with spontaneous speech samples in the vocabularies of 30 children at the 200 and 500 word stages (Salerni et al., 2007). Although not all studies confirm the validity of the CDI questionnaires, it appears to be the best among the measures of vocabulary.
3.4. **VOCABULARY DEVELOPMENT**

An overview of lexical development and vocabulary size is given below with respect to the first, second, and third year, and in later development up to the multi-word stage.

3.4.1. **The lexicon in the first year of life**

The first words in the productive lexicon emerge around eight to 14 months of age (Benedict, 1979; Crain & Lillo-Martin, 1999; Ganger & Brent, 2004). First words are acquired at a quite slow rate of about one or two words per week (Barrett, 1996). Comprehension generally begins earlier than word production, and continues at a rate twice as fast (Benedict, 1979). The first words are highly variable in their sound shape and are often homonymous (French, 1989; Menn & Stoel-Gammon, 1996; Stoel-Gammon & Cooper, 1984). Although they become more identifiable, systematic, and stable with time, they typically include a loss of information, higher context-dependency, and are subject to systematic and unsystematic reductions compared to the adult target.

3.4.2. **The lexicon in the second year of life: The vocabulary spurts**

Between 13 and 19 months of age, the 18 children of Nelson’s sample (1973) acquired ten words in their productive lexicon, and reached the 50-word stage between 14 and 24 months. For the first 50 to 100 lexical entries, word acquisition rate is up to three words per week (Goldfield & Reznick, 1990). Typically, this is the size of the lexicon at the onset of the vocabulary spur at around 18 months of age (Benedict, 1979; Nazzi & Bertoncini, 2003; Nelson, 1973). During the vocabulary spurts, the rate of lexical acquisition can reach up to nine words per day (Goldfield & Reznick, 1990). The vocabulary spurts in the productive lexicon goes hand in hand with improved receptive knowledge of words (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998; Reznick & Goldfield, 1992), however, the spurts in receptive vocabulary seems to occur earlier than in productive vocabulary (Harris & Chasin, 1999). A rapid increase in the comprehensive lexicon from eleven
to 15 months of age, compared to 16 to 22 months was found (Reznick & Goldfield, 1992). It is suggested that the vocabulary spurt does not occur in all children in the same fashion; some children show a continuous, steady vocabulary increase, others exhibit several smaller spurts instead of one naming explosion (Goldfield & Reznick, 1990, 1996; Mervis & Bertrand, 1995). In a re-analysis of previous studies under application of a strict definition of exponential vocabulary growth, Ganger and Brent (2004) test spurt data for the presence of a sudden change in vocabulary acquisition rate, mathematically reflected in an inflection point in the curve. They conclude, however, that only a minority of children (only one in five) fulfils the conditions for a spurt, and that observed correlates of the vocabulary spurt as outlined below are rather a function of vocabulary size reaching 50 to 100 lexical entries (Ganger & Brent, 2004).

Lexical acquisition and category formation are intimately linked; a relationship that peaks at the time of the vocabulary spurt (Gopnik & Meltzoff, 1987). Even 12-month-olds are sensitive to settings in which they are provided with an appropriate label to a novel object (Fulkerson & Haaf, 2006), and object naming continues to play an important role in forming categories for novel objects in the process of early language acquisition throughout early development. For example, 17-month-olds look longer at pictures of familiar objects for which they know the referents when given no referential label in a looking preference setting (Schafer, Plunkett, & Harris, 1999).

Sensitivity to category typicality has been shown in 18- and 24-month-olds, whereas 12-month-olds only prefer a target picture over a distracter if the target displays a typical exemplar of the named category (Meints, Plunkett, & Harris, 1999). This indicates that category definitions broaden coincidentally with the vocabulary spurt.

The proportion of different word classes changes with vocabulary size (Barrett, 1996; Fenson et al., 1994; Harris & Chasin, 1999), although there are problems with assigning early lexical entries to word classes (Lieven, Pine, & Barnes, 1992;
Nelson, 1973), and parents in particular tend to overestimate the noun content of children’s vocabularies (Harris & Chasin, 1999; Pine, 1992). In a receptive lexicon size of 60 words based on data from six children studied from six to 18 months of age, the proportion of object names was 40% and action words accounted for 27% (Harris & Chasin, 1999), confirming the data of Fenson and colleagues (1994). In the production lexicon, the proportion of nouns grows up to a lexicon size of 100 words, then levels out and even decreases in vocabularies above 200 entries (Barrett, 1996). The proportion of verbs increases in a lexicon size from 50-100 to 400-500 words, then levels out, while the proportion of adjectives also starts to increase in vocabularies with 50-100 entries, but only levels out beyond a lexicon size of 500 words (Barrett, 1996).

3.4.3. The lexicon in the third year of life

A typical 2-year-old can produce about 310 words (Fenson et al., 1993). The postspurt lexicon size of 2½- to 3-year-olds increases to about 500 words (Goldfield & Reznick, 1990). Vocabulary growth is often domain-specific; E.V. Clark (1996) reports a child in his third year whose domains “vehicle” and “food” grew at a much higher rate than “animals”. From two years on, children frequently coin new nouns when they need them (e.g., “crow-birds” or “taxi-cars”) and also implement nouns as verbs, as in “to broom” for sweeping, or “to fork” for eating with a fork (E. V. Clark, 1996). Early multiword utterances, as measured with word-based MLU, are often frozen phrases (Lieven, Pine, & Baldwin, 1997), as two or more words, excluding compounds and reduplications, that invariably occur together from the first instance on which they are produced in contrast to purposefully constructed real multiword utterances (Lieven et al., 1992). Eventually, the components of frozen phrases will be used in other contexts and therefore become constructed phrases.
3.4.4. Further lexical development

The lexicon size of 4-year-old children is about 3000 words and still continues to increase rapidly (Goldfield & Reznick, 1990). The sources of word meaning acquisition are constantly refined and extended. By the age of four years, children are able to draw information about a word’s meaning from its lexical category indicated by phonology and syntax (Hall & Moore, 1997). In contrast to 3-year-olds, preschoolers can extend associations to an object of a different kind with the same property, or an object of the same kind with a different property, depending on whether they heard “this is a blue bird” or “this is a bluebird” (Hall & Moore, 1997). Six-year-old school children have a vocabulary of 14,000 words and are, each school year, exposed to about 10,000 new words of which they acquire at least 3,000 (E. V. Clark, 1996). The vocabulary of an 18-year-old is around 60,000 words (Aitchinson, 1994).

3.5. Word learning in context

To conclude this chapter on word learning, lexical development will be linked to the two previous chapters, speech perception and articulation.

It is established that early phonetic abilities are important for later language development. For example, Tsao, Liu, and Kuhl (2004) found that speech discrimination ability in 6-month-olds predicts vocabulary size as measured by the MacArthur CDI in the second year of life. The 6-month-olds were re-assessed in a longitudinal design at 13, 16, and 24 months on the dimensions of word comprehension, word production, and sentence comprehension which were all correlated with speech discrimination (Tsao et al., 2004). In turn, vocabulary predicts speech perception abilities: vocabulary size, word familiarity and lexical neighbourhood were found to explain differences in early phonological awareness (Metsala, 1999).
There is also a correlation between the lexicon and articulatory phonology. For example, late talkers between 18 and 33 months who scored below the 10th percentile on productive vocabulary as assessed by the MacArthur CDI, performed lower on measures of phonetic complexity in intelligible speech when compared to typically developing toddlers (Thal, Oroz, & McGaw, 1995). Also, lexical constraints, in particular word frequency and neighbourhood density, affected the phonological acquisition training of 3- to 7½-year-olds (Gierut, Morrisette, & Champion, 1999). A dense phonological neighbourhood structure was least facilitative to phonological acquisition, whereas word frequency was most facilitative. In a typically-developing sample, vocabulary size and phonotactic probability predicted production accuracy as well as production duration in nonword repetition in children from age three to eight, as well as in young adults (Edwards, Beckman, & Munson, 2004). Low-frequency sequences of two phonemes in nonwords were produced less accurately and slower, with a greater effect in children with smaller productive vocabulary sizes. The results confirm the emergence of phonological knowledge with increasing productive vocabulary in childhood (Edwards et al., 2004).

---

Figure 3.1.: Overview over the development of speech perception, speech production, and lexical acquisition in the first year of life. In speech perception development, arrows mark the onset for language-specific vowel and consonant perception. In articulation development, arrows mark the onset of canonical babbling and other articulation milestones. In lexical development, arrows mark infants’ sensitivity to word boundaries, word recognition, and emergence of the first words (reproduced with the author’s permission from Kuhl, 2007a).
Kuhl (2007) presents a schematic overview of perceptual and production development in relation to lexical development (Figure 3.1.). It ties the first three chapters of this thesis together in establishing a developmental timeline for the milestones of language acquisition in the first year of life. Speech perception, articulation, and vocabulary development do not occur independently of each other. As the figure illustrates, there is overlap and interaction between these components. On the basis of the established knowledge of speech perception and the two aspects of speech production, articulation and word learning, the rationale of the study in this thesis is introduced in Chapter IV.
Chapter IV

Inferring the Structure of Lexical Representations
In this chapter the threads from each of the first three chapters will be picked up and tied together to describe the rationale of the study and outline its aims and hypotheses. The key question of this study concerns the structure and development of lexical representations: how much phonological detail do they entail? Obviously lexical representations cannot be measured directly, thus their structure must be inferred from observable measures. There are two ways to approach the structure of lexical representations - from the perception or from the production side (Beckman & Edwards, 2000). In the first approach, infants’ and children’s sensitivity to phonological details in speech perception are taken as indications of their encoding process and as a reflection of the underlying structure of their lexical representations (see section 4.1.1.). In the second approach, phonological detail in infants’ and children’s utterances is analysed to infer the structure of their lexical representations. This is naturally confounded with children’s limited articulation abilities (Beckman & Edwards, 2000). This thesis is mainly concerned with the first approach.

4.1. THE RELATIONSHIP OF THE LEXICON AND PHONOLOGICAL DETAIL IN SPEECH PERCEPTION

In the following, the relationship between phonological detail in the lexicon and in speech perception tasks will be presented. Then the lexical restructuring hypothesis as a specific position on the nature of this relationship will be presented. On this basis, the main hypotheses of the present study regarding the relationship between the structure of lexical representations and the perception of phonological detail in speech will be stated.

4.1.1. The development of the lexicon-phonology relationship

4.1.1.1. Phonological specificity of early representations

As reported in Chapter III, the first words emerge around nine months of age (Benedict, 1979). Their emergence requires a representation in the lexicon not only for an approximation of the word meaning, but also an encoding of the sound
shape of the word, to be able to produce it with sufficient consistency across contexts. This consistency depends not only on the articulatory abilities of the infant, but also on the phonological details of the early lexical representations. In order to recognise a word, a child must store the phonetic representation, the representation of the referent, and the association between both (Storkel, 2004a). The youngest age at which a preference for familiar words without training was found is at 10½ months (Hallé & de Boysson-Bardies, 1994)\(^2\). The French infants in the study recognised familiar high-frequency words and showed their preference by looking longer at familiar as opposed to infrequent words in a head-turn preference task. At twelve months, this effect was consistent in all infants tested (Hallé & de Boysson-Bardies, 1994).

To investigate the structure of the word representations that allowed 10½-month-olds to recognise familiar words from their ambient language, 11-month-olds were tested on their sensitivity to small phonetic changes in familiar words (Hallé & de Boysson-Bardies, 1996). In a series of word recognition tasks in a head-turn preference procedure, infants were exposed to familiar words, familiar words altered in one phonetic feature, and unfamiliar words. Both the original and altered familiar words were preferred over unfamiliar words, suggesting that altered words were still recognised despite the alteration. The infants showed tolerance to voicing variation of word-initial consonants and word-initial and word-medial variation of manner of articulation but not initial consonant suppression, a phonotactic violation too gross to continue to recognise the words (Hallé & de Boysson-Bardies, 1996). This underspecification of early lexical entries corresponds to early word productions: only the global word shape is preserved while phonetic detail is highly variable (Ferguson & Farwell, 1975).

The reason for this tolerance seems to be related to the task of word learning, since infants can perceptually discriminate every contrast on which they have been tested (see Chapter I). Indeed, in a word learning task which requires pairing

\(^2\) Word recognition after familiarisation at 7½ months was reported in section 3.3.2. (P.W. Jusczyk & Aslin, 1995).
words with objects, 14-month-olds neglect the same phonetic detail they use in speech discrimination tasks so perfectly (Stager & Werker, 1997). In word-object pairings 14-month-olds, but not 8- to 12-month-old infants, quickly acquire object names like “lil” and “neem” and respond to mismatches of objects and object names (Werker, Lloyd, Cohen, Casasola, & Stager, 1998). However, with more similar object names, “bih” and “dih”, switches in word-object pairings go unnoticed in 14-month-olds (Stager & Werker, 1997). This cannot be explained by a simple loss of their early speech discrimination abilities at 14 months, because the infants still discriminate “bih” and “dih” in a discrimination task setting. Therefore, only when children attempt to acquire word meaning, do they fail to attend to information on the phonetic level (Stager & Werker, 1997). This suggests that the lexical representations of the early words are underspecified in their phonetic structure, due to the requirement that infants must store the combined information of word meaning and phonetic structure in the lexicon. A model of this process is presented in the following section.

4.1.1.2. Interlude: The Word Recognition and Phonetic Structure Acquisition Model (WRAPSA)

The Word Recognition and Phonetic Structure Acquisition (WRAPSA) model gives a plausible account for how word recognition and the lexicon evolve on the basis of early speech perception abilities (P.W. Jusczyk, 1993, 1997). Key constructs of the model entail auditory analysers, a weighting scheme, and the matching of the acoustic information to stored instances of words rather than abstract word prototypes (P.W. Jusczyk, 1994a). The auditory analysers extract units from the speech signal on the basis of acoustic signal properties. The extraction of fine-grained acoustic information patterns determines the sophisticated speech discrimination abilities in infants. These patterns are language-general in young infants, but become gradually more and more language-specific (P.W. Jusczyk, 1993, 1994b). This specialisation process signifies the implementation of weighting schemes derived from native language input. These weighting schemes develop from distributional probabilities to transitional
probabilities (Kuhl, 2007). They regulate the acquisition of meaningful sound distinctions in the native language and are mapped onto stored lexical items. Their initial prototypical forms have global character, but become more like words with experience (P.W. Jusczyk, 1997). Infants store information about the acoustic structure of words, refine features that distinguish one lexical item from another, and match the encoded sound structure with lexical information. This process of matching sound and meaning is indicated by the decrease in nonnative speech discrimination abilities over age, signifying the emergence of active weighting schemes, and continuously refines the phonetic structure of lexical entries throughout development (P.W. Jusczyk, 1994b). According to Jusczyk, this process is driven by infants’ attention to how sound patterns are distributed in the native language input, rather than as a result of the need to distinguish the meaning of different lexical entries from another (P.W. Jusczyk, 1994a).

4.1.1.3. Lexical representations in the second year

The interference of word learning in contrast discrimination shown by 14-month-olds (Stager & Werker, 1997; Werker et al., 1998) is remedied in slightly older children. Werker, Fennell, Corcoran, and Stager (2002) found that some of a sample of 17-month-olds and all 20-month-olds were able to discriminate between “bih” and “dih” in the word learning task. This is backed up by neuropsychological evidence of event-related potentials (ERP): 14-month-olds show the same ERP pattern when listening to real words or phonetically similar nonwords such as “bear” and “gare”, but more experienced word learners, 20-month-olds, show different lateral distribution for words and similar nonwords (Mills et al., 2004). At both ages the children were able to distinguish between dissimilar nonwords such as “gare” and “kobe”. As reflected in the refined organisation of neural systems, the representation of sound-meaning mappings in the lexicon undergoes changes between 14 and 20 months.

Further evidence regarding the level of detail of lexical representations comes from word recognition tasks involving correct and mispronounced items. Children’s
orienting towards target pictures was videoed when they listened to familiar words such as “baby” and “apple” or their corresponding one-feature difference mispronunciations “vaby” and “opples” (Swingley & Aslin, 2000). Both types of object labels were recognised by 18- to 23-month-olds, although performance was poorer for the mispronounced labels. Similarly, 18- to 24-month-olds could distinguish words and their one- or two-feature mispronunciations in a preferential looking paradigm both for words learned at a very early age and words acquired only recently, showing that the detail of lexical representations did not depend on how long they had been in the lexicon (Bailey & Plunkett, 2002). Also within the same age range, Dutch infants at 19 months show sensitivity to mispronunciations of familiar words either in one word-initial or word-medial features (Swingley, 2003). Clearly, 19-month-olds encode specific detail in their lexical representations. Evidence for phonetic detail in lexical representations also comes from 24-month-olds who were found to process the items of a word recognition task continuously, showing response delays for labels with phonetic overlap (Swingley et al., 1999, see also section 3.3.2.). The children took longer to look at the target picture of a label when the competing distracter’s label had a phonetically similar onset, for example “dog”-“doll” as opposed to “ball”-“doll”. The 24-month-olds’ responses, indicating continuous processing of phonetic detail in words, were very similar to adults’ responses (Swingley et al., 1999). Continuous processing on the basis of fine-grained phonetic details was also shown in 18- and 21-month-olds presented with words and part words which consisted only of the first 300ms of the word (Fernald et al., 2001). The children recognised the part words as quickly and as reliably as they did the words, although this effect was found to be vocabulary-dependent: children with productive vocabularies of more then 100 words were more accurate in their responses than children with a vocabulary of under 60 words (Fernald et al., 2001). This shows that vocabulary development affects speed and accuracy of word recognition processes.

In this section, the onset of phonetically-based lexical representations has been identified. In children aged 14 months or younger, fine phonetic detail in words is not encoded; their lexical representations appear to have a global nature.
However, some 17-month-olds and children above 18 months can attend to phonetic detail even while acquiring words. It can therefore be said that lexical entries become highly specific between 14 and 17 months of age.

4.1.1.4. Protracted development in childhood

Although the onset of the process of specification of lexical representations occurs between 14 and 18 months, this does not mean that the process is completed in 18-month-olds. In fact, the relationship between attention to phonological detail in speech23 and the lexicon is a protracted, ongoing process, as several studies indicate. For example, an early study of phonemic discrimination abilities of 2-year-olds in a word recognition task found that discrimination of minimal pairs was affected by word familiarity (Barton, 1976). Only common words were used in this study - if a word could not be identified, the experimenter taught the word-picture pairing until it was acquired consistently. The minimal word pairs were all monosyllabic with a one-feature difference either at word-onset or word-coda position, for example “bear” and “pear” or “cat” and “cap” (Barton, 1976). All 25-to 31-month-old children were able to identify the target picture as opposed to the similar sounding distracter picture, with better accuracy in the older children. Despite the lack of more sensitive measures, for example reaction times, this early study indicates a relationship between phonological-based word recognition abilities and the lexicon (Barton, 1976).

A study of the encoding of 3- and 4-year-olds’ lexical representations presented an invariant target word such as “little” with a set of nonwords, that matched the target to different degrees, differing in the first segment (“ittle”), the second (“ettle”), or the third segment (“iggle”), to test which of two strategies, segment-matching or feature-overlap, children used (Gerken et al., 1995). According to the segment-matching approach, the word-nonword comparison would be segment-

23 The shift from using the term “phonetic” to “phonological” to describe the fine-grained details of spoken words is deliberate here, as now the development of lexical representations in and beyond the third year of life will be discussed. Up to the end of the second year of life, the terminology in the above cited studies is inconsistent, as they blend both terms.
based and therefore yield faster responses for differences in earlier segments. The feature-overlap strategy involves holistic encoding which would cause faster responses for differences in more features (e.g., “little” and “liggle” differ in place of articulation and voicing). The data offer some support for still developing segment-based lexical representations: 3- and 4-year-olds, unlike adults, confused words and nonwords not on the basis of segment-matching, but feature-overlap (Gerken et al., 1995). There was, however, also an indication of segment-based lexicon access, which still allows the conclusion that the process of word recognition at three and four years is not yet adult-like. Similarly, 5- and 8-year-olds showed differences in attending to word-initial as opposed to non-initial phonetic information in spoken words and their mispronunciations (Walley & Metsala, 1990). This indicates sequential access to lexical entries in spoken word recognition. It can be inferred that the lexical representations are still undergoing further specification in childhood.

What is the nature of this further specification? A study by Storkel (2002) on the lexical neighbourhood structure in the lexicon of 3- to 5-year-olds revealed that density in the developing lexicon is defined by manner class similarity rather than phoneme similarity and so differs in structure from the fully developed phoneme-based lexicon. This suggests a structural refinement towards more segment-based representation during childhood (Storkel, 2002). Further evidence for protracted specifications of lexical entries comes from a study on the relationship between vocabulary and phonological awareness in 3- to 5-year-old children (Metsala, 1999). The children’s performance in a set of phoneme awareness tasks was predicted by the size of their receptive vocabulary, and vocabulary-related factors such as word familiarity and neighbourhood density. This draws attention to the relationship between phonological detail in speech and speech production, and the absolute lexicon size (Metsala, 1999). Lexicon size and developmental age seem to play a role in the structure of lexical representations, and since this relationship has been shown to be ongoing, it can be called lexical restructuring.
4.1.2. Lexical restructuring and the vocabulary

Lexical restructuring describes the segment-based reorganisation process of lexical entries and is summed up in the Lexical Restructuring Model (LRM) (Metsala, 1997b; Metsala & Walley, 1998; Walley, 1993; Walley, Metsala, & Garlock, 2003). With growing vocabulary, holistic storage of lexical entries becomes increasingly inefficient because of overlapping acoustic properties, especially in dense neighbourhoods (Walley, 1993). The claims of the LRM are that, firstly, word representations in the lexicon are encoded holistically and become gradually more fine-grained; secondly, lexical restructuring depends on vocabulary growth and is therefore protracted into middle childhood; and thirdly, lexical restructuring is a precursor to phoneme awareness (Walley et al., 2003). According to the model, the emergence of lexical restructuring is initiated by the vocabulary spurt, but continues to refine the lexicon throughout childhood, augmenting basic language processing abilities that influence later language development such as the acquisition of reading and writing via phoneme awareness (Metsala, 1997b; Walley, 1993). Evidence for the LRM stems almost exclusively from extensive literature reviews interpreting studies in the light of its claims (Metsala & Walley, 1998; Walley, 1993; Walley et al., 2003).

A direct test of the lexical restructuring model is difficult to find. Empirical evidence for the LRM has been inferred from a range of studies comparing age - but not vocabulary size24 - to the level of sensitivity to phonological detail in spoken word recognition tasks and to the amount of phonological information required for recognition in gating tasks (see section 3.1.2.1.). Lexical restructuring said to be gradual and word-specific, and therefore occurs differentially across the developing lexicon (Metsala & Walley, 1998). Age of acquisition, word frequency, and neighbourhood density are factors which define the differential nature of lexical restructuring (Garlock, Walley, & Metsala, 2001), also vocabulary size, vocabulary growth rate, word familiarity, and sound-similarity between individual

---

24 Interestingly Metsala and Walley (1998, p. 101) list vocabulary size as the first item on a list of factors which they suggest determine lexical restructuring.
lexicon entries (Metsala & Walley, 1998). Studies in the realm of the LRM often take on a quite narrow focus and cannot yield conclusive answers about the motor driving the segment-based lexical reorganisation.

For example, Metsala (1997a) used the gating paradigm with words varying in neighbourhood density and word frequency in 7-, 9-, 11-year-olds and adults. For high frequency words from dense neighbourhoods, performance was similar in all age groups. A strong need for early segmental restructuring can be assumed for these words, explaining the similarity in word recognition across the age groups (Metsala, 1997a). However, for low frequency words neighbourhood density effects and age differences appeared: all participants performed worse for low frequency words from sparse neighbourhoods compared to dense neighbourhoods; and 7- and 9-year-olds needed longer segment durations than 11-year-olds and adults for word recognition. This indicates that segmental restructuring is a protracted process which depends on the frequency of exposure to a word and the number of its lexical neighbours (Metsala, 1997a).

More recently, Bowey and Hirakis (2006) employed a mispronunciation task manipulating the position of the mispronounced segment within a word. According to the propositions of the LRM if word-initial mispronunciations show a recognition advantage over mispronunciations occurring later in the word, this would indicate that the target word has been fully restructured in the lexicon - which should be the case with increasing age (Bowey & Hirakis, 2006). Studying the position effect is also consistent with the segment-matching approach discussed earlier (Gerken et al., 1995). However, Bowey and Hirakis found position effects of word-initial or word-medial feature changes in 5- and 6-year-olds and adults only if the mispronunciations also differed in acoustic-phonetic clarity. This indicates that findings of position effects may be confounded with other factors and should alone not be interpreted as evidence for lexical restructuring (Bowey & Hirakis, 2006), as was the case in an earlier study with 5- and 8-year-olds. The results of this study indicated detection differences for word-initial and word-medial features changes in relation to age and lexical familiarity as defined by age
of acquisition ratings (Walley & Metsala, 1990) and were quoted in favour of protracted lexical restructuring (Walley et al., 2003). Another previous study of 4- and 5-year-olds’ mispronunciation detection found a stronger position effect in the 5-year-olds when the items were presented with pictures (Walley, 1987). This contradicts the proposition of segment-based encoding in the more developed lexicon of the older children, and suggests again that there are other variables operational in addition to the position effect.

Together, these studies and critical considerations show that as yet, there has not yet been a direct and conclusive test of the lexical restructuring model and its hypothesis that vocabulary size determines the segment-based reorganisation of the lexicon. Such a test should investigate the relationship of attention to phonological detail to vocabulary size rather than to age, and this should be done in toddler-aged children who could still show differential vocabulary spurt effects (Bowey & Hirakis, 2006). The present study answers this call.

There are, however, other views on the relationship between vocabulary and phonological specificity in the lexicon besides the LRM, some of which suggest a directional influence opposite to the lexical restructuring hypothesis. The vocabulary spurt onset at around 18 months marks a quantitative change in lexicon size and a qualitative change in the way word meaning is mapped onto referents (Nazzi & Bertoncini, 2003, see also section 3.4.2.). It has been suggested that a change in representation mode accompanies and facilitates this rapid vocabulary increase (Nazzi & Bertoncini, 2003). Before the vocabulary spurt, the pairing of a word’s sound pattern and meaning in the lexicon does not include a phonetically-specified representation, but is rather a global sound pattern mapped onto a meaning. In the authors’ view, the vocabulary spurt is the product of a conceptual change in lexical acquisition which frees capacities to attend to phonetic detail in spoken words (Nazzi & Bertoncini, 2003). In the WRAPSA, Jusczyk (1994a) also sees the increase in vocabulary as a function of segmental encoding and the way infants attend to sound patterns in speech.
This discussion leads to two distinct positions on the relationship between phonological sensitivity and vocabulary. Firstly, lexical restructuring and attention to phonological detail in speech are, at least partially, a function of vocabulary growth. Secondly, sensitivity to phonological detail in speech is the driving force for lexical restructuring and consequently, rapid vocabulary acquisition (B. L. Smith, McGregor, & Demille, 2006). In the following predictions, these positions will be formulated as two models, the vocabulary-driven and the phonology-driven model, and explicit hypotheses will be derived regarding the relationship between vocabulary and phonological sensitivity.

4.2. Predictions for Vocabulary-driven and Phonology-driven Models

This thesis proposes a relationship between vocabulary development and attention to phonological detail in speech perception. The direction of this relationship shall be tested in two alternatives: the lexical restructuring hypothesis (vocabulary-driven model) and an opposing position (phonology-driven model). Attention to phonological detail in speech will be studied with two different measures: language-specific speech perception, and phonological sensitivity. The relationship between language-specific speech perception and phonological sensitivity on the one hand, and vocabulary size on the other, will be examined longitudinally from 30 to 33 and to 36 months of age. A possible reason why some studies found no relationship between vocabulary and phonological specificity in word recognition (Bailey & Plunkett, 2002; Swingley & Aslin, 2000) or one restricted only to a single age group (Werker et al., 2002) could be a delay in the effect. This design permits the study of the influence of predictors across a time span of three months on the criterion in regression-based path analyses.

In the vocabulary-driven model, it is hypothesised that vocabulary at 30 months will predict attention to phonological detail in 33-month-olds, and vocabulary at 33 months will predict attention to phonological detail in 36-month-olds.

---

25 The author acknowledges that these terms were developed after suggestions by Judith A. Bowey and Nenagh Kemp.
In the *phonology-driven model*, it is hypothesised that attention to phonological detail at 30 months will predict vocabulary in 33-month-olds, and attention to phonological detail at 33 months will predict vocabulary in 36-month-olds.

4.3. **Study Aims**

The first aim of the study is to chart language development on its three major tiers - speech perception, articulation, and word learning - in the period from 30 to 36 months of age. This has not been done before in a comprehensive approach in this age group. The previous chapters have shown how these tiers develop and are intertwined in the process of language acquisition. The study to be described here will specify further the relationship between these tiers.

To chart language acquisition in 30- to 36-month-olds comprehensively, newly designed tasks and adaptations of previously used tasks must be established. Thus the secondary aim of this study is to develop and test newly-developed and newly-adapted measures of speech perception, articulation, and word learning. The tier of speech perception entails tasks for mispronunciation detection, rhyme detection, and nonword repetition that will constitute phonological sensitivity and measures of native, nonnative, and tonal speech perception that will yield indices of language-specific speech perception. The tier of speech production is represented by a measure of articulation accuracy. The word learning tier is addressed by a productive and a receptive vocabulary measure. As a control variable, a measure of nonverbal cognitive functioning will also be included.

The third aim of this study is to test the vocabulary-driven and the phonology-driven models. To test these models in their basic versions, only one measure of the two concepts language-specific speech perception and phonological sensitivity will be selected on the basis of the suitability of its operationalisation for the age group. Therefore, the *basic vocabulary-driven model* predicts phonological sensitivity or language-specific speech perception from vocabulary size, whereas
the basic phonology-driven model predicts vocabulary size from phonological sensitivity or language-specific speech perception. Both basic models are tested from 30 to 33 months and from 33 to 36 months, each across a test interval of three months.

To do justice to the complexity of language acquisition, both the vocabulary-driven and the phonology-driven models will also be tested in an augmented version. The augmented versions of both models will include the total set of variables to investigate the presence of other predictors beyond the basic models. They will include the other remaining variable out of phonological sensitivity or language-specific speech perception - depending on which was used in the basic models - on the one hand, and articulation and nonverbal cognitive functioning on the other. Both augmented models are also tested from 30 to 33 months and from 33 to 36 months. The augmented vocabulary-driven model will predict either phonological sensitivity or language-specific speech perception from vocabulary size, from either phonological sensitivity or language-specific speech perception, from articulation, and cognitive functioning. The augmented phonology-driven model will predict vocabulary size from phonological sensitivity, language-specific speech perception, articulation, and cognitive functioning.
Chapter V

Method: Tasks, Task Development and Procedures
5.1. PARTICIPANTS

Participants were Australian English speaking children from the Greater Sydney area. Their language development was followed from 30 months to 36 months of age across three test sessions, each three months apart. The children were recruited through the MARCS Babylab register, and therefore had either formerly participated as babies or at least showed interest in participating in infant research studies at MARCS Auditory Laboratories. The parents first agreed to participate in the longitudinal study over the telephone, and then gave their written consent at each of the three test sessions (see CD Consent Form). Due to parental employment, most children were tested on weekends. A total of 83 children were recruited for the study. Of these, following the initial interview and start of testing in the initial session, 15 were not included further due to unavailability (n = 3), higher motivation for physical activities than for computer language games (n = 5), or previously undisclosed bilingualism (n = 7).

All children were considered to be monolingual Australian English speaking, based on parents’ responses on the questionnaire section of the consent form. Nine of the parents indicated that their children had minor exposure (less than ten hours per week) to languages other than English, including Dutch (1), German (2), Spanish (2), Mandarin (1), Greek (1), and Italian (2); no child had been exposed to Thai (the stimulus language in one of the tasks) prior to the study.

The toddlers were tested at three ages, 30 - 31, 33 - 34, and 36 - 37 months, permitting up to four weeks after their birthday for scheduling purposes (see CD Participant Age). Mean age of testing in weeks was 122, 134, and 146. Great care was taken to balance gender in the sample (30 female, 30 male). It was found that with children at toddler age, most mothers had returned to work full time and were therefore only available on weekends. Over the course of data collection, only eight families discontinued their participation, all due to reasons not related to the study such as death in family (1), moving interstate (1) or overseas (1), mother’s illness (1), risk pregnancy with second child (1), and parental workload or other
commitments (3). This results in an extremely low attrition rate of 5.44%, compared to reported drop-out rates of longitudinal studies ranging from 5% to 60% (Farrington, Gallagher, Morley, St.Ledger, & West, 1993), depending on the nature of the study and the length of intersession intervals (Capaldi & Patterson, 1987; Murphy, 1993). The high retention (N = 60) was achieved through establishing good relationships with the participating families (see CD Thankyou Letter).

For participating in each session, parents received $20 to reimburse their travel expenses, and the children chose from a range of small gifts, partly from sponsor donations (see CD Sponsor Logos; and CD Sponsor Letters). At the first session, children also received a MARCS t-shirt (see CD T-Shirt). At the third session, a certificate with the child’s photo, taken at the second session, was presented (see CD Young Scientist Certificate). This study was approved by the Human Ethics Committee of the University of Western Sydney (HEC 05/007, see CD Ethics Approval).

5.2. **APPARATUS AND EXPERIMENTAL CHAMBER**

Each child was required to complete a number of different tasks, involving a variety of equipment. The apparatus is described in section 5.2.1., and its functional set-up in the experimental chamber in the following section.

5.2.1. **Apparatus**

Two computers were used: an Apple iBook computer (G3; 14 inch monitor; OS MacIntosh 9.2), and a Compaq notebook computer (Evo N1000c; 14 inch monitor; OS Microsoft Windows 2000). The Apple iBook was used for the LSSP task (see 5.3.3.) due to the use of the platform-specific PsyScript in the task, and the Compaq notebook for all other tasks. To allow children to touch the screen during testing and provide a larger picture than the laptops would, a robust Mitsubishi Diamond View 1770HB monitor (17 inch) was connected to both computers and used to present all task-supporting visuals such as pictures (for the
mispronunciation detection, rhyme detection, and the articulation tasks) and short clips (for the LSSP task). To display additional reward stimuli for correct responses during the mispronunciation detection, rhyme detection, and articulation tasks, a second monitor was used, a CRT Optima Monitor (CM521M/KM521; 14 inch). Two 2-position data switches intercepted the connection between monitors and computers to enable switching the VGA output between the two screens (see Figure 5.1.). The distribution of tasks across computers and the combinations of laptops and monitors is shown in Table 5.1.

![Circuit diagram](image)

*Figure 5.1.* Circuit diagram for equipment set-up: Switch 1 connects and disconnects the PC laptop and the Mitsubishi monitor, switch 2 alternates the iBook laptop output between the Mitsubishi and the Optima monitor.

<table>
<thead>
<tr>
<th>Switch box</th>
<th>Connection between…</th>
<th>Tasks and Rewards</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A A, B B</td>
<td>Compaq and Mitsubishi monitor</td>
</tr>
<tr>
<td></td>
<td>Compaq and Mitsubishi monitor</td>
<td>Rhyme detection task</td>
</tr>
<tr>
<td></td>
<td>iBook and Optima monitor</td>
<td>Mispronunciation detection</td>
</tr>
<tr>
<td></td>
<td>iBook and Optima monitor</td>
<td>Articulation accuracy task</td>
</tr>
<tr>
<td>2</td>
<td>A B, A B</td>
<td>iBook and Optima monitor</td>
</tr>
<tr>
<td></td>
<td>Compaq and Mitsubishi monitor</td>
<td>Nonword repetition</td>
</tr>
<tr>
<td></td>
<td>iBook and Mitsubishi monitor</td>
<td>LSSP task</td>
</tr>
<tr>
<td></td>
<td>iBook and Optima monitor</td>
<td></td>
</tr>
</tbody>
</table>

All auditory stimuli were presented with an Edirol stereo speaker (MA-10A), connected alternately to either the Compaq or the iBook computer. The speaker volume was set to a mean sound level of 60dB with a range from 55dB to 65dB.
During the LSSP task, the parent - if assisting the child in the task – was required to listen to masking noise played from a CD through Koss headphones (UR-20). Children’s oral responses were recorded onto a DAT Walkman Sony (TCD-D100) recorder using a unidirectional Sony Electret condenser microphone (ECM-ZS90), mounted on the Mitsubishi monitor about 30 cm from the child. This positioning allowed children’s responses to be picked up the child’s response from any location in the room.

5.2.2. Experimental Chamber

The same test room was used for all the different tasks the children completed, and for all three of the test sessions. This provided a stable test environment which allowed sufficient flexibility to maintain the children’s interest throughout each test session. Positioning of apparatus and furniture in the room can be seen in Figure 5.2. The dimensions of the experimental chamber were 200cm x 300cm x 230cm. The colourful furniture for both participant and experimenter was child-sized to facilitate interaction; only the parent enjoyed a normal size chair. The child was usually seated in front of the Mitsubishi monitor, with the parent sitting directly behind. The experimenter handled both computers and the switches from the left hand side while interacting with the child and recording the responses. From this position, the child was about 50 cm from the Edirol audio speaker. During the testing session, the first time sound was used as a reward was a recording of children’s clapping and cheering in the phonological sensitivity tasks and in the articulation accuracy task if necessary. For this, the sound level was set at 65dB. The child’s reaction to this auditory event was monitored, and intensity level adjustments were made accordingly. A record of this was kept for subsequent sessions by the child.
5.3. EXPERIMENTAL DESIGN AND VARIABLES

At each of the three test sessions, the children were tested for their ability on four variables. These were vocabulary (two tasks, applicability depending on age), phonological sensitivity (three tasks, mispronunciation detection, rhyme detection, and nonword repetition; all presented at each age), articulation, and language-specific speech perception. Additionally, at 36 months, the Stanford-Binet V subtest Fluid Reasoning was used as a measure of nonverbal cognitive functioning.

5.3.1. Vocabulary

Measuring early vocabulary size is essential not only for many questions in language acquisition research (E. Bates et al., 1994; Fenson, Pethick et al., 2000), but also for detecting abnormalities in language development in clinical practice (Arriaga et al., 1998; Dale & Fenson, 1996; Fenson, Bates et al., 2000; Fenson et al., 1994). Since the receptive lexicon of children around three years has usually reached proportions at which it becomes impractical to attempt measuring the total lexicon size (see Chapter III), only productive vocabulary was measured exhaustively in this study. This was done via a vocabulary checklist, administered only at 30 months. For this purpose, a modified version of the MacArthur Communicative Development Inventory (Fenson et al., 1993), was completed by the parent during the 30-month test session. At 33 and 36 months, a score for receptive vocabulary was obtained by using a standardised vocabulary test. The standard Peabody Picture Vocabulary Test III (Dunn & Dunn, 1997) was used for this purpose.

5.3.1.1. The Australian English vocabulary inventory OZI

Background

Vocabulary checklists have proved to be the most efficient and most representative measure to study early vocabulary development. Among these, the most widely used is the MacArthur Communicative Development Inventory CDI
(MacArthur CDI for short), developed in the USA for North American English speaking children (Fenson et al., 1993).

The MacArthur CDI is a list of 680 words via which parents indicate their child’s productive vocabulary by checking which of the words they have heard their child say (Reznick & Goldfield, 1994; Reznick & Goldsmith, 1989; Robinson & Mervis, 1999). The MacArthur CDI has been adapted to British English - the Oxford CDI (Hamilton et al., 2000), and to New Zealand English (Reese & Read, 2000); and has also been translated into many of the world’s languages, including Chinese, Dutch, French, German, Greek, Italian, Japanese, Polish, Spanish, Swedish, Russian, and Thai. The OZI is an Australian English adaptation of the MacArthur CDI, designed with permission of the CDI Advisory Board (see CD Permission Request and CD CDI Board Permission). The OZI is based on and adapted from the MacArthur CDI, with some consideration of the Oxford CDI vocabulary list (see CD Shared Words and CD OZI Questionnaire).

Development

The American English MacArthur CDI (Fenson et al., 1993) is available in two versions: “Words and Gestures” (WG) for infants from eight to 16 months, and “Words and Sentences” (WS) for toddlers from 16 to 30 months. In the following we are only concerned with the toddler version of the MacArthur CDI. It consists of two parts: Part I, a list of the most common words children use at that age, and Part II, an extension beyond a simple vocabulary count into sentence structuring and grammatical skills.

The primary goal of developing the OZI was to adapt the MacArthur CDI to contemporary Australian English. Another goal in the development of the OZI was to shorten the time parents spend on filling out the questionnaire. This was achieved by concentrating on a list of nouns, verbs and descriptives only, and omitting the following sections from Part I of the MacArthur CDI: section 16,

---

26 The CDI III for children from 30 to 37 months of age has not yet been published and is only available upon request via the MacArthur CDI webpage (http://www.sci.sdsu.edu/cdi/cdi3_e.htm).
words about time; section 17, pronouns; section 18, question words; section 19, prepositions and locations; section 20, quantifiers and articles; section 21, helping verbs; and section 22, connecting words. This resulted in deleting a total of 114 lexical items. In Part II of the MacArthur CDI, parents are required to indicate the sentence complexity level their child has reached on sample sentences differing in length and grammatical sophistication, as well as to indicate their child’s use of irregular word forms and endings including overregulizations. Additionally, parents are asked to provide three of the longest sentences they have heard their child say, on which the calculation of the Mean Length of Utterance (MLU) is based. In the OZI, only the sections on MLU, irregular word forms, and word endings were retained, as the lexical entry count was of greater importance than that of complexity level of language ability; and MLU can sufficiently predict sentence production27 (Parker & Bronson, 2005; Rice, Redmond, & Hoffman, 2006). Following the example of the short forms of the MacArthur CDI, CDI-WORDS (Corkum & Dunham, 1996; Fenson, Pethick et al., 2000), these deletions reduced the administration time of the OZI for the parents to 20 minutes overall.

Those sections retained in the OZI were adapted to the language environment of Australian toddlers by replacing, deleting, and adding items. Replaced were, for example, in section 2 “alligator” with “crocodile”, “squirrel” with “possum”; section 3 “sled” with “pram”; section 5 “soda” with “cordial”; section 6 “diaper” with “nappy”; section 10 “sidewalk” with “footpath”; and section 11 “downtown” with “city” and “gas station” with “petrol station”. There were a total of 34 replacements. Deleted were, for example, in section 2 “moose”; section 5 “jello”; section 6 “mittens” and “snowsuit”; and section 9 “basement”. There were a total of 14 deletions. Added were a total of 11 items, for example in section 10 “wall”; section 11 “barbecue”; section 14 “cuddle”, “tell”, and “smell”; and section 15 “easy”, “nasty”, and “yummy”. Contrasting the total of sections 1 to 15 in the two questionnaires, the MacArthur CDI contains 566 items compared to 559 in the Australian Vocabulary Checklist (see CD Comparison of Item Number).

27 The exclusion of the grammatical sections in the OZI is also justified by the substantial correlations found between vocabulary, sentence complexity, and grammar (Dale, Dionne, Eley, & Plomin, 2000).
Inconsistencies in the number of items per section between the two questionnaires that are not a result of replacements, deletions, or additions, are due to merging of two words that are used interchangeably, for example “veranda” and “porch” in section 9.

\[ \text{Figure 5.3a.}: \text{Pilot OZI data from 24-month-olds: As the significant differences in the section scores permitted to conclude, the OZI total score is also significantly higher than the CDI total score (significance is marked with *).} \]

\[ \text{Figure 5.3b.}: \text{Pilot OZI data from 30-month-olds: As the significant differences in the section scores permitted to conclude, the OZI total score is also significantly higher than the CDI total score (significance is marked with *).} \]

**Pilot Study and Norm Collection for the OZI**

Before using the OZI as measure of vocabulary size in the present study, a preliminary study was conducted to compare CDI and OZI and determine the degree of equivalence despite the adaptation for cultural and linguistic relevance. One CDI and one OZI questionnaire each were sent out to parents of 24-month-olds (n = 32; 14 females, 18 males) and 30-month-olds (n = 32; 17 females, 15 males). In each age x gender subgroup half the parents completed the CDI first, then the OZI, and the other half the OZI, then the CDI. The two tests correlated highly at both ages (r = .99 in 24-month-olds, r = .97 in 30-month-olds). In addition, the greater appropriateness of the OZI for Australian parents was demonstrated in both age groups. As can be seen in Figures 5.3a. and 5.3b., the Australian toddlers scored significantly higher on the OZI in both age groups (see CD Update OZI for further information and data regarding the OZI pilot test). After approval by the CDI Advisory Board was granted to use the OZI for research purposes and to develop it into the official Australian English version of the worldwide MacArthur
CDI family, norm collection begun at MARCS Auditory Laboratories in 2004. So far data from around 300 children has been accumulated. Their vocabulary is not only recorded as a total score but also itemised, providing true age of acquisition norms for Australian English (see section 3.3.3. for criticism on previous age of acquisition data). Preliminary vocabulary size data are shown in Figure 5.4. and further detail regarding the age of acquisition for each of the 559 words of the OZI are shown in CD Update OZI.

![Graph](image)

*Figure 5.4.:* Preliminary vocabulary development in Australian toddlers (16 to 30 months old) measured with the OZI (N = 294). The data are preliminary due to low participant numbers with high individual variation, therefore SE was not depicted (for further information, see CD Update OZI).

**Administration and Scoring**

The OZI was completed by the parent present during the first test session and scored afterwards by the experimenter.

### 5.3.1.2. Peabody Picture Vocabulary Test III

**Background**

Form A of the Peabody Picture Vocabulary Test III (PPVT III) was used at 33 and 36 months. The PPVT III (Dunn & Dunn, 1997) is an individually administered, untimed, and norm-referenced test of receptive vocabulary development. It consists of 204 test items, grouped into 17 blocks of twelve items each, and arranged in increasing difficulty. Each item consists of four black-and-white drawings, presented on one picture plate to the participant. The participant
is asked to point out the picture that best represents the target word. Each error is recorded by the experimenter. The designated age of the PPVT participants ranges from 2½ to over 90 years, and allows for different starting points according to age, with four sets of age-appropriate training items. Administration of the PPVT III follows two basic rules: the basal set rule and the ceiling set rule. The basal set rule states that the participant must achieve one or no errors in the first set of twelve test items, to allow use of the norms. The ceiling set rule determines the end of a test session; the session ends once there are eight or more errors in any particular set.

**Administration and Scoring**

Although the PPVT III is standardised, the reference norms could not be used in this study because the basal set rule was not met by all children at the first test time at 33 months (n = 32, 15 females, 17 males) for reasons that were suspected to lay in nature and age of the drawings. However, as this was a research study rather than a clinical exercise, and since scores of the same child across ages were to be compared, standardised scores were not desirable. Therefore, the raw score was used instead of the standard score. The experimenter controlled the format and content of the conversation with each participant as much as possible while accommodating each child individually. The standard conversation was as follows: “We are going to look at some pictures together. I’m going to show you four pictures and ask you to show me one of them. You show me the one.” The pointing response to “Can you show me...?” was practiced until satisfactory with the training items “Ball” and “Sleeping” before moving on to test set 1 (see Figure 5.5.). When asking for each target word, the definite article for nouns was omitted, in compliance with the PPVT manual, in order to disguise the difference between a verb and noun. The first pointing response was recorded every time, even though sometimes the children corrected their initial response. However, in cases of lack of

![Figure 5.5.: “Can you show me...?” The toddler points to the picture that best represents the word the experimenter requested.](image)
attention to the task, each target word was requested up to three times before the item was recorded as an error. Praise was given to any legal response rather than only for correct responses in accordance with the standard testing procedure. Administration of the PPVT III to this age group took about ten minutes; most children did not go beyond test set 6. The task rarely had to be presented in two blocks as the majority of children enjoyed working with the PPVT, and completed it in a single test block. The raw vocabulary score was calculated by the total number of items attempted minus the number of errors.

5.3.2. Phonological Sensitivity

Background
Commonly used test methods of phoneme awareness for 4-year-old preschoolers are tasks of phoneme recognition, phoneme blending, phoneme counting, phoneme deletion, phoneme segmentation, nonword repetition, onset-rime identification and rime identification (Van Bon & Van Leeuwe, 2003), phoneme substitution (Bialystok et al., 2003), and invented spellings (Mann et al., 1987) (see Chapter I). To obtain a valid measure for the equivalent of phoneme awareness in toddlers, three phoneme awareness tasks were selected out of this spectrum of regular tasks that could be adapted to a younger age group: a mispronunciation task, a rhyme detection task, and a nonword repetition task. The psycholinguistic construct for these tasks can be seen as a fledgling, a developing form, of phoneme awareness, and can therefore be called phonological sensitivity (e.g., Bowey, 2001; Burgess & Lonigan, 1998). Phonological sensitivity was measured here at 30, 33, and 36 months with the three tasks mentioned above, and then, justified by the phoneme awareness concept, a single phonological sensitivity score was derived from principal components analyses conducted at 30, 33, and 36 months (see Chapter VI).

Stimulus Materials
The stimuli for the mispronunciation detection, rhyme detection, and nonword repetition tasks were recordings of two young female Australian English speakers. Both speakers were familiar with recording procedures. The recording manuscripts
were prepared to prevent list effects by shuffling and repeating items, and including buffer items. The items were recorded onto DAT (Walkman Sony TCD-D100) with a tie-clip Sony microphone (ECM-T145), and then transferred to computer hard drive (Hewlett Packard Vectra VL-400, Delta 66 sound card) using Cool Edit 2000 and keeping the sampling rate of 44100 Hz, mono, and 16 bit resolution. Each file was amplified to 95% to normalise loudness. The fade in and out function was applied to each file, and they were automatically noise-filtered for frequencies under 75 Hz through a Matlab 7.1 batch routine. Speaker A produced the mispronunciation detection items, and speaker B produced both the rhyme detection and the nonword repetition items. All phonological sensitivity stimuli were run from the Compaq and presented via the Edirol stereo speaker (MA-10A).

5.3.2.1. Mispronunciation Detection

Background
Sensitivity to mispronounced words is an ability that infants acquire between 14 (Bailey & Plunkett, 2002) and 18 months (Werker et al., 2002). There is a link between mispronunciation detection skills and vocabulary size (Werker et al., 2002) because children discriminate minimal word pairs better if they know the words well (Barton, 1976, 1980). Word initial phoneme substitutions are detected earlier than changes in other word positions (P. W. Jusczyk, Goodman, & Baumann, 1999), however, 19-month-old Dutch infants are also sensitive to word-medial substitutions (Swingley, 2003). Naturally, the number of deviating features per word is linearly related to word recognition (White, Morgan, & Wiers, 2004). A mispronunciation detection task with a one-feature difference in word-initial and word-medial position is an age-appropriate predecessor of phoneme substitution tasks used to test phoneme awareness in preschoolers and school children and has been used in similar age groups (e.g., Bailey & Plunkett, 2002; Swingley, 2003). It was therefore included into the set of phonological sensitivity tests used here.
Development

A mispronunciation task previously used with 4-year-olds (Bowey & Postle, 2003) was the basis for this task. Following consideration of error data on this task from Bowey and Postle (2003) and given the age of acquisition values of the MRC database of the University of Western Australia (Coltheart, 1981; Coltheart & Wilson, 1981), a list of 20 words was created, with each word represented in a correctly pronounced as well as a mispronounced form (see CD Scoring Sheet). The total of 40 test items was preceded by six practice items (three correctly and three mispronounced) which included one initial- and two medial-phoneme substitutions. During the second and third year of life, children normally substitute the liquids /r/ and /l/ with /w/ and de-aspirate initial stop consonants like /kb/ in “car” to /k/ (Eilers & Oller, 1976). This made items like “window-window”, “rainbow-wainbow”, and “bucket-pucket” especially interesting for a potential correspondence between perceptual errors and production substitutions (see Chapter 2.1.3.).

Table 5.2.: Overview of the mispronunciation detection task items grouped into four sections: practice items, and test items with easy, medium, and high difficulty. Each word is listed with its phoneme substitution. Stress in the syllable of the substituted phoneme is marked with S, secondary stress and/or unstressed syllables are coded with U. The position of the substituted phoneme is marked with 1 for word-initial, and 2 for word-medial.

<table>
<thead>
<tr>
<th>Practice</th>
<th>Stress Position</th>
<th>Easy</th>
<th>Stress Position</th>
<th>Medium</th>
<th>Stress Position</th>
<th>Difficult</th>
<th>Stress Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>app/bble</td>
<td>S 2 t/koothpaste</td>
<td>S 1  c/gamel</td>
<td>S 1  isl/nand</td>
<td>U 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pupp/cky</td>
<td>U 2 t/kurtle</td>
<td>S 1  peac/gock</td>
<td>U 2  ball/noon</td>
<td>U 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m/oney</td>
<td>S 1 guit/kar</td>
<td>U 2  m/hailbox</td>
<td>S 1  eye/wash</td>
<td>U 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S 1 b/pucket</td>
<td>U 2  firen/ban</td>
<td>U 2  liz/vard</td>
<td>U 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S 1 tah/ple</td>
<td>U 2  m/nushroom</td>
<td>S 1  w/window</td>
<td>S 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>U 2 rahb/ppit</td>
<td>U 2  hamm/ner</td>
<td>U 2  r/wainbow</td>
<td>S 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>r/looster</td>
<td></td>
<td></td>
<td>S 1 carr/lloot</td>
</tr>
</tbody>
</table>

The experimental items were classed according to the difficulty of the substitution into three groups: easy, medium, and difficult (for phoneme acquisition ages and recognition of place of articulation changes, see Chapter II). Since position effects were not of interest here, the words were only loosely matched for syllable stress,
position of the substituted phoneme, and the substituted phoneme itself (for
details, see Table 5.2.).

The items were always presented in the same pre-arranged order with the goal to
prevent randomly missing data by being able to exclude the last few items if the
majority of children did not finish the task. The mispronunciation detection task
words were recorded in isolation as well as within two questions: “Can you show
me the …?” and “Where is the …?” These questions were then used in a
preliminary study to test whether children were familiar with the picture items. In
the mispronunciation detection task, a photo was presented with every auditory
stimulus to provide visual support for the task to the child (see CD Pictures). To
establish the child-friendliness of the depicted objects, a preliminary study was run
with 14 30-month-olds (six female, eight male). The children were requested to
point to the item in question from a set of four photos. Most items were correctly
identified by all children, except for the item pairs “eyelash-eyewash” and “island-
island”, which were already classed as difficult, and could only be pointed out by
50% of the pilot study participants (see CD Picture Pilot Results). Nevertheless,
both were still included to prevent ceiling effects in the 36-month-olds, as it could
be safely assumed that they would be acquired soon by most participants, and the
mispronunciation detection task did not explicitly require children to know the
meaning of the items.

Administration and Scoring

The child was instructed to “listen whether
the lady says the word right”, answering with
“yes” or “no” (Figure 5.6.). In an earlier
version of the mispronunciation detection
task, questions offering two alternative
answers such as “Is the lady right or wrong?”
proved to be not suitable for toddlers at this
age, as they showed a response bias to the
latter alternative. Administration of the
misprediction detection task usually took a maximum of ten minutes, as the
majority of children enjoyed finding “mistakes” in the recorded words. Correct
responses were intermittently rewarded with an animated picture displayed on the
second monitor to the child’s right, along with the sound of children clapping and
cheering. The responses were scored manually by the experimenter and converted
into a mispronunciation detection index according to the formula hits minus false
positives divided by the total number of trials (Table 5.3.).

Table 5.3.:  
Mispronunciation detection responses and their coding in regard to the discrimination index formula: the
child replied to the question: “Did the lady say it right?” Main target is the detection of mispronounced
items.

<table>
<thead>
<tr>
<th>Child’s response</th>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correctly pronounced items</td>
<td>Correct rejection</td>
<td>False positive</td>
</tr>
<tr>
<td>Mispronounced items</td>
<td>Miss</td>
<td>Hit</td>
</tr>
</tbody>
</table>

5.3.2.2. Rhyme Detection

Background
Onset and coda rime are commonly part of phoneme awareness test batteries
(Anthony et al., 2002; Bowey, 1990; Foy & Mann, 2003). Rhyme awareness has a
global nature (Bryant, 1998; Cardoso-Martins, 1994) and therefore has less
predictive value for early reading skills than tasks focusing on smaller phonological
units, for example phoneme segmentation tasks (Goswami & Bryant, 1990;
Hulme, 2002; Hulme et al., 1998; Muter et al., 1998). However, there is the
possibility that the value of rhyme awareness has been underestimated (Bryant,
1998; Goswami, 2002; Hulme et al., 2002; Hulme et al., 1998). Besides the debate
about the value of rhyme awareness in predicting reading ability (Bowey, 1990;
Bryant, MacLean, & Bradley, 1990; Kirtley, Bryant, MacLean, & Bradley, 1989),
there is also a relationship between rhyme awareness and vocabulary size via
Nevertheless, rhyme awareness is a intrinsic component of phonological sensitivity
(Anthony et al., 2002), and has been previously tested in developmental settings
with 2-year-olds (Anthony et al., 2002; Kehoe & Stoel-Gammon, 2001). For these
reasons, it was selected to measure phonological sensitivity in this study.
Development

The rhyme oddity task from Bowey (n. d.) was adapted for the rhyme detection task here. Rhyme items were selected according to toddlers’ vocabulary, for example “gun” was replaced with “bun”, “wig” with “big”, and “rake” with “lake”. In an initial version of the task, each trial consisted of three words, a rhyming word pair and a non-rhyming word. Participants were required to identify the “odd” word (see CD Oddity Scoring Sheet). Pictures of the items were presented with oral presentations of the word in sequences of three in a Microsoft Powerpoint 2000 presentation. This rhyme oddity task was pilot-tested with six 24- to 30-month-olds. They tended to focus primarily on the pictures rather than listening for the non-rhyming word and appeared not to be able to keep three items in their phonological short term memory. As a result, the task was completely re-designed to work with toddlers\(^\text{28}\).

In the re-designed task, the level of difficulty was decreased by making it a simple detection task; the children had to decide whether a word pair rhymed or not. The word items were re-used and arranged into 28 pairs, 50% rhyming and 50% non-rhyming pairs plus four practice items, all presented in pre-arranged order with rhyming and non-rhyming pairs immediately following each other (see CD Detection Scoring Sheet). This was to prevent missing data in the case of children who did not finish the task. To provide a visual basis for the detection task, two pairs of colourful snakes were introduced to the toddlers. The snakes matching in colour looked “the same” and sounded “the same” (i.e. their corresponding word pairs rhymed); the snakes with different colours looked and sounded “different” (i.e. their word pairs did not rhyme). The audio-only items in the rhyme detection task along with the two pairs of snakes presented visually were presented from the Compaq using Microsoft Powerpoint 2000.

\(^{28}\) The original rhyme oddity task is successfully in use in a related longitudinal project with 3-, 4- and 5-year-olds, conducted in collaboration between the University of Western Sydney and the University of Melbourne since 2005.
Administration and Scoring

During the practice items of the “snake game”, the child learned to map rhymes onto the identical snakes (the slightly different shade of the matching snakes’ head was to reflect the fact that the “head” of the word was still different) and non-rhymes onto the different snakes (Figure 5.7a. and 5.7b. respectively) before proceeding to the test trials. On each trial, the child was asked which snakes said the words. The child’s response to the presented word pairs was to touch the corresponding pair of snakes on the screen (Figure 5.8.). Correct responses were rewarded with an animated picture on the second monitor to the child’s right, accompanied by a recording of children clapping and cheering, manually elicited from the iBook by the experimenter. Administration of the rhyme detection task took ten minutes, and was often split by a 15 minute-break halfway through each test session. This break was required, as most children struggled with the concept of rhyme and became bored. After the break, the last item presented in the previous session was repeated to refresh the child’s memory of the mapping of word and snake pairs. The rhyme detection task was scored manually in percent correct of the total number of items. To check for response biases such as “accept all” or “decline all”, the number of correctly identified rhymes and correctly declined non-rhymes was also counted separately.

5.3.2.3. Nonword Repetition

Background

Nonword repetition tasks are part of the standard repertoire of phoneme awareness batteries, and their scores are commonly used in relation to vocabulary
size (Metsala, 1999), phonological memory (Gathercole, Willis, Baddeley, & Emslie, 1994), and phonotactic probability (Edwards et al., 2004; Munson, Edwards, & Beckman, 2005b; Munson, Kurtz, & Windsor, 2005). The more ‘wordlike’ a nonword is and the more familiar the corresponding word (the word from which the nonword has been constructed by modification), the more likely correct repetition is (Gathercole et al., 1994). What nonword repetition tests exactly measure, is debated (Bowey, 1997; Gathercole, Willis, & Baddeley, 1991; Snowling, Chiat, & Hulme, 1991), especially since the task includes a production component. Nonword repetition scores are influenced by the lexicon (Bowey, 2001; Dollaghan, Biber, & Campbell, 1995; Metsala, 1999), by wordlikeness and word frequency (Frisch, Large, & Pisoni, 2000; Gathercole et al., 1994; Munson, Kurtz et al., 2005), and by neighbourhood density (number of words with the difference of a single phoneme addition, deletion, or substitution) and neighbourhood frequency (frequency of the items in the neighbourhood) (Lipinski & Gupta, 2005; Vitevich & Luce, 2005). There is a range of different tests for nonword repetition available (Archibald & Gathercole, 2006; Bowey, 1996, 2001; Gathercole et al., 1994), none of which has been reported as having been administered to an age group younger than three years (Metsala, 1999). A nonword repetition task was used here at 30, 33, and 36 months, and its development is described in the next section.

Development

A list of nonsense words from Bowey (n. d.), formerly used for nonword repetition with 4-year-olds, was adapted to younger children. All consonant clusters were excluded from the list, and consonants whose pronunciation is more difficult were replaced with easier ones - for example, /w/, /ð/, and /ð/ were not used (Ingram, 1989). The items were matched to common words familiar to children, avoiding phonotactically illegal combinations in

---

29 In 2006, Applied Psycholinguistics dedicated a complete issue (27/4) to the debate on nonword repetition. Vance, Stackhouse, and Wells also give a good summary on the task and its application (2005).
English. The resultant nonword repetition task contained a total of 20 items of increasing difficulty due to increasing number of syllables, plus four practice items (see CD Scoring Sheet). There were six one-syllable, six two-syllable, and four three-syllable nonwords. The practice items contained two one-syllable and two two-syllable nonwords. Bowey’s “Nonword Repetition Game” (n.d.) for 4-year-olds is driven by a story about the puppet “Pemnie” which speaks a different language and wants to know whether the child can repeat particular words. In the current task, a big soft toy pyjama case became “Pemnie” (see Figure 5.9.); Pemnie’s words, the nonword items, were presented with the program Irfanview version 3.85 (Skiljan, 2003) via the Compaq PC. Irfanview is free web-based software, and was used to play recorded nonwords of the nonword repetition task following a pre-designed playlist saved as text document.

Administration and Scoring

Administration of the nonword repetition task took five minutes, depending on the child’s level of compliance. Pemnie was introduced to the child in a play-like manner and explained the game to follow (see Figure 5.10.). All children realised that Pemnie did not actually say the nonsense words, so it was explained that Pemnie had words in his language on the computer, and would play them through a loudspeaker (see section 5.2.1.). Each nonword was presented a maximum of two times. Sometimes it was necessary to involve the parent who would repeat the recorded nonword a third time before the child was willing to respond - a situation that became clear during the practice trials and could be accommodated for. To give the parent visual assistance in such conditions, the nonword list was displayed on the wall in the test room (see CD Nonword List). The children’s responses were recorded on a DAT recorder via a stationary condenser microphone (see section 5.2.1.). Recordings of each child’s responses were scored manually after each test session and marked either as correct after the first or second presentation; or incorrect, in which case the child’s version of the nonword was phonetically transcribed onto a response sheet. The child was
required to pronounce each phoneme exactly as in the original nonword to receive a “correct” score. In cross-reference with the articulation task, it was checked whether incorrectly repeated items in the nonword repetition task contained phonemes that were not correctly produced when naming pictures. If so, the mispronunciations or refusal to repeat a nonword may derive from immature articulator control (Gathercole et al., 1994) rather than from inaccuracies in perception and processing of phoneme sequences.

5.3.3. Language Specific Speech Perception

Background
Language-specific speech perception, LSSP for short, was developed to measure the degree of phonological specialisation in speech discrimination in the native language relative to the performance in nonnative language (Burnham, 2003). Scores for nonnative speech discrimination (NN) are subtracted from native speech discrimination scores (N); a high N-NN indicates a strong orientation towards the native language and relatively less orientation to nonnative speech sounds; a low N-NN score demonstrates approximately equivalent performance in native and non-native perception, thus less specialisation into the native language; and negative N-NN scores reflect greater attention to the nonnative language or confused responding (see Chapter 1.2.4). The LSSP discrimination index is calculated according to a variation of the $d'$ formula. This calculation provides a maximum score of +1 for the maximum number of hits and no false positives, a score of 0 for responses at chance level with an equal number of hits and false positives, and a negative score for more false positives responses than hits.

$$\text{Discrimination Index} = \frac{\text{correct change trials}}{\sum \text{change trials}} - \frac{\text{correct no-change trials}}{\sum \text{no-change trials}}$$

Stimulus Material Development and Selection
The stimuli for this study were part of a larger sample of Thai sounds, recorded for a number of different studies at the University of Western Sydney and the
University of Melbourne. Three female native Thai speakers were originally recorded, pronouncing randomised sound lists (see Table 5.4.). The stimuli were recorded directly onto a Hewlett Packard Vectra PC (VL-400) with a Delta 66 sound card via a MACKIE 1202-VLZ PRO mixer in a sound-attenuated recording booth, using a Rode NT2 microphone (sampling rate of 44100, mono, and 16 bit resolution). Noise less than 75Hz was filtered out during the recording; the Windows PCM files in wav-format were amplified to 95% to normalise loudness, and the fade in/fade out function was applied to each stimulus in Cool Edit 2000 (see CD Thai Stimulus List). All recorded stimuli were presented to an expert panel consisting of two Thai native speakers (one male, one female), and one trained phonetician in order to select the three best exemplars per sound for this study (for details of the selection procedure, see CD Thai Stimulus Selection and CD Stimulus Rating Sheet).

The final selection of sounds for the LSSP task contrasted the bilabial stops [ba], [pa], and [pʰa] in mid-level tone, the dental/alveolar plosives [da], [ta], and [tʰa] also in mid-level tone, and the velar plosive [ka] in low (1), high falling (2), and low rising tone (4) (see section 1.1.2.2., Table 1.1.). For each sound, the three best

<table>
<thead>
<tr>
<th>Sound</th>
<th>Tone (numeral coding)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ba]</td>
<td>0 1 2 3 4</td>
</tr>
<tr>
<td>[pa]</td>
<td>0 1 2 3 4</td>
</tr>
<tr>
<td>[pʰa]</td>
<td>0 1 2 3 4</td>
</tr>
<tr>
<td>[da]</td>
<td>0 1 2 3 4</td>
</tr>
<tr>
<td>[ta]</td>
<td>0 1 2 3 4</td>
</tr>
<tr>
<td>[tʰa]</td>
<td>0 1 2 3 4</td>
</tr>
<tr>
<td>[ka]</td>
<td>0 1 2 3 4</td>
</tr>
<tr>
<td>[kʰa]</td>
<td>0 1 2 3 4</td>
</tr>
</tbody>
</table>

Table 5.4.: Matrix of recorded Thai speech sound and tone class combinations.

<table>
<thead>
<tr>
<th>Sound</th>
<th>Tone (numeral coding)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ba0</td>
<td>1 606 576 667 632 533 607 613 550 535</td>
</tr>
<tr>
<td>pa0</td>
<td>2 626 525 646 634 484 528 558 502 538</td>
</tr>
<tr>
<td>pʰa0</td>
<td>3 629 534 629 599 508 585 593 425 502</td>
</tr>
</tbody>
</table>

Table 5.5.: Sound duration (in ms) for the best three exemplars each with numeral tone coding (0 = mid-level tone, 1 = low level, 2 = high falling, and 4 = low rising).

<table>
<thead>
<tr>
<th></th>
<th>ba0</th>
<th>pa0</th>
<th>pʰa0</th>
<th>da0</th>
<th>ta0</th>
<th>tʰa0</th>
<th>ka1</th>
<th>ka2</th>
<th>ka4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>606</td>
<td>576</td>
<td>667</td>
<td>632</td>
<td>533</td>
<td>607</td>
<td>613</td>
<td>550</td>
<td>535</td>
</tr>
<tr>
<td>2</td>
<td>626</td>
<td>525</td>
<td>646</td>
<td>634</td>
<td>484</td>
<td>528</td>
<td>558</td>
<td>502</td>
<td>538</td>
</tr>
<tr>
<td>3</td>
<td>629</td>
<td>534</td>
<td>629</td>
<td>599</td>
<td>508</td>
<td>585</td>
<td>593</td>
<td>425</td>
<td>502</td>
</tr>
</tbody>
</table>

M = 620.3 545 647.3 621.7 508.3 573.3 588 492.3 525

exemplars had been selected by the panel. In the native condition, [pa] - [pʰa] and [ta] - [tʰa] were contrasted. In the nonnative condition, [ba] - [pa] and [da] - [ta] were contrasted (for sound exemplar duration, see Table 5.5.). In the tone
condition, an easy and a difficult tonal discrimination were tested: [ka2] - [ka4] and [ka1] - [ka4] (see CD Counterbalancing Conditions). To create a masking tape for the parent, exemplars of all sounds were randomly concatenated without pauses, and copied onto one channel of a stereo PCM wav file, using Cool Edit 2000. Music from the “Forrest Gump” soundtrack was fed into the other channel, also without pauses between tracks. Such a procedure was followed so that when the parent assisted the child in the LSSP task, the experimental stimuli would be effectively masked by the tape so that the child’s behaviour was not unintentionally influenced by the parent (Pinto, Fernald, McRoberts, & Cole, 1998).

**LSSP Task: Program, Apparatus and Task Structure**

The program to run the LSSP task was written in PsyScript 5.1.d1 (T. Bates, 2002; T. Bates & D’Oliviero, 2003), a free web-based experimental program from Macquarie University Sydney by a MARCS lab programmer (Haszard Morris, 2004). PsyScript is based on Applescript, and version 5.1.d1 required the use of the MacIntosh OS 9 operating system.\(^{30}\) The LSSP task program (see CD LSSP Task Program Script) integrates subscripts for speech contrasts and reward clips, using MacIntosh Smile Script Editor. To analyse the output files of the LSSP task program, a LSSP task analysis program (Chen, 2005) was written in Matlab. This calculates the discrimination indices for each condition and converts the data into a Microsoft Excel file (see CD LSSP Task Output Script).

The speech sound pairs were contrasted in an AnX go-/no-go procedure in which participants are required to judge whether the following sounds are the same or different than the previous ones. Sounds were presented continuously with an interstimulus interval of 300 ms to elicit linguistic rather than just acoustic processing (Werker & Logan, 1985; Werker & Tees, 1984b), and to keep the trial length in the experimental phase as short as possible. Each trial consisted of two to eight presentations of stimulus A, followed by a maximum of three repetitions of

\(^{30}\) Now PsyScript version 5.3, compatible with MacIntosh OS X, is now available and can be downloaded at http://www.marcs.mq.edu.au/~tim/psyscript/.
stimulus X before the response window expired. To indicate perception of a sound change, the child hit a button on top of a response box in front of them (see CD Button Box). This type of response was chosen because Trehub (2003) found that manual engagement within a discrimination task (a computer touch screen in her case) worked well with 2-year-olds. Control trials without a sound change were also included. For these a correct response (correct rejection) involved not pressing the button in the trial. The LSSP task program recorded hits and misses as well as reaction time, measured as a time window starting from the onset of the X stimulus and lasting for three repetitions. Each trial ended with either a button press or timing-out of the three change trial repetitions.

The LSSP task consisted of three phases; a task familiarisation phase, a training phase, and an experimental phase. Each trial was started by the experimenter once the child showed task focus. In the task familiarisation phase, recordings of two different animal sounds (rooster and cow) were used to train the child by pressing the button to a sound change. All four trials were change trials. In the training phase, a vowel change in Australian English words, “rag” versus “rug”, was used. The child proceeded to the experimental phase when they responded correctly to six out of the last eight trials\(^{31}\), with a maximum of two correct rejections. The experimental phase contained six blocks of target speech sounds; the native, nonnative consonant and nonnative tone conditions were each repeated once with inverted AnX contrast order. At the start of each block there were four demonstration trials to familiarise the child with the type of contrast. There were eight experimental contrasts in each block of test trials. In the demonstration trials for each block, one trial contained no change; and in the test trials, there were two no-change trials. Preliminary testing revealed that the task was demanding for the children. For this reason stimuli for only one Thai speaker were chosen, only one exemplar of each sound was randomly selected per trial rather

\(^{31}\) Compared to commonly used criteria for passing the conditioning stage in head-turn paradigms which are nine out of ten (Grieser & Kuhl, 1989; Kuhl, 1983) or seven out of eight (Mattock & Burnham, 2006) with a chance level of 0.05, the criterion six out of eight trials results in a probability of 0.11 at chance level. This can lead to overinclusion of participants in the LSSP task, but was necessary due to the age-related high task difficulty that became apparent during pilot testing.
than varying between the three exemplars available per sound, and the order of contrast presentation (i.e., the order of sounds within a pair) was only alternated between blocks. To maintain the child’s interest in the task, a reward system was incorporated into the LSSP task program, using short animated video clips of four to seven seconds duration. Both correct discrimination (hits - button presses) and correct rejection (no button press) were rewarded with running a video clip. On the next correct trial the same story continued with up to eight parts making up a complete story within a block. The animation froze during trial presentation and was re-animated only by correct responses, either hits or correct rejections, presumably increasing the child’s motivation to see the story continued.

**LSSP Task Administration and Scoring**

Each child was randomly assigned one of the twelve conditions. The condition assignment remained the same for each participant for all three test sessions to maintain consistency and to allow analyses over time.

Before the task familiarisation phase, the child practiced hitting the button as fast as possible. In the four change trials of the task familiarisation phase, pressing the button in response to sound change was practiced with animal sounds. Conveying the objective of the LSSP task was assisted by explaining the same-different notion: “We are going to listen to some sounds together. The sounds are going to be the same for a while. Then there is a different sound. Listen for the new sound and press the button as fast as you can.” (see Figure 5.11.). For the first two trials in the task familiarisation phase, the experimenter would demonstrate the desired response with the child’s hand, and then encourage an independent reaction by the child. In quite a few cases, the children showed reluctance to press the button at all. As the parents were “deaf” to the experimental conditions due to the masking tape played to the parent over headphones in such cases, the child could use the
parent’s hand to press the button. This was not consistent across participants and influenced the reaction time, therefore reaction time data were not used in the analyses.

In the training phase, six out of eight correct responses, including correct rejections, were required before proceeding to the experimental phase. In the demonstration trials of each experimental block, the experimenter reminded the child to “listen for the new sound” and to “wait until the sound is different” in change trials, whereas the child was reminded with “remember, the lady may trick us” in no-change demonstration trials. During the experimental blocks, the experimenter only interacted to re-direct the child’s attention back to the task between trials. Administration of the LSSP task program took around 20 minutes and was always split half-way through by a break and an alternative task after the third block.

5.3.4. Articulation Accuracy

Background
As outlined in Chapter II, there are links between the development of speech perception and articulation (Burnham, 2003), and between vocabulary and articulation (Schwartz & Leonard, 1982; B. L. Smith et al., 2006; Walley, 1993). To investigate this relationship further, a measure of articulation accuracy at 30, 33, and 36 months was included in the study.

Development
The articulation accuracy task was based upon the Queensland Articulation Test (Kilminster & Laird, 1978), QAT for short.

Table 5.6.: Original wordlist of the Queensland Articulation Test (reproduced with permission from Kilminster & Laird, 1978)

<table>
<thead>
<tr>
<th>Sound</th>
<th>Initial</th>
<th>Medial</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>pig</td>
<td>apple</td>
<td>cup</td>
</tr>
<tr>
<td>b</td>
<td>hall</td>
<td>baby</td>
<td>web</td>
</tr>
<tr>
<td>m</td>
<td>mouse</td>
<td>hammer</td>
<td>drum</td>
</tr>
<tr>
<td>w</td>
<td>window</td>
<td>flower</td>
<td></td>
</tr>
<tr>
<td>f</td>
<td>fish</td>
<td>elephant</td>
<td>knife</td>
</tr>
<tr>
<td>v</td>
<td>vacuum</td>
<td>television</td>
<td>stove</td>
</tr>
<tr>
<td>Ø</td>
<td>thumb</td>
<td>nothing</td>
<td>mouth</td>
</tr>
<tr>
<td>ð</td>
<td>that</td>
<td>feather</td>
<td>smooth</td>
</tr>
<tr>
<td>t</td>
<td>tap</td>
<td>letter</td>
<td>boat</td>
</tr>
<tr>
<td>d</td>
<td>dog</td>
<td>spider</td>
<td>bed</td>
</tr>
<tr>
<td>s</td>
<td>sock</td>
<td>glasses</td>
<td>house</td>
</tr>
<tr>
<td>z</td>
<td>zip</td>
<td>scissors</td>
<td>toes</td>
</tr>
<tr>
<td>l</td>
<td>leaf</td>
<td>lolly</td>
<td>ball</td>
</tr>
<tr>
<td>n</td>
<td>knife</td>
<td>banana</td>
<td>spoon</td>
</tr>
<tr>
<td>c</td>
<td>shoe</td>
<td>washing</td>
<td>fish</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ş</td>
<td>chair</td>
<td>matches</td>
<td>watch</td>
</tr>
<tr>
<td>ç</td>
<td>jam</td>
<td>vegemite</td>
<td>cage</td>
</tr>
<tr>
<td>j</td>
<td>yes</td>
<td>onion</td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>red</td>
<td>carrot</td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>car</td>
<td>bucket</td>
<td>look</td>
</tr>
<tr>
<td>g</td>
<td>gun</td>
<td>tiger</td>
<td>dog</td>
</tr>
<tr>
<td>η</td>
<td>swinging</td>
<td>swing</td>
<td></td>
</tr>
<tr>
<td>h</td>
<td>house</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The QAT has been used to chart and norm the articulation development of Australian English speaking children from three to nine years of age (Kilminster & Laird, 1978). The test involves all 24 consonants in the English language, in all phonotactically legal word positions. The original test items were selected according to the vocabulary knowledge of 3-year-olds (Kilminster & Laird, 1978), close to the target age of this study (Table 5.6.). When establishing norms for the acquisition of Australian English consonants with the QAT, Kilminster and Laird (1978) considered a sound as acquired when 75% of the children pronounced it correctly in all three positions. The articulation accuracy task here served a very different aim, and so the list was reduced to include sounds in initial position only. Since one consonant, /ŋ/, does not exist in word initial position in English, this resulted in 22 items in the articulation accuracy task. Some items were replaced by more appropriate or easier words containing the same consonant, for example “gun” was changed to “girl”, and “yes” was changed to “yellow”. New photographs were shot for each item on the word list (see CD Pictures) and then piloted in a preliminarily test with four 2-year-olds to ensure that the target words could be identified by young children. The results were satisfactory.

The program Irfanview version 3.85 (Skiljan, 2003), free web-based software, was used to display the photos in the articulation accuracy task to the toddlers. Irfanview provides fast loading of the photos from a pre-designed playlist saved as text document (see CD Irfanview Screen). The photos were presented in the same order to each child, to prevent random missing data items in the case that a child was not able to finish the task.

Administration and Scoring
Administering the articulation accuracy task took from five to ten minutes, depending on how much the children liked naming pictures. For the first few photos, the experimenter asked the child “what do you see?” but this became superfluous...
later in testing as the children then usually pre-empted the question with their responses. In case of response refusal, children were asked three times before an item was considered missing. Responses were recorded on a DAT recorder via a stationary condenser microphone (see 5.2.1. Apparatus). Recordings were scored manually after each test session and marked as correct, incorrect, or missing on a response sheet (see CD Scoring sheet). The naming responses were scored as correct if the initial consonant and the following vowel were pronounced correctly as judged by the experimenter listening - if necessary several times - to the recording. Inaccurate responses and response refusal were collapsed together, as children tend to avoid saying words whose pronunciation they struggle with (Schwartz & Leonard, 1982; Schwartz, Leonard, Loeb, & Swanson, 1987). The final articulation accuracy score was percent correct of total trial number.

5.3.5. General Cognitive Functioning

The Stanford Binet V (SB5) (Roid, 2003), is an individually-administered assessment of intelligence and cognitive abilities appropriate for age two onwards. Applying the complete test battery for a full scale measure of global intellectual ability involves an average testing time of 45 to 75 min. The SB5 provides separate measures for verbal and nonverbal cognitive abilities. Here the subtest “Fluid Reasoning-Object Series/Matrices” was selected as a measure of general nonverbal cognitive flexibility. In this, object patterns and matrices require the child to determine underlying rules and/or relationships according to the principles of inductive and/or deductive reasoning.

The test was administered here only at 36 months. Standard testing procedures were followed closely; praise was given for legal responses rather than only for correct responses (Figure 5.13a. and 5.13b.). Administration of the Fluid
Reasoning subtest took about ten minutes, depending on the child’s level of compliance. Testing began with the basal set for 36-month-olds; the stop rule came into effect after a score of zero out of the first three items (which never occurred) or four consecutive errors. To avoid losing any information by converting raw scores to norm brackets, the raw scores were retained for the analyses (see CD Scoring Sheet).

5.4. **Overall Procedure**

Measuring four constructs in six tasks required a total time of two hours, including a 15 minute break after one hour, in which experimenter, parent, and child played ball outside the test chamber. The total testing time within this amounted to 75 min - the rest was playtime, as between each task, the child played in the toy corner of the test room for a few minutes. Despite careful planning to make the test room as toddler-compatible as possible, it was not always possible to execute a task in the ideal position at the desk (as described in 5.2.). To keep the child interested, it was often necessary to change the place where the child completed the task. Alternating between sitting on the floor and sitting in the chair in front of the black Mitsubishi monitor proved to be the best way to sustain the child’s attention. Generally, the PPVT, the Fluid Reasoning, and the second half of the LSSP task were administered on the floor, whereas the articulation, mispronunciation detection, rhyme detection, nonword repetition tasks and the first half of the LSSP task were usually performed with the child sitting at the table. The child also frequently completed tasks on the parent’s lap or in the experimenter’s seat in front of the laptops.

After a short play phase to assist the child settling into the test room situation, the tasks usually began with the articulation task: picture naming worked well as an ice-breaker. After that, the child repeated the nonwords after “Pemmie”, often with the parent’s help. Then the first half of the LSSP task was run, followed by the PPVT on the floor at 33 and 36 months. Before the break, rhyme detection was at least begun. After the break, rhyme detection was completed, followed by
mispronunciation detection. Finally, the second half of the LSSP task was either administered at the table or on the floor, the second solution making the setting more similar to watching TV, as the button box cable extended easily to most points in the test room.
Chapter VI

Results
VI. RESULTS

The results of this study are presented in three parts. First, section 6.1. presents descriptive and inferential statistics for each individual variable in the study. The value for an individual analysis of each variable lies in the fact that many of the variables in this study were new or newly-adapted from existing tests. Second, section 6.2. illustrates the correlations between these variables. Third, the results of regression-based path analyses are presented. These regression analyses were performed to test the vocabulary-driven and phonology-driven models in the basic and augmented versions (see Chapter IV). Section 6.4. concludes by providing a general summary of the results.

6.1. RESULTS FOR INDIVIDUAL TESTS

In this section, results for each variable are considered individually. Sections 6.1.1. through to 6.1.5. present the results pertaining to vocabulary, phonological sensitivity, language-specific speech perception (LSSP), articulation accuracy, and nonverbal cognitive functioning.

The total sample size recorded in this study was relatively large (n = 60). For large sample sizes (n > 50) many inferential statistics are considered robust against non-normality, especially since in this study all variables have equal participant numbers (Hills, 2003), particularly regression analyses (Box & Watson, 1962). Nevertheless, the data were tested for normality distribution with the D’Agostino and Pearson K square normality test (D'Agostino, Belanger, & D'Agostino, 1990; DeCarlo, 1997). This test was chosen because the Kolmogorov-Smirnov and the Shapiro-Wilk tests for normality are too conservative for samples larger than 50 participants (D’Agostino et al., 1990), whereas the D’Agostino and Pearson test neither under- nor overrejects and is adequate in comparison with all major tests for normality (Dufour, Farhat, Gardiol, & Khalaf, 1998). If the D’Agostino and Pearson K square test, calculated in SPSS 14.0. (SPSS Inc., 2005), was significant

32 To run the D’Agostino and Pearson K square normality test on SPSS 14.0., a script needs to be implemented (see CD Normality Test Script).
for any measure ($p < .001$), the measure was considered not to be normally distributed. This was not the case for any measure (see CD Data Screening).

6.1.1. Vocabulary

Two tests were employed to measure vocabulary size: at 30 months, the parental questionnaire OZI was used, followed by the PPVT III at 33 and 36 months (see CD Raw data). Both vocabulary measures were normally distributed. After descriptive analyses of the measures, the vocabulary data of both tests were converted into $z$-scores and used in further analyses (see CD $z$-Scores).

6.1.1.1. Australian English vocabulary inventory OZI

The data for the OZI are illustrated in Figure 6.1. As can be seen, scores are uniformly relatively high ($n = 60; M = 431.32; SD = 99.7$). This may be because the test was used at the upper end of its targeted age range of 16 to 30 months. The Mean Length of Utterance, MLU, was measured across all participants as part of the OZI ($n = 60, M = 5.41, SD = 1.5$).

![OZI scattergram](image_url)

*Figure 6.1.*: The scattergram of the individual OZI vocabulary scores ($n = 60$) shows the data distribution. The mean vocabulary size ($M = 431.32$) is plotted as a horizontal line across the data.
6.1.1.2. Peabody Picture Vocabulary Test III

Mean PPVT scores at 33 and 36 months are shown in Figure 6.2. (n = 60, $M_{33\text{ months}} = 37.02, SD = 12.0$; n = 60, $M_{36\text{ months}} = 46.62, SD = 12.9$). A priori testing for a linear trend with planned within-participant contrasts (alpha = .05) was conducted in PSY 2000, a statistical program from the University of New South Wales (Bird, Hadzi-Pavlovic, & Isaac, 2000), showed a significant vocabulary growth between 33 and 36 months ($F_{(1,59)} = 82.8; p < .01$). Figure 6.2. illustrates the developmental gain in vocabulary (see CD PPVT Development).

![Figure 6.2.](image)

*Figure 6.2.:* The graph shows the linear increase in PPVT vocabulary scores between 33 and 36 months. Error bars indicate standard error ($SE_{33} = 1.6, SE_{36} = 1.7$).

6.1.2. Phonological sensitivity

Phonological sensitivity was measured by three subtests: mispronunciation detection, rhyme detection, and nonword repetition (see CD Raw Data). All phonological sensitivity measures were normally distributed. Each of the subtests was first analysed separately across age (6.1.2.1.) and then a principal component analysis was conducted to uncover commonalities (section 6.1.2.2.) and in readiness for the regression analyses (section 6.3.).
6.1.2.1. The phonological sensitivity subtests

Mispronunciation detection was scored as a hits-minus-false positives index, with every correct detection of a mispronounced word treated as a hit, and mistaking a correctly pronounced word for a mispronunciation scored as a false positive. Rhyme detection was scored as proportion correct of the possible total score, as was nonword repetition. Preliminarily, relative proportions were scored, using the total of attempted items rather than the absolute item total. However, most children finished the tasks, therefore it was then decided to score the few missing items as errors in the rhyme detection and the nonword repetition tasks. For all three tasks, the scores varied between 0 and 1. All three phonological sensitivity subtests were normally distributed. Data were analysed in three separate single factor (age) ANOVAs with planned comparisons of linear and quadratic age trends (alpha = .05). As shown in Table 6.3. and Figure 6.3., results revealed linear age increases for mispronunciation detection, rhyme detection, and nonword repetition (see CD Mispronunciation, Rhyme, and Nonword Development). None revealed quadratic trends.

![Graph showing linear increase over age](image)

*Figure 6.3.*: The graph shows the linear, non-quadratic increase over age in all three phonological sensitivity tasks, mispronunciation detection \( F(1,59) = 73.8; SE_{30} = .03, SE_{33} = .03, SE_{36} = .03 \), rhyme detection \( F(1,59) = 5.19; SE_{30} = .02, SE_{33} = .02, SE_{36} = .02 \) and nonword repetition \( F(1,59) = 9.37; SE_{30} = .03, SE_{33} = .02, SE_{36} = .02 \). Error bars indicate standard error.
Table 6.1:
Means, standard deviations, and F values for mispronunciation detection, rhyme detection, and nonword repetition across age (significant F values are marked in bold).

<table>
<thead>
<tr>
<th></th>
<th>30 m.s</th>
<th>33 m.s</th>
<th>36 m.s</th>
<th>F_{linear}</th>
<th>p_{linear}</th>
<th>F_{quadratic}</th>
<th>p_{quadratic}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mispron.</td>
<td>.27 (.22)</td>
<td>.41 (.26)</td>
<td>.54 (.24)</td>
<td>73.8</td>
<td>&lt;.01</td>
<td>&lt;.001</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Rhyme</td>
<td>.46 (.13)</td>
<td>.51 (.14)</td>
<td>.52 (.18)</td>
<td>5.19</td>
<td>&lt;.05</td>
<td>1.46</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Nonword</td>
<td>.55 (.19)</td>
<td>.62 (.17)</td>
<td>.63 (.16)</td>
<td>9.37</td>
<td>&lt;.05</td>
<td>1.36</td>
<td>&gt;.05</td>
</tr>
</tbody>
</table>

This showed that all three tasks were sensitive to developmental changes, and suggests that there is a linear increase in phonological sensitivity skills across this age bracket.

6.1.2.2. Phonological sensitivity component scores

It is common practice when using phoneme awareness as a variable for further analyses to derive a composite score from its subtests administered to school children (e.g. Hulme et al., 2002). This strategy has also been reported for the concept of phonological sensitivity in preschool children in which confirmatory factor analyses were used to derive a one-factor-model for phonological sensitivity (Anthony et al., 2002). However, the sample size of the present study did not satisfy the requirements of confirmatory factor analyses, therefore principal component analyses (PCA) were used instead. Principal component analysis is the most common factor analysis, predominantly used to reduce data while retaining its variability (Tabachnick & Fidell, 2001). As phonological sensitivity will be used as predictor and criterion in further analyses to test the vocabulary-driven and phonology-driven models, the aim is to compile the three subtests into one or two component scores at 30, 33, and 36 months. Rather than simply reducing the number of variables, principal component analyses assign weighting to each of the components and therefore preserve the individual contribution of each component at 30, 33, and 36 months which minimises the loss of information. The basic prerequisite for principal component analysis is the presence of moderate to strong correlations. As shown in Table 6.2., the three phonological sensitivity measures are moderately correlated with each other.
Table 6.2:
Correlation matrix between the three phonological sensitivity measures (mispronunciation detection, rhyme detection, and nonword repetition) at 30, 33, and 36 months; significant correlations (two-tailed) are marked in bold font. Note how rhyme detection at 30 months did not correlate with any other phonological sensitivity measure.

<table>
<thead>
<tr>
<th></th>
<th>Rhyme 30</th>
<th>Rhyme 33</th>
<th>Rhyme 36</th>
<th>Mispron 30</th>
<th>Mispron 33</th>
<th>Mispron 36</th>
<th>Nonword 30</th>
<th>Nonword 33</th>
<th>Nonword 36</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rhyme 30</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
</tr>
<tr>
<td>Rhyme 33</td>
<td>-0.008</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
</tr>
<tr>
<td>Rhyme 36</td>
<td>0.089</td>
<td>0.387</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
</tr>
<tr>
<td>Mispron 30</td>
<td>-0.021</td>
<td>0.264</td>
<td>0.274</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
</tr>
<tr>
<td>Mispron 33</td>
<td>-0.058</td>
<td>0.156</td>
<td>0.187</td>
<td>0.428</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
</tr>
<tr>
<td>Mispron 36</td>
<td>-0.231</td>
<td>0.134</td>
<td>0.199</td>
<td>0.408</td>
<td>0.561</td>
<td>~</td>
<td>~</td>
<td>~</td>
<td>~</td>
</tr>
<tr>
<td>Nonword 30</td>
<td>0.088</td>
<td>0.045</td>
<td>0.106</td>
<td>0.142</td>
<td>0.486</td>
<td>0.224</td>
<td>~</td>
<td>~</td>
<td>~</td>
</tr>
<tr>
<td>Nonword 33</td>
<td>-0.065</td>
<td>0.014</td>
<td>0.289</td>
<td>0.298</td>
<td>0.489</td>
<td>0.444</td>
<td>0.405</td>
<td>~</td>
<td>~</td>
</tr>
<tr>
<td>Nonword 36</td>
<td>0.005</td>
<td>0.139</td>
<td>0.271</td>
<td>0.271</td>
<td>0.442</td>
<td>0.467</td>
<td>0.361</td>
<td>0.664</td>
<td>~</td>
</tr>
</tbody>
</table>

As seen in Table 6.3, one component (Eigenvalue >1) was extracted for the variable phonological sensitivity at 33 and 36 months, retaining individual contributions weighted in coefficients for each measure (see CD Components 33 and Components 36). Single component scores for phonological sensitivity at 33 and 36 months will be used in the following tests of the vocabulary-driven and phonology-driven models.

Table 6.3:
Matrix of component loadings and component coefficients for the three measures of phonological sensitivity at 30, 33, and 36 months, extracted with principal component analyses: component loadings express the correlations, coefficients assign the weighting that each phonological sensitivity measure received when calculating the component score.

<table>
<thead>
<tr>
<th>Phonological sensitivity</th>
<th>30 months</th>
<th>33 months</th>
<th>36 months</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Loading</td>
<td>Coefficient</td>
<td>Loading</td>
</tr>
<tr>
<td>Mispronunciation detection</td>
<td>.805</td>
<td>.719</td>
<td>.867</td>
</tr>
<tr>
<td>Nonword repetition</td>
<td>.699</td>
<td>.601</td>
<td>.827</td>
</tr>
<tr>
<td>Rhyme detection</td>
<td>-.006</td>
<td>-.44</td>
<td>.285</td>
</tr>
</tbody>
</table>

However, the picture of phonological sensitivity was more complex at 30 months. The principal component analysis returned two components (Eigenvalue >1); mispronunciation detection and nonword repetition emerged as one component while rhyme detection emerged as a separate component (see CD Components 30). As the overall performance level was below chance in the rhyme detection task and the measure did not correlate with any other, it was concluded that rhyme

---

33 Loadings of the rotated component matrix and component coefficients for component 1 only (see in-text argument). The second component, centered on rhyme detection, was not included in further analyses and is therefore not reported here (see CD Components 30).
VI. RESULTS

detection at this age may not measure a language skill, but possibly task attention, and should be omitted as a component for phonological sensitivity. Therefore, at 30 months, only the first component score was retained for the following regression analyses, representing the weighted contribution of mispronunciation detection and nonword repetition while minimising the influence of rhyme detection. One may argue rhyme detection should then also be excluded from phonological sensitivity at 33 and 36 months in order to avoid argumentative problems in result interpretation. To address this issue, all analyses have also been calculated with a consistently composed phonological sensitivity score out of only mispronunciation detection and nonword repetition31. Analyses showed that an impoverished phonological sensitivity score does not impact on the results. Yet, it would be premature to argue that rhyme detection contributes nothing unique and should therefore be excluded. The weighting of rhyme detection within the composite score increases over age as it loads better onto phonological sensitivity, indicating a unique input. In summary, phonological sensitivity seemed to be internally consistent, and one component score per age can be derived to be used in further analyses (see CD Component Scores).

6.1.3. Language-specific speech perception LSSP

The scores for native (N), nonnative consonantal (NN), and nonnative tonal (T) speech contrast discrimination were normally distributed. The discrimination indices were calculated by subtracting the correct rejections over the total number of rejections from hits over the total number of possible hits (see section 5.3.3.). The LSSP measures are N-NN and N-T. N, NN, and T (see CD Raw Data) were tested in three separate single factor (age) ANOVAs for linear and quadratic age trends. As there are no previous data on LSSP measures in toddlerhood, it was difficult to predict how the single measures would behave from 30 to 36 months. However, it was expected that due to the preceding phase of rapid vocabulary acquisition around this age N would remain stable or even increase, while NN and T would show a decline. Consequently, there should be an increase in LSSP scores

31 See CD Two Components 33 and 36 for PCA results and CD Vocabulary-driven Models 33(2) and 36(2) as well as CD Phonology-driven Model 36(2) for regression analysis results.
(N-NN and N-T) between 30 and 36 months. Mean N, NN, and T scores and LSSP measures N-NN and N-T at each age are shown in Table 6.4. N, NN, and T all significantly increased linearly over age (see CD N, NN, T, N-NN, and T Development) and there were no significant quadratic trends.

Table 6.4:
Means, standard deviations, and F values for N, NN, T, N-NN and N-T across age (significant F values are marked in bold).

<table>
<thead>
<tr>
<th></th>
<th>n = 60</th>
<th>30 m.s</th>
<th>33 m.s</th>
<th>36 m.s</th>
<th>F&lt;sub&gt;linear&lt;/sub&gt;</th>
<th>P&lt;sub&gt;linear&lt;/sub&gt;</th>
<th>F&lt;sub&gt;quadratic&lt;/sub&gt;</th>
<th>P&lt;sub&gt;quadratic&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>26 (.29)</td>
<td>.38 (.34)</td>
<td>.48 (.27)</td>
<td>24.0</td>
<td>&lt;.01</td>
<td>.11</td>
<td>&gt;.05</td>
<td></td>
</tr>
<tr>
<td>NN</td>
<td>21 (.23)</td>
<td>.31 (.31)</td>
<td>.40 (.28)</td>
<td>19.0</td>
<td>&lt;.01</td>
<td>.02</td>
<td>&gt;.05</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>25 (.30)</td>
<td>.36 (.32)</td>
<td>.50 (.29)</td>
<td>32.2</td>
<td>&lt;.01</td>
<td>.12</td>
<td>&gt;.05</td>
<td></td>
</tr>
<tr>
<td>N-NN</td>
<td>.05 (.32)</td>
<td>.07 (.32)</td>
<td>.08 (.26)</td>
<td>.29</td>
<td>&gt;.05</td>
<td>.03</td>
<td>&gt;.05</td>
<td></td>
</tr>
<tr>
<td>N-T</td>
<td>.01 (.38)</td>
<td>.02 (.32)</td>
<td>-.02 (.26)</td>
<td>.39</td>
<td>&gt;.05</td>
<td>.27</td>
<td>&gt;.05</td>
<td></td>
</tr>
</tbody>
</table>

These results plus observations during testing suggest that the LSSP task is a difficult one which is increasingly mastered over age. The improvement therefore may signify greater ability to perform the LSSP task, rather than developmental changes in discrimination ability. These results suggest that the potential contribution of LSSP towards a model of the relationship between phonology and vocabulary in toddlerhood may be small. Only the N-NN measure of LSSP was retained for the following analyses of the proposed models (see section 6.3), because it has been successfully used before (Burnham, 2003), and the negative LSSP score N-T at 36 months indicated confused or random responding.

![Figure 6.4.](image)

Figure 6.4.: The graph shows the linear, non-quadratic increase over age in all three LSSP tasks, N (F<sub>1,59</sub> = 24.0; SE<sub>30</sub> = .04, SE<sub>33</sub> = .04, SE<sub>36</sub> = .04), NN (F<sub>1,59</sub> = 19.0; SE<sub>30</sub> = .03, SE<sub>33</sub> = .04, SE<sub>36</sub> = .04) and T (F<sub>1,59</sub> = 32.2; SE<sub>30</sub> = .04, SE<sub>33</sub> = .04, SE<sub>36</sub> = .04). Error bars indicate standard error.
6.1.4. Articulation accuracy

Articulation accuracy was scored as proportion correct out of a total of 22 items. All children completed the task, therefore calculating a relative score out of the number of attempted items was not necessary (see CD Raw Data). Items participants refused to name during the task were scored as errors (for reasoning, see Chapter 5.3.4). Data screening showed normal distribution for articulation accuracy.

The development of articulation accuracy was tested with three separate single factor (age) ANOVAs for the presence of linear and/or quadratic trends with planned within-participants contrasts (alpha = .05). The task was sensitive to the developmental increase in articulation accuracy as there was a linear trend (n = 60, $M_{30\ months} = .52$, $SD = .14$; $M_{33\ months} = .62$, $SD = .13$; $M_{36\ months} = .68$, $SD = .13$; $F_{(1,59)} = .89, p < .01$); see CD Development and Figure 6.5.), but no quadratic trend ($F_{(1,59)} = 1.42, p > .05$).

![Graph showing linear increase in articulation accuracy scores between 30 and 36 months. Error bars indicate standard error ($SE_{30} = .02$, $SE_{33} = .02$, $SE_{36} = .02$).]

Figure 6.5.: The graph shows the linear increase in the articulation accuracy scores between 30 and 36 months. Error bars indicate standard error ($SE_{30} = .02$, $SE_{33} = .02$, $SE_{36} = .02$).

6.1.5. General cognitive functioning

At 36 months, a measure of general cognitive ability was taken. Raw scores of the subtest “Fluid Reasoning-Object Series/Matrices” of the SB5 battery (Roid, 2003)
were used (n = 60, M = 9.5, SD = 2.2; see Figure 6.6. and CD Raw Data). For two male participants out of a total of 60, these scores could not be obtained due to logistic reasons. As their absence is unrelated to the measured variable (Allison, 2001), the missing data were seen as missing completely at random. They were imputed with the mean of male participants in order to preserve potential gender effects, which appeared to be only marginal: girls (n = 30, M = 9.7, SD = 2.0) and boys (n = 28, M = 9.3, SD = 2.3). According to the D’Agostino and Pearson K square test the data were not normally distributed but as the p value for neither skewness nor kurtosis was significant by itself, this could be neglected. The reason for non-normality was likely to lie in two outliers (participants 27 and 44), however, all methods of adjusting outliers, for example to one unit above the next less extreme score, are debated as they reduce variability in the data and promote regression to the mean (Tabachnick & Fidell, 2001).

![Figure 6.6.](image)

**Figure 6.6.** The scattergram shows the raw data distribution of the general cognitive functioning scores (n= 60), used as indication for nonverbal IQ. The Mean (M = 9.5) is represented by the horizontal line.

---

35 In these two cases, cognitive functioning tests could not be taken at the third test session and the families lived too far away to return for such a short task. Imputation of missing data with the variable mean has shortcomings as it reduces variance, yet it is widely practiced (Wood, White, & Thompson, 2004). However, a second analysis without these subjects (listwise exclusion was chosen as an alternative to mean imputation) showed equivalent results (see CD Norm 5B for normality assumptions and CD Vocabulary-driven Model 36(58) and Phonology-driven Model 36(58) for regression analyses of the augmented models). Therefore it can be concluded that the missing proportion of the data was so small that it did not affect results and permitted valid inferences.
6.1.6. **Summary of individual test results**

Many of the tests in this study were newly adapted and had not been used with toddlers before. Children’s performance on these tests over age is consistent, suggesting that they are valid measures of the construct in question. Administration of the OZI at its upper validation range did not skew the data significantly. Each of the three subtests of phonological sensitivity showed a developmental increase in language skills so it appears that their adaptation from tasks used for preschoolers was successful. Apart from rhyme detection at 30 months, all contributed uniquely to the phonological sensitivity component scores. Despite careful considerations and preliminary test sessions, the LSSP task was not sufficiently age-appropriate; however, this may not be due to the underlying validity of the LSSP construct, but rather due to the complexity of the task. Articulation accuracy proved to be a satisfactory measure of the improvement in phonological production. Overall, each variable was measured satisfactorily by the carefully designed and selected tasks, and can continue to be used for the following data analyses.

6.2. **Correlations between the variables**

To examine the relationships between all variables of this study, Pearson’s product-moment correlations were calculated (see Table 6.5.).

There is a weak, but robust relationship between the OZI and the two PPVT measurements. The weakness of the correlation may be due to the relatively high OZI scores. However, the significance of the correlation still indicates that the OZI measures similar aspects of the lexicon within participants as the PPVT, even though the types of lexica differed. Obviously, the retest reliability for the standardised PPVT was high, as shown in the strong correlation of $r = .792$ between the PPVT measurements at 33 and 36 months.
VI. Results

Of specific interest for the hypothesised vocabulary-driven and phonology-driven models is the relationship between phonological sensitivity and vocabulary. The correlation coefficients between these two variables ranged from low to moderate, indicating a close relationship at 30, 33, and 36 months (see Figure 6.7.). The phonological sensitivity component score showed good retest reliability across six months with moderate intercorrelations between 30, 33, and 36 months (see Table 6.5.).
VI. Results

Table 6.5:
Correlation matrix of all variables at 30, 33, and 36 months: The four major variables are listed from left to right. There was a robust relationship between all measures of phonological sensitivity and all vocabulary tests (significant correlations are marked with bold font). As could be expected from the performance of the LSSP scores, there are no meaningful correlations between the two LSSP measures N-NN, or N-T and any other variable. Consistent robust correlations appeared between articulation accuracy, phonological sensitivity, and vocabulary. The variable nonverbal cognitive functioning did not correlate to a significant extent with any other variable at any age.

<table>
<thead>
<tr>
<th>Vocabulary</th>
<th>Phonological Sensitivity</th>
<th>Language-specific speech perception</th>
<th>Articulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPVT 33</td>
<td>0.379</td>
<td>~</td>
<td></td>
</tr>
<tr>
<td>PPVT 36</td>
<td>0.352</td>
<td>0.792</td>
<td>~</td>
</tr>
<tr>
<td>PS 30</td>
<td>0.386</td>
<td>0.421</td>
<td>0.467</td>
</tr>
<tr>
<td>PS 33</td>
<td>0.419</td>
<td>0.432</td>
<td>0.454</td>
</tr>
<tr>
<td>PS 36</td>
<td>0.511</td>
<td>0.384</td>
<td>0.468</td>
</tr>
<tr>
<td>N-NN 30</td>
<td>0.164</td>
<td>0.193</td>
<td>0.107</td>
</tr>
<tr>
<td>N-NN 33</td>
<td>0.210</td>
<td>0.214</td>
<td>0.080</td>
</tr>
<tr>
<td>N-NN 36</td>
<td>0.073</td>
<td>-0.147</td>
<td>-0.030</td>
</tr>
<tr>
<td>N-T 30</td>
<td>0.118</td>
<td>0.231</td>
<td>0.092</td>
</tr>
<tr>
<td>N-T 33</td>
<td>0.060</td>
<td>0.159</td>
<td>0.040</td>
</tr>
<tr>
<td>N-T 36</td>
<td>0.083</td>
<td>-0.015</td>
<td>0.145</td>
</tr>
<tr>
<td>Art. 30</td>
<td>0.521</td>
<td>0.357</td>
<td>0.363</td>
</tr>
<tr>
<td>Art. 33</td>
<td>0.512</td>
<td>0.516</td>
<td>0.505</td>
</tr>
<tr>
<td>Art. 36</td>
<td>0.455</td>
<td>0.220</td>
<td>0.262</td>
</tr>
<tr>
<td>Nonv. IQ</td>
<td>0.159</td>
<td>0.068</td>
<td>0.052</td>
</tr>
</tbody>
</table>
The LSSP measures did not correlate significantly with any other measure, which was to be expected following the analyses of N, NN, and T. Correlations between N-NN and N-T and other variables were weak, apart from N-T at 36 months correlating with phonological sensitivity at 30 and 36 months; and both N-NN and N-T correlated with articulation at 33 months and N-T with articulation at 36 months.

![Figure 6.7: Correlation scatterplots illustrate the relationship between phonological sensitivity and vocabulary at 30, 33, and 36 months. Note how the uniformly high OZI scores distorted the data cloud.](image)

Articulation accuracy showed significant correlations with the vocabulary and with phonological sensitivity. The correlation coefficients between articulation and vocabulary ranged from low to moderate. The articulation accuracy score at 33 months especially correlated consistently with all three vocabulary scores. The correlation coefficients between articulation and phonological sensitivity were mostly moderate, and bordering on strong when correlated within the same age group. Probably due to the consistently high scores in articulation accuracy at 36 months, correlation coefficients were low. Overall, articulation showed an unexpectedly strong and consistent relationship with the two major variables, vocabulary and phonological sensitivity. Nonverbal cognitive functioning did not show any significant relationship with any other variable in this study but was nevertheless included for control purposes in the augmented models.
6.3. **Regression-based Path Analyses**

Typically, hierarchical regressions are used to study the relationship between vocabulary and tests of phonology such as nonword repetition (Bowey, 2001; Metsala, 1999). To examine the relationship across development between vocabulary and phonological sensitivity specifically, and between vocabulary and all other variables in a wider scope, regression-based path analyses were employed here. Two sets of analyses were conducted for each model: the time 1 analysis concerned predicting performance at 33 months from performance at 30 months, and the time 2 analysis concerned predicting performance at 36 months from performance from 33 months.

6.3.1. **Data screening**

The number of participants (n = 60) in this study met the minimal criteria for running regression analyses (Tabachnick & Fidell, 2001). Normality, linearity, and homoscedasticity can be assumed on the basis of the previous individual analyses. No variables to be used in the regression analyses were highly intercorrelated. As time 1 and time 2 were to be predicted from a range of variables in basic and augmented models, two separate data screenings were run: for time 1 (see CD Time 1), predicting vocabulary at 33 months from vocabulary, phonological sensitivity, LSSP, and articulation accuracy at 30 months, the Mahalanobis distance was not significant (critical Chi-Square (4) = 18.467; p < .001), although there was an outlier (case 52) with a residual of almost three standard deviations. For time 2, predicting vocabulary at 36 months from vocabulary, phonological sensitivity, LSSP, and articulation accuracy at 33 months, the Mahalanobis distance was also not significant (see CD Time 2). Therefore, all assumptions for multiple regression analyses are met.

Two sets of models were investigated in regression-based path analyses, basic and augmented models. The **basic vocabulary-driven model** predicts phonological sensitivity from vocabulary size, whereas the **basic phonology-driven model** predicts vocabulary size from phonological sensitivity. The **augmented vocabulary-driven**
model predicts phonological sensitivity from vocabulary size, language-specific speech perception, articulation, and cognitive functioning. The augmented phonology-driven model predicts vocabulary size from phonological sensitivity, language-specific speech perception, articulation, and cognitive functioning. These basic and augmented models were examined in hierarchical multiple regression analyses as set out below.

6.3.2. Basic models

In the basic models, only the variables phonological sensitivity and vocabulary were used to test the predictions of the vocabulary-driven model and its alternative, the phonology-driven model. This resulted in four hierarchical regression analyses, woven together in pairs to regression-based path analyses across time 1 and time 2. To control for autoregression, the variables’ equivalents from the previous time were included first in the regression.

Testing the basic vocabulary-driven model

Results of the hierarchical regression analyses of the vocabulary-driven model at 33 months and at 36 months are shown in Table 6.6. (see CD Vocabulary-driven Model 33 and 36). As can be seen the only significant prediction of phonological sensitivity at 33 months and at 36 months was phonological sensitivity at the earlier age - the effect of autoregression. Vocabulary predicted phonological sensitivity neither at time 1 nor at time 2. Although this test of the propositions of the lexical restructuring hypothesis did not reveal vocabulary as a predictor of phonological sensitivity, such a relationship may still be present at a younger age.
Table 6.6.:
The basic vocabulary-driven model, as put forth by the lexical restructuring hypothesis, predicted phonological sensitivity at 33 and 36 months from vocabulary after controlling for autoregression (significant predictors are marked with bold font).

<table>
<thead>
<tr>
<th>Block</th>
<th>Variables</th>
<th>$R^2$</th>
<th>$R^2$ change</th>
<th>final $\beta^{36}$</th>
<th>$F$ change</th>
<th>sign $F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Phon. Sens. 30</td>
<td>.405</td>
<td>.405</td>
<td>.558</td>
<td>39.5</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2</td>
<td>Vocabulary 30</td>
<td>.440</td>
<td>.035</td>
<td>.204</td>
<td>3.61</td>
<td>.062</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Block</th>
<th>Variables</th>
<th>$R^2$</th>
<th>$R^2$ change</th>
<th>final $\beta$</th>
<th>$F$ change</th>
<th>sign $F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Phon. Sens. 33</td>
<td>.523</td>
<td>.523</td>
<td>.658</td>
<td>63.7</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2</td>
<td>Vocabulary 33</td>
<td>.530</td>
<td>.006</td>
<td>.088</td>
<td>.763</td>
<td>.386</td>
</tr>
</tbody>
</table>

Testing the basic phonology-driven model

Results for the basic phonology-driven model are shown in Table 6.7. (see CD Phonology-driven Model 33 and 36). Both vocabulary and phonological sensitivity at 30 months predicted vocabulary at 33 months significantly ($p < .05$). When predicting vocabulary at 36 months, only the autoregressed variable vocabulary at 33 months was a significant predictor.

Table 6.7.:
The basic phonology-driven model, predicting vocabulary at 33 and 36 months from phonological sensitivity after controlling for autoregression (significant predictors are marked with bold font).

<table>
<thead>
<tr>
<th>Block</th>
<th>Variables</th>
<th>$R^2$</th>
<th>$R^2$ change</th>
<th>final $\beta$</th>
<th>$F$ change</th>
<th>sign $F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Vocabulary 30</td>
<td>.144</td>
<td>.144</td>
<td>.255</td>
<td>9.75</td>
<td>.003</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Block</th>
<th>Variables</th>
<th>$R^2$</th>
<th>$R^2$ change</th>
<th>final $\beta$</th>
<th>$F$ change</th>
<th>sign $F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Vocabulary 33</td>
<td>.628</td>
<td>.628</td>
<td>.733</td>
<td>97.9</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2</td>
<td>Phon. Sens. 33</td>
<td>.643</td>
<td>.15</td>
<td>.138</td>
<td>2.47</td>
<td>.121</td>
</tr>
</tbody>
</table>

Thus there was a phonology-driven effect on vocabulary size at least at 33 months but this weakened at 36 months. This could indicate that phonological drive is indeed the direction of causality and that this may be operational at even younger ages than those tested here.

---

$^{36}$ Although there are arguments to report unstandardised regression coefficients, if the scale of the measures is comparable (Grayson, 2004), standardised regression coefficients are reported here.
6.3.3. Augmented models

In the augmented models, additional to the variables vocabulary and phonological sensitivity, the complete range of language skills from 30 to 36 months was examined, including articulation accuracy, one LSSP measure (N-NN) and nonverbal cognitive functioning. The vocabulary-driven model as well as the phonology-driven model were tested in four hierarchical multiple regression analyses, collated into two regression-based path analyses. Again, to control for autoregression, the variables' equivalents from the previous time were included first. As there was no theoretical basis to guide the order of variable input beyond vocabulary and/or phonological sensitivity, the hierarchical multiple regression analyses were run in two blocks as in the basic models: the first block contained the variable controlling autoregression, and the second block contained the remaining set of variables in addition. As this specific combination of hierarchical and multiple regression analyses could only be conducted separately in SPSS, block 1 both for the vocabulary-driven and phonology-driven models was inserted from the two basic models above. For block 2, standard multiple regression analyses were conducted with all five variables phonological sensitivity, vocabulary, articulation, N-NN, and nonverbal cognitive functioning.

Testing the augmented vocabulary-driven model

Results for the augmented vocabulary-driven model are shown in Table 6.8. (see CD Vocabulary-driven Model 33 and 36). Apart from autoregression effects at time 1 and time 2, phonological sensitivity at 30 predicting phonological sensitivity at 33, and phonological sensitivity at 33 predicting phonological sensitivity at 36 months, only articulation accuracy at 30 months significantly predicted phonological sensitivity at 33 months. These results show that how well toddlers listen to phonological detail in spoken language was predicted by how well they produced initial phonemes in well-known words. This effect was no longer present at time 2 between 33 and 36 months.
Table 6.8.:
The augmented vocabulary-driven model combined hierarchical and multiple regression analyses (significant predictors are marked with bold font).

<table>
<thead>
<tr>
<th>Block</th>
<th>Variables</th>
<th>R²</th>
<th>R² change</th>
<th>β at step</th>
<th>final β</th>
<th>F change</th>
<th>sign β</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Phon. Sens. 30</td>
<td>.405</td>
<td>.405</td>
<td>.558</td>
<td>.558</td>
<td>39.5</td>
<td>.015</td>
</tr>
<tr>
<td></td>
<td>Vocabulary 30</td>
<td>.546</td>
<td>.546</td>
<td>.306</td>
<td>.306</td>
<td>16.5</td>
<td>.015</td>
</tr>
<tr>
<td></td>
<td>Articulation 30</td>
<td>.468</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LSSP: N-NN 30</td>
<td>.052</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Predicting phonological sensitivity at 36 months

<table>
<thead>
<tr>
<th>Block</th>
<th>Variables</th>
<th>R²</th>
<th>R² change</th>
<th>β at step</th>
<th>final β</th>
<th>F change</th>
<th>sign β</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Phon. Sens. 33</td>
<td>.523</td>
<td>.523</td>
<td>.723</td>
<td>.723</td>
<td>63.7</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Vocabulary 33</td>
<td>.536</td>
<td>.536</td>
<td>.629</td>
<td>.629</td>
<td>12.5</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Articulation 33</td>
<td>.104</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LSSP: N-NN 33</td>
<td>-.020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nonverbal IQ 36</td>
<td>.027</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Testing the augmented phonology-driven model

Results for the augmented phonology-driven model are shown in Table 6.9. (see CD Phonology-driven Model 33 and 36). Phonological sensitivity was a strong predictor of vocabulary size at 33 months. This effect was already present in the basic phonology-driven model above (see Table 6.7.) and so is robust despite the addition of other language variables. This suggests that the better toddlers listen to phonological detail in speech, the more words they know. This relationship between the two variables ceased to be significant at time 2, suggesting an end to a period in which vocabulary growth benefits from attention to phonological detail in speech.

Table 6.9.:
The augmented phonology-driven model combined hierarchical and multiple regression analyses (significant predictors are marked with bold font).

<table>
<thead>
<tr>
<th>Block</th>
<th>Variables</th>
<th>R²</th>
<th>R² change</th>
<th>β at step</th>
<th>final β</th>
<th>F change</th>
<th>sign β</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Vocabulary 30</td>
<td>.144</td>
<td>.144</td>
<td>.379</td>
<td>.379</td>
<td>9.75</td>
<td>.003</td>
</tr>
<tr>
<td>2</td>
<td>Phon. Sens. 30</td>
<td>.258</td>
<td>.258</td>
<td>.216</td>
<td>.216</td>
<td>4.77</td>
<td>.125</td>
</tr>
<tr>
<td></td>
<td>Articulation 30</td>
<td>.022</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LSSP: N-NN 30</td>
<td>.161</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Predicting vocabulary at 36 months

<table>
<thead>
<tr>
<th>Block</th>
<th>Variables</th>
<th>R²</th>
<th>R² change</th>
<th>β at step</th>
<th>final β</th>
<th>F change</th>
<th>sign β</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Vocabulary 33</td>
<td>.628</td>
<td>.628</td>
<td>.792</td>
<td>.792</td>
<td>97.9</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2</td>
<td>Vocabulary 33</td>
<td>.658</td>
<td>.658</td>
<td>.725</td>
<td>.725</td>
<td>20.8</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>Phon. Sens. 33</td>
<td></td>
<td></td>
<td>.075</td>
<td>.075</td>
<td></td>
<td>.494</td>
</tr>
<tr>
<td></td>
<td>Articulation 33</td>
<td></td>
<td></td>
<td>.115</td>
<td>.115</td>
<td></td>
<td>.336</td>
</tr>
<tr>
<td></td>
<td>LSSP: N-NN 33</td>
<td></td>
<td></td>
<td>-.116</td>
<td>-.116</td>
<td></td>
<td>.179</td>
</tr>
<tr>
<td></td>
<td>Nonverbal IQ 36</td>
<td></td>
<td></td>
<td>-.001</td>
<td>-.001</td>
<td></td>
<td>.994</td>
</tr>
</tbody>
</table>

6.4. **SUMMARY OF RESULTS**

For the first time language development in 30- to 36-month-olds has been studied comprehensively on multiple dimensions: phonological sensitivity, vocabulary size, articulation accuracy and language-specific speech perception. As the individual analyses revealed, all variables apart from the LSSP task show appropriate age-related modulation and appeared to capture the intended language skills. Thus, in general, useful instruments have been designed and implemented for this age group. Regression-based path-analyses over age showed that phonological sensitivity was a strong predictor of vocabulary size at 33 months, whereas articulation accuracy predicted phonological sensitivity. Thus, the link between attention to phonological detail in speech and the lexicon proposed by the lexical restructuring hypothesis does occur, but in exactly the opposite direction. In addition, articulation, not vocabulary, best predicted phonological sensitivity. In Chapter VII, the implications of these findings will be discussed further, especially in the light of the lexical restructuring hypothesis.
Chapter VII

Discussion
7.1. **Vocabulary-driven or Phonology-driven Model?**

In this section, the direction of the relationship between vocabulary and attention to phonological detail in speech will be discussed by contrasting the results for the two alternative models, the vocabulary-driven model, and the phonology-driven model. First, the results from the basic models will be explained in light of the lexical restructuring hypothesis and the WRAPSA model. Then, the relationship between vocabulary and attention to phonological detail in speech will be discussed in the augmented models which include the additional factors of language-specific speech perception and articulation.

7.1.1. **Discussion of the Basic Models**

In regression-based path analyses predicting from 30 months to 33 months and from 33 months to 36 months, the two versions of the basic models tested whether the variables, vocabulary and phonological sensitivity, were predictors for the criteria - either phonological sensitivity or vocabulary. Two different basic models were entertained: the vocabulary-driven model and the phonology-driven model.

The *vocabulary-driven model* corresponds to the lexical restructuring model which hypothesises that vocabulary size predicts attention to phonological detail in speech. This hypothesis was not supported by the data, neither from 30 to 33 months, nor from 33 to 36 months. Vocabulary size did not predict phonological sensitivity when controlling for autoregression in the basic vocabulary-driven model. This does not necessarily mean that there is no relationship between vocabulary size and phonological sensitivity. When predicting phonological sensitivity at 33 months from vocabulary size at 30 months, the significance level of $F$ change ($p = .062$) shows a tendency towards significance, probably inhibited by the strong autoregression of phonological sensitivity at 30 months. Nevertheless these results do suggest, however, that the direction of the relationship between vocabulary and phonological detail in the lexicon may not be optimally described by the lexical restructuring model. Jusczyk (1994a) suggested
that the restructuring of lexical items on a phonetic level may be driven by infants’ attention to how sound patterns are distributed in the native language input, rather than as a result of the need to distinguish the meaning of different lexical entries from one another. So it may not be a certain lexicon size of 50 to 100 words around the vocabulary spurt that provides the imperative for a child to reorganise lexical entries and encode them on a more specific basis as implied by the lexical restructuring model (Metsala & Walley, 1998; Walley, 1993). Rather the opposite may be the case. This was tested in the basic phonology-driven model regression analyses.

The phonology-driven model tested the hypothesis that phonological sensitivity predicts vocabulary size. When predicting vocabulary at 33 months, phonological sensitivity at 30 months contributed significantly - clearly above the predictive value of the autoregressed variable. This is clear support for the hypothesis that attention to phonological detail in speech influences vocabulary size. Over a period of three months, phonological sensitivity affects the size of the lexicon. This finding is consistent with the WRAPSA model (P.W. Jusczyk, 1993, 1997). However, the effect weakened in the next regression-based path analysis: when predicting vocabulary at 36 months from phonological sensitivity at 33 months, phonological sensitivity at 33 months did not significantly predict vocabulary size at 36 months. Thus the relationship described by the phonology-driven model seems to apply especially for younger children and may phase out before the third birthday. The link between phonology and vocabulary may well be stronger in a smaller and more restricted lexicon. It is possible that lexicon size functions as a mirror for lexicon-internal mechanisms but loses its reflecting qualities with continuing vocabulary growth. In light of this the finding that the phonology-driven model is significant from 30 to 33 months but not from 33 to 36 months suggests that it would be of interest to investigate this effect in even younger children than those tested here.
7.1.2. Discussion of the augmented models

Both the vocabulary-driven and the phonology-driven model were also tested in augmented versions of the models which include the complete range of language skills tested.

In the augmented model of the *vocabulary-driven* hypothesis, the prediction of the lexical restructuring model again was not supported. Neither from 30 to 33 months nor from 33 to 36 months did vocabulary predict phonological sensitivity. This allows the conclusion that the lexical restructuring hypothesis is not supported here regarding the influence of lexicon size on attention to phonological detail. However, these results are neutral with regard to the main claim of the lexical restructuring model that lexical representations change from a holistic format to a segment-based encoding as age or vocabulary increases (Walley et al., 2003).

Another interesting and unexpected effect appeared in the augmented vocabulary-driven model: articulation predicted phonological sensitivity from 30 to 33 months. This effect weakened and lost significance in the 33 to 36 months analysis. The role of articulation as a predictor for phonological sensitivity is discussed in section 7.3.

In the augmented model of the *phonology-driven* hypothesis, the significance of phonological sensitivity as a predictor of vocabulary size from 30 to 33 months is confirmed. Even when the factors, articulation and language-specific speech perception, are added phonological sensitivity still predicts vocabulary size. This is especially remarkable as the multiple regression analysis used in the augmented model did not pre-assign any model-guided weighting on any of the factors which could have strengthened the position of phonological sensitivity. In fact, including additional factors in the augmented model did not affect the β coefficient (β = .324) compared to the basic model (β = .322). Analogous to the basic model, from 33 to 36 months the effect of phonology on vocabulary weakened and lost
significance. It can be concluded that the phonology-driven model for vocabulary operates between 30 and 33 months and possibly also in younger children.

7.2. CRITICAL REVIEW OF INDIVIDUAL MEASURES

Since this study also had a strong exploratory element, the newly-developed and newly-adapted individual measures will be critically reviewed in the following section, especially in the light of their applicability to younger age groups than those tested here (see also section 7.5.).

7.2.1. Vocabulary

Two tests were used in the present study to measure vocabulary, the OZI and the PPVT III. The OZI is the first specifically Australian English vocabulary measure and presents a viable alternative to the use of American English vocabulary questionnaires in Australia. OZI data for vocabulary size in Australian children from 16 to 30 months and for the age of acquisition of all vocabulary questionnaire items were collected here and are a functional and valuable by-product of this thesis. Indeed norms will continue to be collected and the resulting norms will provide a significant contribution both to the clinical and scientific community beyond the scope of this project.

The choice to use the OZI at the upper end of its validation range (16 to 30 months) rather than using the PPVT III at the lower end of its validation range (30 months to above 90 years) in this study posed conceptual problems and required conversion of the vocabulary scores into z-scores. Nevertheless this choice is justified by the fact that more than half of the children in the sample did not meet the basal set rule of the PPVT at 33 months. This figure is likely to have been even greater if the PPVT had been used at 30 months and would have prevented the use of scaled reference scores from the PPVT manual.

The OZI is a sensitive measure of productive vocabulary, and is efficient and cost-effective to use. As an offshoot of this thesis, MARCS Auditory Laboratories
intends to create an online version of the questionnaire in the near future to facilitate norm collection. In order to provide continuity in measuring vocabulary development from its onset, the MARCS Babylab currently administers the OZI to children under 16 months to extend the validation range of the OZI at its lower end and collect norms.

7.2.2. Phonological Sensitivity

The phonological sensitivity measure proved to be a robust and sensitive tool to capture children’s attention to phonological detail in speech, uniting tasks in its three components that have previously been successful in measuring phoneme awareness in preschool and school children. Deriving phonological sensitivity from three different tasks provided a global, naturalistic, and robust measure for the attention to phonological detail in speech. The component that showed the steepest linear improvement over age was the mispronunciation detection task. Thus if studies similar to the one conducted here were extended to younger children in future studies (see 7.5.), it is possible that the phonological sensitivity measure could be reduced to include only mispronunciation detection. This adaptation would be particularly apt given that the nonword repetition involves speech production measurement and rhyme detection was already very difficult for the 30- to 36-month-olds. Therefore both these measures would be very difficult to adapt to younger children.

7.2.3. Language-specific speech perception

Language-specific speech perception (LSSP) was intended to be a complementary measure to phonological sensitivity in the speech perception tier. Its three components, native consonant, nonnative consonant, and nonnative tone discrimination, all showed linear developmental increases. This was unexpected in comparison with the previous application of the LSSP measure which showed a decline in nonnative speech discrimination during reading acquisition (Burnham, 2003). A decline in nonnative speech discrimination abilities could be expected during developmental periods that intensify specialisation in the native language
such as reading acquisition or vocabulary acquisition. To obtain the LSSP index, scores for nonnative consonant discrimination (NN) are subtracted from native consonant discrimination scores (N), with a high N-NN indicating a strong orientation towards the native language and a low N-NN score demonstrating approximately equivalent performance in native and non-native perception, thus less specialisation in the native language. Negative N-NN scores reflect greater attention to the nonnative language or confused responding. In the present study, however, NN neither declined, nor even remained stable across age, it actually improved, as did N scores, resulting in relatively stable N-NN scores over age. This was unexpected, assuming that rapid vocabulary acquisition would represent a period of intensified specialisation in the native language (Burnham, 2003). Furthermore, the tone scores (T) showed a strong increase over age. It is a possibility that tone, being both novel and salient to the children, captured their interest more than native or other nonnative contrasts. However, this seems unlikely in 30- to 36-month-olds, given the specialisation in the native language they already show in their lexicon size and in their articulation skills. Instead, the fact that all scores, N, NN, and T, increased over age may give a clue to what is happening here.

The LSSP task in general may have additional task-inherent and age-sensitive components such as response consistency, attention span, and understanding of reward contingency. Especially in young children limited memory capacities could lead to confused response behaviour as observed in some of the participants in this study (see Packiam Alloway, Gathercole, & Pickering, 2006). These task-inherent components presumably improve over age and could thus be responsible for the developmental increase in N, NN, and T. In addition, the 300 ms interstimulus interval could have been too short to ensure linguistic processing of the stimuli (Werker & Logan, 1985; Werker & Tees, 1984b). In preliminary test trials, 300 ms was the longest interval duration during which 24- and 30-month-old children did not visibly lose attention within a trial despite a maximum trial length of eight sound exemplar repetitions before a change occurred. Thus it is quite possible that there was acoustic rather than linguistic processing of all of the sound stimuli,
resulting in similar scores for all. Perhaps this was even more the case for tone: indeed, the tone scores at 36 months were higher than N and NN scores.

When N-NN was included as a LSSP measure in the augmented vocabulary-driven and phonology-driven model, it is not surprising that it did not contribute to the prediction of either phonological sensitivity or vocabulary size. However, the same LSSP task set-up has been implemented with preschool and school children, yielding meaningful results (Erdener, 2007). These task-related issues could be addressed and resolved in an adaptation of the LSSP task for younger children, using an eye tracking device instead of a response button. Analogous to how the conditioned head-turn paradigm has been successfully used with 5- to 18-month-olds (Polka, Jusczyk, & Rvachew, 1995; Werker, Polka, & Pegg, 1997), detection of sound change in speech contrasts could be indicated by eye movement rather than by head-turns. In comparison to the complex response of pressing a button, eye movement control is gained early in development as studies with 3- and 4-month-olds show (Griffin, 2004; Lecuyer, Berthereau, Taieb, & Tardif, 2004) and does not require the same level of cognitive and memory involvement. The children could be trained to expect a reward trailer in combination with sound change at a particular position on screen, and the eye tracker could record the change in looking direction with high accuracy37. Using the eye tracker paradigm offers the possibility to separate perceptual responses from motor responses in the LSSP task.

7.2.4. Articulation accuracy

The articulation measure showed a linear improvement over age and was normally distributed. It correlated strongly with the vocabulary measures and phonological sensitivity. This was not surprising as strong links between phonological detail in

---

37 The latest generation of eye trackers does not even require the mounting of any head equipment or interference-sensitive magnetic transmitters reporting the head position in space. They use near infra-red light-emitting diodes to create non-harmful reflection patterns in the eyes. These corneal reflection patterns are identified and collected by the camera via image-processing algorithms. While compensating for head movements, the three-dimensional position in space is calculated for each eye ball to identify the gaze point on screen, reporting response times and gaze position for both eyes with high accuracy (Tobii Technology, 2006).
speech perception and phonological detail in word production were presented and reviewed in Chapter II. It was, however, surprising that articulation predicted phonological sensitivity from 30 to 33 months in the augmented vocabulary-driven model, but not in the phonology-driven model. Articulation lost significance in the 33 to 36 months prediction. Compared to the basic vocabulary-driven model from 30 to 33 months, the $\beta$ for vocabulary lost three quarters of its predictive quality when articulation is added ($\beta_{\text{basic}} = .204; \beta_{\text{augmented}} = .049$). The following comparison of the regression equations for the basic vocabulary-driven model and the augmented vocabulary-driven model in the 30 to 33 months prediction shows this.

*Basic vocabulary-driven model:*

$$\text{Phon. Sens.} = \text{Phon. Sens.} \times 0.558 + \text{Vocab.} \times 0.204 + \text{constant a}$$

*Augmented vocabulary-driven model:*

$$\text{Phon. Sens.} = \text{Phon. Sens.} \times 0.360 + \text{Vocab.} \times 0.049 + \text{Articulation} \times 0.468 + \text{LSSP} \times 0.052 + \text{constant a}$$

Perhaps it is the case that the influence of vocabulary on phonological sensitivity is filtered through production phonology. The possibility that articulation is involved in perceptual phonology and the lexicon will be discussed in the following section.

7.3. **IN A NUTSHELL: THE ROLE OF ARTICULATION**

Articulation accuracy may offer a window into the developing phonological system. Inaccurately articulated words can still be based upon adult-like lexical representations including phonological specifications (Storkel, 2004b). In the development of articulation accuracy, it has been found that phonological cues of the ambient language affect preschool children’s learning of nonword nouns (Storkel, 2001). Thus they learned words more quickly when they contained high phonotactic probabilities which demonstrates a facilitating effect of phonotactic probability in lexical acquisition. A similar effect has been found for verbs (Storkel, 2003). In both studies, the effect of phonotactic probabilities was especially present in a picture naming condition in which the children produced
the previously learned nonwords. They seem to assume that the most likely label for a novel object or action is a common rather than a rare sound sequence (Storkel, 2003). As word-learning studies such as these demonstrate, the relationship between articulation and perceptual phonology appears to develop via the lexicon. This link between the lexicon and productive phonology appears to be operational after the 50-word stage (Storkel & Morissette, 2002).

Storkel and Morissette (2002) describe the feedback loop between the lexicon and phonology as follows: it is unlikely that children form an adult-like phonological representation of a word containing unknown sounds when they hear it for the first time. In combination with the lexical representation and via production of the underspecified sound shape of the word, the phonological representation becomes refined which in turn influences lexical representations. The relationship between phonology and the lexicon is described as bidirectional: the lexicon influences phonological acquisition and phonology influences lexical acquisition (Storkel & Morissette, 2002). For example, children with functional phonological delay display difficulties in creating lexical representations for words with high phonotactic probabilities (Storkel, 2004b).

Lexicon size does indeed appear to influence productive phonology. B. L. Smith, McGregor, and Demille (2006) measured children’s phonological abilities via consonant inventory size, number of correctly produced consonants, and general phonological processes. They found that the productive phonology of 24-month-olds with advanced vocabularies corresponded to 30-month-olds with similar lexicon size, rather than to 24-month-olds with smaller vocabularies. Thus, children with advanced vocabularies showed advanced phonological abilities in their speech production compared to children with average-sized vocabularies (B. L. Smith et al., 2006).

---

38 Vocabularies were classed in this study as advanced and the speaker as precocious if scores above the 85th percentile on the MacArthur CDI were obtained (Fenson et al., 1993).
Articulation also influences the perception of phonological detail in speech. A group of 3-year-olds who failed to produce the phoneme /r/ accurately in all instances tested also showed perceptual insensitivity to this particular phoneme, but had no difficulties in performing phoneme sensitivity tasks with control phonemes (Thomas & Sénéchal, 1998). On the other hand, the group who produced /r/ accurately in all items had no difficulty with /r/ in phoneme sensitivity tasks. These two groups were assessed over a period of five years. Articulation quality of the phoneme /r/ at age three predicted phoneme sensitivity from three to five years after controlling for vocabulary and letter knowledge and still explained unique variance in phoneme sensitivity at age eight (Thomas & Sénéchal, 2004).

The finding that articulation accuracy plays an important role in the development of phonological sensitivity in the present study confirms and extends these earlier findings.

7.4. The Complete Picture

The present study offers the first test of the lexical restructuring hypothesis that is not restricted to a particular task or effect such as the position effect (Bowey & Hirakis, 2006). Phonological sensitivity, measured as an amalgam of three different tasks, predicts vocabulary size across three months of rapid language development. Three months is a long period in a 2- to 3-year-old child’s life, especially as linguistic development is progressing so rapidly at this stage, so the fact that the prediction holds across these ages signifies the strength of this relationship.

Vocabulary size has been viewed as the force that drives the need for segmental restructuring of the lexicon due to competition mechanisms (Metsala, 1997a). Once vocabulary reaches the size of 50 to 100 entries - Fernald, Swingley, and Pinot’s work (2001) suggests this size more precisely to be around 60 lexical entries - words are stored on the basis of segments. As this coincides with the lexicon size at which
the vocabulary spurt starts, the vocabulary spurt has been suspected to initiate lexical restructuring (Walley, 1993). However, if as Ganger and Brent (2004) suggest, the vocabulary spurt is an artefact, would this be the downfall of the LRM? If there is, following Ganger and Brent’s redefinition, no exponential spurt, just rapid vocabulary increase, then there is no qualitative conceptual change in lexical acquisition (Nazzi & Bertoncini, 2003), just regular developmental growth. However, if lexical restructuring is not initiated by vocabulary size - as the results of this study suggest - the notion of a vocabulary spurt is theoretically unnecessary.

The current results are consistent with the following explanation which does not include the vocabulary spurt as a necessary component. The lexical acquisition task of word-object-mappings becomes easier between 14 and 20 months, and the capacity to attend to phonological detail in spoken language is gradually expanded. Attending to phonological details in words during word meaning acquisition initiates segment-based encoding of lexical entries which in turn facilitates further lexical acquisition. Instead of a qualitative change in lexical acquisition, children continue to refine their phonological sensitivity skills during word learning, and as they improve, their vocabulary grows more quickly. However, it appears that the influence of phonological sensitivity skills phases out before the third birthday and can be assumed to be strongest during toddlerhood in the second year of life.

Studies in the light of the LRM indicate a relationship between lexicon-internal mechanisms such as age of acquisition, neighbourhood density, and word frequency and the representation of phonological detail. For example, findings on lexical neighbourhood density suggest such a lexicon-internal mechanism that restructures entries and/or guides their acquisition. Words in denser lexical neighbourhoods are more likely to be restructured earlier than words with only few lexical neighbours (Storkel, 2002, 2004a). Storkel found in 3- and 5-year-old preschool children phoneme-based segmentation for words in dense neighbourhoods, but manner class organisation for words in sparse
neighbourhoods (Storkel, 2002). This effect can be attributed to the greater phonotactic probability of words in dense neighbourhoods (see section 3.1.2.2.), as the two are correlated (Vitevich et al., 1999). On the other hand, neighbourhood density can be interpreted as an operator independent of phonotactic probability which facilitates lexical restructuring from within the lexicon (Storkel, 2004a). This poses the chicken-and-egg question of the direction of the relationship between phonological detail and the lexicon: is the lexicon the motor of attention to phonological detail in speech or is phonological sensitivity the motor of lexical growth?

Smith, McGregor, and Demille (2006) offer a solution to this dilemma. They discuss a third possibility: the relationship between phonological sensitivity and the lexicon could be bidirectional or even cyclical. The more young children listen for phonological details in words, the more segment-based their lexical entries become. The more segmented their lexical entries are, the better they attend to phonological details in words, and the faster they acquire new words. The more words children know, the more they need to listen for phonological details in words to be able to encode and store them efficiently and accessibly in the lexicon.

7.5. Implications for future research

With respect to current research into word learning and lexicon content in infants, Swingley points out that there are three main avenues along which the current knowledge status would advance through refined findings (Swingley, 2005): firstly, analysis of infant-directed speech input could provide the basis for the way infants perceive, segment, analyse, and process the speech signal, such processing being in itself the second avenue; and thirdly, computational modelling of infant language acquisition could integrate infant-directed speech input and infant language knowledge (Swingley, 2005).

In order to model language acquisition, models must be proposed that can be readily tested. A cyclical or bidirectional relationship between attention to
phonological detail in speech and vocabulary can only be truly tested in a longitudinal design using structural equation modelling, but the participant numbers required for this statistical method with a similar number of factors as in the present study exceeds 100 (Marsh, Hau, Balla, & Grayson, 1998). This would be quite an ambitious venture, especially if the developmental follow-ups were conducted at shorter time intervals than three months. However, benefiting from the indicative results of the present study, the measures could be modified and applied in specific experimental paradigms, with individual measurement time shortened.

Questions for further research aiming to test the lexical restructuring hypothesis are:

1. What is the nature of the relationship between attention to phonological detail in speech and the lexicon in 14- to 20-month-olds? What direction of the lexicon-phonology relationship will be indicated if this age group is studied longitudinally?

2. Is there evidence for a vocabulary spurt if strict definitions and calculations are applied? A consistent developmental chart of vocabulary development in 14- to 20-month-olds could settle the debate about the presence of a vocabulary spurt.

3. How would the language-specific speech perception measure operate if it were adapted to an eye tracking paradigm? It has the potential to complement the phonological sensitivity measure well, but this remains to be established.

The present study is the first empirical test of the lexical restructuring hypothesis in direct relation to vocabulary size, moving one step closer to identifying the true nature of the relationship between attention to phonological detail in speech and
the lexicon. Further studies addressing the above questions are required to close in on the directional issue in the lexicon-phonology relationship.
REFERENCES


REFERENCES


REFERENCES


REFERENCES


References


REFERENCES


Farrington, D. P., Gallagher, B., Morley, L., St.Ledger, R. J., & West, D. J. (1993). Minimizing attrition in longitudinal research: Methods of tracing


REFERENCES


REFERENCES


REFERENCES


REFERENCES


186


REFERENCES


Appendix CD is available for viewing at the University of Western Sydney Library