The Acquisition of Contrast:
A Longitudinal Investigation of Initial s-Plosive Cluster Development in Swedish Children

Fredrik Karlsson

Umeå University
Umeå 2006
Abstract

This Thesis explores the development of word-initial s+plosive consonant clusters in the speech of Swedish children between the ages of 1:6 and 4:6. Development in the word-initial consonant clusters is viewed as being determined by 1) the children’s ability to articulate the target sequence of consonants, 2) the level of understanding of which acoustic features in the adult model production are significant for the signalling of the intended distinction, and 3) the children’s ability to apply established production patterns only to productions where the acquired feature agrees with the adult target, to achieve a contrast between rival output forms. This Thesis employs a method where output forms are contrasted with attempted productions of potential homonym target words. Thus, development is quantified as an increase in the manifestations of phonetic features where it agrees with the adult norm, coupled by a decrease in the same feature in output forms where it is inappropriate according to the specifications of the phonological system of the ambient language. Acoustic investigations of cues of voicing, aspiration, place of articulation and syllable onset complexity, and auditory investigations of place, manner and syllable onset complexity were conducted. The Thesis has four outcomes. One, a description of the perceptual quality of the productions in terms of place, manner, voicing and syllable onset complexity is presented. Two, a developmental sequence of stable acquisition of these features is proposed; manner is shown to be acquired first, followed by syllable onset complexity and place of articulation. Evidence is provided that the voiced/aspirated distinction is still being acquired at the end of the investigated age period. Three, the developmental use of acoustic cues of place and voicing are described. Voice Onset Time and Spectral Skewness are shown to be used by children in order to increase the likeness to the adult target in terms of voicing and place of articulation. Aspiration Amplitude is shown to be used as an auxiliary cue to Voice Onset Time. The place cues Spectral Tilt Change, F2, Spectral Mean and Spectral Variance were shown to be used in order to refine already produced consonants rather than approach the adult target model. Four, the Thesis provides evidence of periods of confusions in the output of children. With the reductions of these patterns of confusion, evidence is provided of children’s re-organisation of their internal representation of the consonant to be produced.

KEYWORDS: phonological contrast, speech development, place and manner of articulation, aspiration, structural complexity, acoustic cues, homonym production
Sammanfattning


För det fjärde ges i avhandlingen evidens för perioder då barnens produktion uppvisar förväxlingar vad avser distributionen hos vissa fonetiska drag. Minskat förekomster av sådana förväxlingar visar på en omstrukturering av barnens underliggande representation för den konsonant som ska produceras.
First, I would like to thank my two supervisors, Kirk Sullivan and Peter Czigler. Kirk Sullivan provided excellent comments on the various draft versions of this Thesis and also did not hesitate to tell me when the quality of the text I produced did not match the quality of the idea I was trying to describe. There have been many discussions on how to write over many cups of coffee, and I have tried to learn from all of your insights. Thank you Kirk.

My second supervisor, Peter Czigler, managed to find many errors and logical lacunae in the draft versions of this thesis that neither me nor Kirk had been able to spot. Without your comments, Peter, I would probably be the only one to be able to understand the text in this thesis.

Professor Jan van Doorn gave valuable comments on the data collection method at the onset of the investigation and continuously sent me suggestions on articles to read. Thank you also, Jan, for checking on my thesis progress every time we met. I know that these enquiries reminded me of my deadline and helped keep up the work pace needed to meet it.

I would also like to thank other colleagues at the department of Philosophy and Linguistics who helped me find my way towards the resulting thesis. Görel Sandström read a draft version of what was to become Chapter 4, gave suggestions on how to restructure it in order to increase readability and suggested enhancements to the visualisation of the data set investigated. Erik Eriksson read and commented on the abstract of this thesis and many of the conference papers on which the discussion in this thesis are based. Thank you also, Erik, for the many chats during too long coffee breaks which made me think of other things than the thesis. With Felix Schaeffler, my fellow student for most of my doctoral studies, I also had many discussions over afternoon coffee and many memorable evenings during conferences.

This thesis would, however, not have been completed if it were not for the support offered by my family and friends. Johanna Kjellgren, the mother of my beautiful daughter, provided formidable support during the thesis writing, especially in the last hectic months of thesis writing. This thesis would simply not have been completed without your help. Family members and friends from outside of the department ensured me access to a sphere (much needed at times) where I was NOT asked about the progress of my thesis. Thank you all for reminding me that life exists also outside of the University.
Finally, I would of course like to thank my daughter Alva who, without her knowing it, participated in a series of undocumented case studies in data-driven changes in the perception of aspiration. You have reminded me many times about the good things in life by your seemingly endless stream of smiles and laughter.
# Table of Contents

1 Developing phonological contrasts 1
   1.1 From perception to production through stages of potential homonymy . 3
   1.2 Development of syllable and monosyllabic-word production ability . . 4
   1.3 Methodological Issues .................................................. 5
   1.4 Issues addressed in this Thesis ........................................ 6
   1.5 Structure of the Thesis .................................................. 7

I Background and methodology 9

2 Aspects of s+Plosive clusters 11
   2.1 Processes in s+plosives productions by children .................. 11
   2.2 Voicing in plosives ...................................................... 13
       2.2.1 Voicing in adult-directed speech ................................. 13
       2.2.2 Voicing in speech directed to children .................... 15
       2.2.3 The child’s navigation in the multidimensional space of adult
            voicing ............................................................ 16
       2.2.4 Progression from a unimodal to a bimodal VOT distribution . 16
       2.2.5 Prevoicing production in development .......................... 17
   2.3 Place and manner of articulation ..................................... 17
       2.3.1 Perceptual stability of phonological features in the ambient
            language ............................................................. 18
       2.3.2 Progression in phonological features at the phoneme level ... 19
       2.3.3 Acoustic correlates of PLACE OF ARTICULATION in plosives .. 20
       2.3.4 Acoustic correlates of PLACE OF ARTICULATION in fricatives .. 21
   2.4 Conclusion ............................................................... 22

3 Data collection and the UC³ Corpus 23
   3.1 Participants ............................................................... 24
       3.1.1 Criteria ............................................................. 24
       3.1.2 Participant recruitment ........................................... 24
       3.1.3 Pre- and post investigation briefing ............................ 25
Table of Contents

3.1.4 Participant identity encoding process .................................. 25
3.1.5 Treatment of participant background information ..................... 26
3.2 Data collection ........................................................................ 26
3.2.1 Recording procedure ............................................................. 26
3.2.2 Data inclusion criteria ............................................................ 27
3.3 Markup .................................................................................. 28
3.3.1 Markup procedure ................................................................. 28
3.3.2 Markup configuration ............................................................. 28
3.3.3 Markup criteria ...................................................................... 30
3.4 Data organisation and data extraction method ............................... 34
3.5 The UC3 Corpus ..................................................................... 34
3.6 Discussion ............................................................................ 35

II Voicing contrast in production .................................................. 37

4 Quantifying Development of Phonological Voicing using Measurements of Voice Onset Time ..................................................... 39
4.1 Presentation of the VOT data set ................................................. 39
4.2 Data categorisation scheme ....................................................... 40
4.2.1 Perceptually based VOT CATEGORY ....................................... 42
4.2.2 Division by AGE CATEGORY .................................................. 43
4.3 Distribution of VOT across Voicing Categories ............................ 43
4.3.1 Progression from short lag to long lag VOT ............................. 44
4.3.2 PLACE OF ARTICULATION and AGE dependency in the manifestation of voiceless unaspirated plosives ................................. 47
4.3.3 On the frequency of prevoicing production across target voicing categories .............................................................................. 48
4.4 Investigating VOT length across development ............................. 49
4.4.1 Means and medians within cells ............................................. 49
4.4.2 Investigating the distribution of VOT measurements ............... 50
4.4.3 Interpretation of the progression in distribution shape ............. 51
4.5 Investigating group differences .................................................. 55
4.6 Discussion ............................................................................ 56

5 Use of aspiration to signal voicelessness ....................................... 59
5.1 Investigating the Measurable RELATIVE ASPIRATION AMPLITUDE .................................................. 60
5.2 Investigation of group differences .............................................. 64
5.3 Discussion ............................................................................ 64
5.4 Conclusions drawn from the use of relative amplitude .................. 66

6 Linguistic uses of voicing cues complementing Voice Onset Time ............................................................................. 67
6.1 Interaction of aspiration force and VOT ..................................... 67
6.2 The use of f0 onset frequency for signalling of a voicing distinction .................................................. 69
6.3 Discussion ............................................................................ 73
# Table of Contents

## III Structural complexity

7 Acquisition of structurally complex syllable onsets

7.1 Structural complexity

7.1.1 Frequency of occurrence

7.1.2 False positives

7.1.3 Production stability in the manifestation of complex syllable onsets

7.2 Duration of the produced syllable onset

7.2.1 Statistical testing of AGE and produced onset complexity effects

7.3 Discussion

## IV Acquisition of manner and place of articulation

8 Acquisition of manner of articulation

8.1 Manner of articulation in productions of simple syllable onsets

8.2 Manner of articulation in complex syllable onsets

8.3 Achievement of a stable production of manner

8.4 Discussion

9 Acquisition of place of articulation

9.1 Place of articulation in simple syllable onsets

9.2 Place of articulation in complex syllable onsets

9.3 Achievement of stable production of place

9.4 Discussion

10 Acoustic investigation of refinement in place of articulation

10.1 Acoustic correlates of place of articulation in plosives

10.1.1 Spectral diffuseness

10.1.2 Spectral skewness

10.1.3 Change in spectral tilt

10.2 Acoustic correlates of PLACE OF ARTICULATION in fricatives

10.2.1 Spectral mean

10.2.2 Variance

10.2.3 Skewness

10.2.4 Kurtosis

10.2.5 Onset frequency of $F_2$

10.3 Discussion

## V Synthesis of the acquisition

11 Synthesis of the findings

11.1 Towards a developmental sequence

11.2 Acoustic correlates to development
11.3 Underlying representation confusion ................................. 161
11.4 Suggestions for future research ................................. 162

A Computation of acoustic correlates of phonetic features 181
A.1 Derived temporal quantities ........................................... 181
  A.1.1 Phone and syllable duration ........................................... 181
  A.1.2 Voice Onset Time ........................................... 181
A.2 Derived spectral quantities ........................................... 182
  A.2.1 Degree of $F_2$ co-articulation ........................................... 182
  A.2.2 Relative aspiration amplitude ........................................... 182
  A.2.3 Spectral skewness ........................................... 182
  A.2.4 Spectral kurtosis ........................................... 182
  A.2.5 Spectral tilt change ........................................... 182

B Publications and presentations while a PhD student 185
List of Figures

2.1 Frequency of reductions found in previous investigations ............ 12
2.2 Frequency of prevoicing found in the investigations .................. 18

3.1 Segmentation criteria used for the markup of plosive release and voicing onset ......................................................... 31
3.2 Example markup of post release friction ........................................ 32
3.3 Histogram of Age coverage ...................................................... 32
3.4 Histogram of the distribution of analysed utterances across Age .... 33
3.5 Histogram of target word frequency ........................................... 33

4.1 Number of Voice Onset Time measurements by Place of Articulation and Voicing Category ................................. 40
4.2 Voice Onset Time measurements by Age of the participants. .... 41
4.3 Distribution of VOT categories across target voicing and aspiration quality ................................................................. 44
4.4 Distribution of VOT categories across target Voicing Category between ages 18–36 months ................................................. 45
4.5 Distribution of VOT categories across target Voicing Categories between ages 36–54 months ................................................. 46
4.6 The distributions of VOT measurements of aspirated targets .... 52
4.7 Histogram of residuals of the VOT model ................................. 55

5.1 Histogram of residuals for the mixed-effects model of Relative Aspiration Amplitude by Age and target plosive Voicing Category. 62

6.1 Speculated effect of relative aspiration amplitude on VOT perception .... 70

7.1 Bar chart showing relative frequency of complex onsets against Age for targets /sak/, /pak/, /bak/ and /spak/ ...................................... 81
7.2 Bar chart showing relative frequency of complex onsets against Age for targets /sal/, /kal/, /gal/ and /skal/ ................................... 81
7.3 Bar chart showing relative frequency of complex onsets against Age for targets /so/, /to/, /do/ and /sto/ ........................................ 82
List of Figures

7.4 Bar chart showing relative frequency of complex onsets against Age for targets /sor/, /por/, /bor/ and /spor/ ................................. 82
7.5 Frequency of too long sequences in the onset by Age of the Participant. 83
7.6 Frequency of too long sequences in the onset by target word. ....... 84
7.7 Progressive relative frequency of complex onsets with two components for target words /sak/, /pak/, /bak/ and /spak/ ......................... 85
7.8 Progressive relative frequency of complex onsets with two components for target words /sal/, /kal/, /gal/ and /skal/ .......................... 86
7.9 Progressive relative frequency of complex onsets with two components for target words /so/, /to/, /do/ and /sto/ ............................ 86
7.10 Progressive relative frequency of complex onsets with two components for target words /sor/, /por/, /bor/ and /spor/ ......................... 87
7.11 Box-and-whisker plots of the duration of (from the top) one, two and three-consonant onsets ........................................... 90
7.12 Box-and-whisker plots of the duration of the syllable onsets produced for the target words /sal/, /kal/, /gal/ and /skal/ ......................... 91
7.13 Box-and-whisker plots of the duration of the syllable onsets produced for the target words /sak/, /pak/, /bak/ and /spak/ ......................... 91
7.14 Box-and-whisker plots of the duration of the syllable onsets produced for the target words /sor/, /por/, /bor/ and /spor/ ......................... 93
7.15 Box-and-whisker plots of the duration of the syllable onsets produced for the target words /so/, /to/, /do/ and /sto/ ............................ 93

8.1 Relative frequency of occurrence for each Manner of Articulation in simple productions of /sal/, /kal/, /gal/ and /skal/ ....................... 98
8.2 Relative frequency of occurrence for each Manner of Articulation in simple productions of /sak/, /pak/, /bak/ and /spak/ ....................... 99
8.3 Relative frequency of occurrence for each MANNER OF ARTICULATION in simple productions of /sor/, /por/, /bor/ and /spor/ .............. 101
8.4 Relative frequency of occurrence for each MANNER OF ARTICULATION in simple productions of /so/, /to/, /do/ and /sto/ ....................... 102
8.5 MANNER OF ARTICULATION of articulation for two-consonant onsets produced for complex target words .............................. 104
8.6 MANNER OF ARTICULATION for two-consonant onsets produced for the /sak/, /sal/, /so/ and /sor/ target words ............................. 106
8.7 MANNER OF ARTICULATION for two-consonant onsets produced for the /pak/, /kal/, /to/ and /pok/ target words ............................. 106
8.8 MANNER OF ARTICULATION for two-consonant onsets produced for the /bak/, /gal/, /do/ and /bok/ target words ............................. 107
8.9 Progressive relative frequency of adult-like MANNER OF ARTICULATION for all elements in the onset for target words /sal/, /kal/, /gal/ and /skal/ ...................................................... 108
8.10 Progressive relative frequency of adult-like MANNER OF ARTICULATION for all elements in the onset for target words /sak/, /pak/, /bak/ and /spak/ ...................................................... 108
List of Figures

8.11 Progressive relative frequency of adult-like MANNER OF ARTICULATION for all elements in the onset for target words /sor/, /por/, /bor/ and /spor/.

8.12 Progressive relative frequency of adult-like MANNER OF ARTICULATION for all elements in the onset for target words /so/, /to/, /do/ and /sto/.

9.1 Relative frequency of occurrence for each PLACE OF ARTICULATION in simple productions of /sal/, /kal/, /gal/ and /skal/.

9.2 Relative frequency of occurrence for each place of articulation in simple productions of /skak/, /pak/, /bak/ and /spak/.

9.3 Relative frequency of occurrence for each place of articulation in simple productions of /sor/, /por/, /bor/ and /spor/.

9.4 Relative frequency of occurrence for each place of articulation in simple productions of /so/, /to/, /do/ and /sto/.

9.5 PLACE OF ARTICULATION for two-consonant onsets produced for [sk]-target words.

9.6 PLACE OF ARTICULATION for two-consonant onsets produced for complex [sp]-target words.

9.7 PLACE OF ARTICULATION for two-consonant onsets produced for complex [st]-target words.

9.8 Progressive relative frequency of adult-like PLACE OF ARTICULATION for all elements in the onset for target words /sal/, /kal/, /gal/ and /skal/.

9.9 Progressive relative frequency of adult-like PLACE OF ARTICULATION for all elements in the onset for target words /skak/, /pak/, /bak/ and /spak/.

9.10 Progressive relative frequency of adult-like PLACE OF ARTICULATION for all elements in the onset for target words /sor/, /por/, /bor/ and /spor/.

9.11 Progressive relative frequency of adult-like PLACE OF ARTICULATION for all elements in the onset for target words /so/, /to/, /do/ and /sto/.

10.1 Variance of the release spectrum against AGE of the participant. The data set is divided by PLACE OF ARTICULATION of the target plosive.

10.2 Variance of the release spectrum against AGE of the participant. The data set is divided by PLACE OF ARTICULATION transcribed by the author.

10.3 Kurtosis of the release spectrum against AGE of the participant. The data set is divided by PLACE OF ARTICULATION of the target plosive.

10.4 Kurtosis of the release spectrum against AGE of the participant. The data set is divided by PLACE OF ARTICULATION transcribed by the author.

10.5 Skewness of the release spectrum against AGE of the participant. The data set is divided by PLACE OF ARTICULATION of the target plosive.
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.6</td>
<td>Skewness of the release spectrum against <strong>Age</strong> of the participant. The data set is divided by <strong>Place of Articulation</strong> transcribed by the author.</td>
</tr>
<tr>
<td>10.7</td>
<td>Categorisation of plosives by the spectral tilt change in from the release to the onset of voicing by <strong>Place of Articulation</strong> of the target plosive.</td>
</tr>
<tr>
<td>10.8</td>
<td>Categorisation of plosives by the spectral tilt change in from the release to the onset of voicing by <strong>Place of Articulation</strong> of the transcribed plosive.</td>
</tr>
<tr>
<td>10.9</td>
<td>Spectral mean of the fricative spectrum against <strong>Age</strong> of the participant. The data set is divided by <strong>Place of Articulation</strong> of the target plosive.</td>
</tr>
<tr>
<td>10.10</td>
<td>Spectral mean of the fricative spectrum against <strong>Age</strong> of the participant. The data set is divided by <strong>Place of Articulation</strong> of the transcribed plosive.</td>
</tr>
<tr>
<td>10.11</td>
<td>Variance of the fricative spectrum against <strong>Age</strong> of the participant.</td>
</tr>
<tr>
<td>10.12</td>
<td>Spectral variance of the fricative spectrum against <strong>Age</strong> of the participant. The data set is divided by <strong>Place of Articulation</strong> of the transcribed plosive.</td>
</tr>
<tr>
<td>10.13</td>
<td>Skewness of the fricative spectrum against <strong>Age</strong> of the participant.</td>
</tr>
<tr>
<td>10.14</td>
<td>Spectral skewness of the fricative spectrum against <strong>Age</strong> of the participant. The data set is divided by <strong>Place of Articulation</strong> of the transcribed plosive.</td>
</tr>
<tr>
<td>10.15</td>
<td>Kurtosis of the fricative spectrum against <strong>Age</strong> of the participant.</td>
</tr>
<tr>
<td>10.16</td>
<td>Spectral kurtosis of the fricative spectrum against <strong>Age</strong> of the participant. The data set is divided by <strong>Place of Articulation</strong> of the transcribed plosive.</td>
</tr>
<tr>
<td>10.17</td>
<td><strong>F₂</strong> onset frequency in pre-vocalic fricatives against <strong>Age</strong> of the participant.</td>
</tr>
<tr>
<td>10.18</td>
<td><strong>F₂</strong> onset frequency in pre-vocalic fricatives perceived as alveolar against <strong>Age</strong> of the participant</td>
</tr>
<tr>
<td>10.19</td>
<td>Absolute difference between <strong>F₂</strong> onset frequency and <strong>F₂</strong> in mid-vowel in pre-vocalic fricatives against <strong>Age</strong> of the participant.</td>
</tr>
</tbody>
</table>
List of Tables

3.1 The target word corpus .................................................. 23
3.2 Information on the participants ........................................ 26
3.3 Spontaneity labels used when extracting words productions from the original recording ........................................... 29
3.4 Context labels used when extracting words productions from the original recording ........................................... 29

4.1 Percentage of productions in cells by Age, Voicing Category and VOT Categories ........................................... 47
4.2 Sample size, mean and median values for the produced VOT ........ 53
4.3 Sample size and standard deviation for produced VOT ............ 54

5.1 Frequency of amplitude measurements by Place of Articulation and Voicing Category ........................................... 61
5.2 Number of productions of measurable relative amplitude divided by Voicing Category, Place of Articulation and Age Category ........................................... 61
5.3 Relative frequency of productions of measurable aspiration by voicing category, Place of Articulation and Age ........................................... 62
5.4 Frequency, mean and standard deviation of the relative amplitude of the fricated portion by age and Voicing Category ........................................... 63

6.1 Mean, median and standard deviation of f0 by target Voicing Category, Place of Articulation and Age ........................................... 71
6.2 Mean, median and standard deviation for f0 across produced segment’s Voicing Category and Age ........................................... 71
6.3 Descriptive statistics for f0 at voicing onset across produced segment’s VOT Category and Age Category. The statistics presented are cell size, mean, median and standard deviation ........................................... 72

7.1 Age of stable production of two-element onsets in sC target words .... 87
7.2 Age of reduction of CC onset productions in target words with C syllable onset structure ........................................... 87
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Table Number</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>8.1</td>
<td>Distribution of productions across MANNER OF ARTICULATION for each target word within the quadruple /s1/, /k1/, /g1/ and /sk1/</td>
<td>99</td>
</tr>
<tr>
<td>8</td>
<td>8.2</td>
<td>Distribution of productions across MANNER OF ARTICULATION for each target word within the quadruple /sk/, /pk/, /bk/ and /spk/</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>8.3</td>
<td>Distribution of productions across MANNER OF ARTICULATION for each target word within the quadruple /so/, /po/, /bo/ and /sp/</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>8.4</td>
<td>Distribution of productions across MANNER OF ARTICULATION for each target word within the quadruple /so/, /to/, /do/ and /sto/</td>
<td>102</td>
</tr>
<tr>
<td>8</td>
<td>8.5</td>
<td>Summary table of MANNER OF ARTICULATION for the first and second consonant in clusters produced for simple target onsets</td>
<td>103</td>
</tr>
<tr>
<td>8</td>
<td>8.6</td>
<td>Summary table of MANNER OF ARTICULATION for the first and second consonant produced for simple target onsets.</td>
<td>105</td>
</tr>
<tr>
<td>8</td>
<td>8.7</td>
<td>Age of stable production of adult-like MANNER OF ARTICULATION of target words.</td>
<td>110</td>
</tr>
<tr>
<td>8</td>
<td>8.8</td>
<td>Delay in stable production of adult-like MANNER OF ARTICULATION compared to production of complex syllable onsets.</td>
<td>111</td>
</tr>
<tr>
<td>9</td>
<td>9.1</td>
<td>Distribution of productions across PLACE OF ARTICULATION for each target word within the quadruple /s1/, /k1/, /g1/ and /sk1/</td>
<td>117</td>
</tr>
<tr>
<td>9</td>
<td>9.2</td>
<td>Distribution of productions across PLACE OF ARTICULATION for each target word within the quadruple /sk/, /pk/, /bk/ and /spk/</td>
<td>117</td>
</tr>
<tr>
<td>9</td>
<td>9.3</td>
<td>Distribution of productions across PLACE OF ARTICULATION for each target word within the quadruple /so/, /po/, /bo/ and /sp/</td>
<td>118</td>
</tr>
<tr>
<td>9</td>
<td>9.4</td>
<td>Distribution of productions across PLACE OF ARTICULATION for each target word within the quadruple /so/, /to/, /do/ and /sto/</td>
<td>121</td>
</tr>
<tr>
<td>9</td>
<td>9.5</td>
<td>Summary table of combinations of PLACE OF ARTICULATION for the first and second consonant in clusters produced for [sk]-target onsets.</td>
<td>122</td>
</tr>
<tr>
<td>9</td>
<td>9.6</td>
<td>Summary table of combination of PLACE OF ARTICULATION for the first and second consonant in [sp] target onsets.</td>
<td>123</td>
</tr>
<tr>
<td>9</td>
<td>9.7</td>
<td>Summary table of combination of PLACE OF ARTICULATION for the first and second consonant in [st] target onsets.</td>
<td>124</td>
</tr>
<tr>
<td>9</td>
<td>9.8</td>
<td>Summary table of combination of PLACE OF ARTICULATION for the first and second consonant in clusters produced for simple target onsets.</td>
<td>126</td>
</tr>
<tr>
<td>9</td>
<td>9.9</td>
<td>Age of stable production of adult-like place of articulation of target words.</td>
<td>127</td>
</tr>
<tr>
<td>9</td>
<td>9.10</td>
<td>Delay in stable production of adult-like PLACE OF ARTICULATION compared to production of MANNER OF ARTICULATION in complex syllable onsets.</td>
<td>130</td>
</tr>
</tbody>
</table>
Chapter 1

Developing phonological contrasts

The phonological systems of the world’s languages show great diversity in the system of distinctive speech sounds. A contrast between two speech sounds upheld by speakers of one language may be not be considered different by speakers of another language. Further, differences may be present in the boundaries used for an acoustic feature present in many languages (e.g. Lisker & Abramson, 1964 and Cho & Ladefoged, 1999), which may cause speakers to perceive a given speech element differently in ambiguous cases (e.g. Caramazza, Yeni-Komshian, Zurif & Carbone 1973).

Due to this diversity in language systems, the establishment of a model for the acquisition of the speech capacity by children presents a challenge to the speech research community due to the complexity of the task. The normally developing child is put to the task of, without formal instruction, deriving which elements are to be considered the basic contrasts between speech elements within the language to be learnt from the speech stream in its acoustic, unsegmented, form.

There have been a number of attempts to characterise the process of phonological acquisition using theoretical models of the components involved. Some of the early models proposed that speech development in children operated solely on the level of articulatory ability, i.e. the speed and precision by which the articulatory movements were made. In an early model, Smith (1973) proposed that:

“[...] only the analysis which presupposes an adult competence for the child and which treats the child’s phonology as a mapping from the adult’s system attains descriptive adequacy.” (Smith 1973, p. 5)

Smith argued that changes in the manifestation of a speech element occurs across the board at the time when they first arise. Thus, children starting to produce aspiration would produce aspiration in all appropriate contexts at the same time. Models similar to Smith (1973) in their proposition of adult-like competence for the child were proposed by several researchers during the seventies and the early eighties (e.g. Ingram (1976) and Donegan & Stampe (1979), see Maxwell (1984) for a review of proposed models).

These early models of speech development were subsequently shown to be not accurate. For instance, Macken (1980) provided evidence of instance-based reorganisations of output forms in Smith’s (1973) own data (See also Braine 1976) indicating that the
child’s underlying representations differ from the adult surface form. This insight led to theoretical models of children’s phonology which incorporated a perceptual encoding module or filter between the adult input and the child’s internal representation (See e.g. Maxwell 1984). It was, therefore, recognised that speech development may involve maturation of both speech production ability and the detail of the specification of the underlying representation of the utterance.

In view of the language specific boundaries for acoustic properties in the input, the child may thus be able to separate the linguistically relevant signal from the noise consisting of both linguistically irrelevant properties in general and properties not significant for the contrast within the specific language of the adult.

The tools available to the child in the task of sifting through potential linguistically relevant acoustic cues have been shown to be diverse in nature. Children have been shown to utilise statistical cues such as differences in distribution probabilities of segment sequences across and within word boundaries. Children have been observed in experimental investigations to be able to extract words in an artificial speech stream based on the probability of clusters occurring at word boundaries in the ambient language (Saffran, Aslin & Newport 1996, Mattys, Jusczyk, Luce & Morgan 1999, Mattys & Jusczyk 2001).

Further, the prosodic cue of stress placement has been shown by Mattys et al. (1999) to afford the perception of a word onset in an unsegmented speech stream. Mattys et al. (1999) showed that the acoustically more salient stressed syllable followed by a weaker unstressed syllable formed a dominant cue for children in the finding of potential word-segmentation points. Thus, it seems that the perceptual salience of the acoustic cue may form a factor in the child’s construction of an hypothesis about the relevant acoustic cues for a specific contrast.

Investigations of the speech directed towards children have shown that adults may simplify their speech pattern, possibly in order to increase the perceptual friendliness of the speech signal and thus enhance the child’s ability to extract information from it. It was shown by for instance Fernald (1989) that the intonation contour in child-directed speech is distinguishable from that of adult-directed speech.

Differences between child-directed speech and adult-directed speech are, however, seen not only in prosody. At the segmental level, Shockey & Bond (1980) showed that processes described by reduction rules occur more frequently in speech addressed to children compared to speech addressed to adults. Thus, it may be that adults may choose to alter their utterance towards a form that they perceive as more simple, instead of providing a more accurate model utterance for the child to base their attempted productions on. Significant for this thesis is that Shockey & Bond (1980) reported high frequencies of occurrence of reduction processes that occur frequently in the speech of children, such as production of clusters as a singleton consonant and deletion of fricatives. Thus, the model production presented to the child may be adjusted by the adult producer, decreasing the perceptual gap between the output form of the child and that of the input form presented to the child.

At a general level, Fernald (2000) proposed that adults may modify their speech pattern in order to manifest a maximum predictability for the child observer. Fernald (2000) argued that child directed speech should be viewed as a form of hyperspeech (Lindblom 1990), which may assist comprehension by providing a speech signal that is accessible to infants.
Developing phonological contrasts

The proposed link between the productions of mothers and perception in infants has been shown to hold for specific acoustic features, for instance the perception of vowels. Liu, Kuhl & Tsao (2003) reported that discriminatory ability of vowels in children is correlated with the mother’s vowel space area; mothers were shown to modify their productions in order to provide a tokens that are perceptually more salient or easily processed compared to adult-directed speech. This process indicates a fruitful cooperation between adult and child in facilitating the formation of the language-specific contrasts in perception.

1.1 From perception to production through stages of potential homonymy

Due to the data-driven nature of the acquisition of language specific contrasts, understanding of the perception of a contrast naturally predates deliberate attempts of manifestation of these contrasts. Here, limitations in both the child’s articulatory ability and the underspecified phonological understanding of what is to be produced may cause the child’s output forms to differ from model productions made by adults. In some cases, the output form produced by the child may correspond perceptually with adult’s productions of another potential target word. In addition, the child’s output forms of both of these rivalling target words may be perceptually very similar or identical to the adult observer. Priestly (1980) argued that these situations may give rise to homonymy, half-contrast or pseudohomonymy in the productions.

Priestly defined homonymy to be productions where both the child and the adult perceiver view productions of the rivalling target words to be identical. Half-contrasts were defined by Priestly as when the child is not aware of the contrast, but the difference is perceived by the adult observer. Pseudohomonymy was defined by Priestly to occur when the child is aware of the contrast, but is not able to produce it in a manner that is perceived by the adult observer.

One indication concerning the status of the potential homonymy may come from adult feedback by imitation of the perceptual results of the child’s output. A number of researchers have provided evidence of apparent pseudohomonymy by referring to dialogues where the child indicates a phonological specification beyond the produced form (Gierut 1991, McLeod & van Doorn 1996, Priestly 1980). The evidence is often described as a reaction by the child that indicates a failure to recognise that the imitation provided by the adult is acoustically similar to the child’s own production. In instances of productions of pure homonymy, the child has not yet developed a full phonological specification of what is to be produced, and, therefore, cannot distinguish between the rivalling output forms. Thus, when fed back with an erroneously produced output form, the child does not have enough information to distinguish between the target form normally produced by the adult and the form fed back and, therefore, does not react to the switch in output form specification.

In pseudohomonymy, however, the child has an underlying specification containing enough detail in order to distinguish the form fed back from the forms usually perceived. Therefore, when fed back an utterance that is acoustically similar to what the child itself produced, the child may indicate that the output forms differ, sometimes with some indignation.
Success in noticing differences between minimal pairs that are not noticed in the case of homonymy indicates a more evolved phonological system that has incorporated more of the contrasts available in the ambient language. With each expansion in deliberate production ability, a period of expansion of the acoustic cues noticed and explored for linguistic significance occurs that provides the groundwork for an increasingly adult-like production.

This Thesis explores one such period in the development of child speech, namely, the acquisition of a production ability for s+Plosive consonant clusters. This period investigated covers the transition from the potential perception of the syllable onset towards an increased awareness with the syllable onset parts. The background against which this investigation is framed, a review of the literature on syllable onset development from the onset of syllable-like productions, is presented in the next section.

1.2 Development of syllable and monosyllabic-word production ability

In the literature on the early productions of children considerable attention has been given to the rise of number of productions perceived as syllable-like by the adult observer. Initially, these syllable-like utterances occur in sequences of identical, reduplicated, perceptually identical syllable-like components.

The phonological status of early syllable-like productions as attempted CV production has been considered in previous research. Kent & Murray (1982) noted that the early research conducted on these productions tended to assume a segmentation process of the utterance into phone-sized units in the child’s perception. They argued that it is possible that the segments do not exist as models in the child’s production and questioned the function of the produced CV-like sequences as attempts to imitate speech:

“What linguists perceives as a CV syllable in an infant’s babbling is a movement sequence, usually with concurrent phonation, in which one or more articulatory structures move from a position that constricts the vocal tract to one that does not constrict the vocal tract. The fact that most CV syllables are heard as stop + vowel syllables indicates that the initial constricted position is a complete obstruction, serving to shut off momentarily all air flow through the vocal tract.” (Kent & Murray 1982, p 362)

Thus, the shape of the resulting output form of the child’s syllable-like productions is mainly governed by the opening and the closing of the mandible during phonation (Kent & Murray 1982, MacNeilage & Davis 2000). In addition, data from Thelen (1981) suggest that the approximate onset of reduplicated babbling occurs in a stage of development where repetitive movements frequently occur in developing motor skills. It is, therefore, possible to view the structure of the syllable-like productions produced in the repetitive babbling stage as more influenced by the characteristics of the child’s motor skills development in articulators than by the syllable-internal structure of the target words to which the child is exposed. It has been argued that these productions, as they are a mechanical consequence of repetitive behaviour, should not be considered manifestations of a syllable that have been imperfectly produced, implying
Developing phonological contrasts

a phonological status of production (Oller 2000). Oller (2000) referred to children’s early productions of syllable-like as *canonical syllables*, indicating that the shape of these early syllable-like sequences may resemble the output forms of adult syllables, but the phonological specificity of these productions may differ substantially from that of adult productions.

Such rudimentary syllable-like productions will later on, through the perception–production loop fed by salient adult productions (Vihman & de Boysson-Bardies 1994), form the basis for early productions used to convey meaning. As mentioned in Section 1, this process has been reported to involve productions of difficult word-forms to be perceptually similar to the output forms of other target words, producing homonyms or pseudohomonyms. In the development away from these perceptual mergers, the output form becomes increasingly adult-like in order to uphold the necessary contrasts within the ambient language. In words where a consonant contrast in voicing, place and manner of articulation is upheld, these features will have to be increasingly produced in an adult-like manner, and in the correct contexts, to facilitate an expansion of the word forms usable in communication. Further, in languages in which complex syllable onsets occur in adult speech, the child has to notice them and expand the set of syllable onsets produced to incorporate those with more than one consonant and with adult-like acoustic feature specifications.

1.3 Methodological Issues

Due to the uncertainty in the status of the phonological and phonetic specifications of an attempted target word in early productions, many of the issues involved in the study of speech productions in the first year applies also to the study of later development. Specifically, there is the issue of which features the chosen methodology may hide, or introduce.

Oller (2000) argued that the early investigations of children’s onset of vocalisations that were based on transcriptions, or diaries, of the child’s productions may have introduced a *shoehorning* effect. That is the properties of the adult observers linguistic perception may have constrained what was perceived, transcribed or recorded in a diary. It has also been argued that the development of a perception system for speech sounds involves a transition from a general perception of most contrasts, towards a perception system where non-native contrast are attenuated (see Vihman, 1996 for a review). Thus, as development progresses, the child looses its initial ability to distinguish between output forms that differ in non-distinctive acoustic features in the ambient language.

The adult perceiver, having gone through the process of attenuation of non-native contrasts, is thus hindered by their own phonetic categories and phonological system to perceive details in productions that fall outside of the empirically established categories. Oller (2000) argued that the adult observer may attempt to *shoehorn* the child’s productions into a category that is familiar to the observer, but may not be an accurate description of the actual production made by the child. Oller (2000) noted that transcriptions of early child vocalisations, when read aloud by a trained phonetician may differ substantially from the form of the original utterance made by the child. Thus, transcriptions of early productions run the danger of being *shoehorned* into an adult category, reducing the empirical value of the observations.
The same problem may also arise in transcriptions of later productions where the output form produced by the child differs substantially in acoustic form from that of adult productions. In such cases, the adult observer may place the production in a category that is an inadequate description of the actual production. Thus, an observer-bias due to the contrasts in the observer’s native language is introduced. There is, therefore, a need for a more objective measurement of the child’s progression and the properties of the output forms than is afforded by the human observer.


Methodologically this Thesis combines phonetic transcription with acoustic measurements and interprets the acoustic data in terms of speech development trends against a theory that provides a background for evaluation of the measurements. It is argued that this theoretical basis for evaluation may come from more than one source.

First, the literature on acoustic cues to phonetic or phonological distinctions in adult speech provides a point of reference towards which children’s productions should develop. In cases where multiple cues have been identified, the child’s progression may follow different cues at different stages of development.

Second, data suggesting an advantage in production economy or articulatory complexity may provide a foundation for an hypothesis regarding the reasons behind the observed progression. It is proposed that if development is observed in a direction that is in contrast with default production patterns, this may be included as empirical evidence for the construction of an interpretation of the acoustic measurements obtained.

1.4 Issues addressed in this Thesis

In the exploration of the child’s transition into the distinctive use of increasingly specified output forms, two themes are pursued in this thesis. The first theme aims to provide a description of the development of word-initial s+Plosive consonant clusters in Swedish children. The methodology employed involves a contrastive analysis of productions of target words with an s+ plosive syllable onset with the output forms of target words that are homophones of a reduced version of the s+plosive consonant. The consonants produced in the syllable onset are analysed in terms of their quality
Developing phonological contrasts in voicing, aspiration, place of articulation and manner of articulation through both auditory observations and measurements of phonetically relevant acoustic cues to these features. A production pattern is not seen as having been established until it is present in places that are appropriate and not present in places that are inappropriate in the ambient language. Estimations of the variability is provided by accounts of the stability of production and developmental trends in the variance found in the acoustic measurements.

The second theme is the explanation of, and development of hypotheses, for the developmental trends observed in the development s+Plosive consonant clusters in Swedish children. Acoustic measurements of children’s productions increase our knowledge of the developmental sequence and permit the formation of hypotheses for the observed progressions.

1.5 Structure of the Thesis

This Thesis is divided into five parts. In Part I, the background to the study and the methodology are presented. Chapter 2 provides the theoretical foundation for the study by overviewing the literature on the development of phonological constraints and motivating the aspects to be investigated in the acquisition of Swedish s+plosive clusters. Chapter 3 presents the procedure for data gathering and analysis. This includes the rationale of the corpus design, information about the participants and their recruitment, the ethical aspects of the study, details about the extraction and annotation of recordings and the segmentation criteria for the acoustic cue measures marked and analyzed. Appendix A provides definitions of the acoustic quantities calculated from the marked acoustic cues.

In Part II, the acquisition of the contrastive use of voicing across target words containing voiced, voiceless unaspirated and voiceless aspirated plosives in the syllable onset is investigated. Chapter 4 considers the use and development of the cue of Voice Onset Time — time delays between the transient associated with oral release of the plosive and the onset of f0. In Chapter 5 the amplitude of the aspirated portion as an alternative cue to voiceless is investigated, and in Chapter 6, the use of aspiration amplitude and f0 onset frequency as auxiliary cues to the voicing distinction is explored.

In Part III, the rise of structural complexity in the children’s productions of s+plosive target words is investigated. Comparisons are made between productions where the intended target production has a consonant cluster onset and productions of comparable target words with a single-consonant syllable onset. The comparable target words have identical nuclei and codas.

In Part IV, the acquisition of the stable production of consonants with an adult-like manner of articulation (Chapter 8) and an adult-like place of articulation (Chapter 9) based on perceptual data is analysed. The analysis includes the manifestation of these features in singleton and complex syllable onsets produced in both adult-like and not adult-like contexts. The productions contrasted in Chapter 8 and Chapter 9 are productions of target words used in this Thesis and presented in Chapter 3. In Chapter 10 the developmental trends and distinctive usage of established acoustic cues to place of articulation in plosives and fricatives are explored; this chapter extends the analysis of developmental trends in terms of place of articulation, which was based on transcription data.
Part V synthesizes the contrastive use of output forms and the development of an increasingly adult-like productions and presents the conclusions that can be drawn from the study reported in this Thesis.
Part I

Background and methodology
Chapter 2

Aspects of s+Plosive clusters

This chapter overviews previous research in contrast marking and the acquisition of the plosives and fricative /s/ in s+Plosive clusters and motivates the selection of the aspects to be investigated in the acquisition of the Place of Articulation, Manner of Articulation, complexity and voicing in these clusters.

2.1 Processes in s+plosives productions by children

In the initial stages of the production of consonant clusters, the output forms produced by children have been shown to be reduced both in quality and structural complexity. Between the production of simple CV syllables and adult-like production of a CCV syllable, a number of intermediate stages have been proposed. For s+plosive consonant clusters, the well-known early bias towards deletion of the consonants has been noted by researchers including Linell & Jennische (1980), McLeod, van Doorn & Reed (2001a, 2001b).

As a consequence of this process, intended productions of /stɔr/ may be produced as /ʃɔ/: Production of only the fricative, producing /ʃɔ/, is also possible, but occur much less frequently due to the low frequency of fricative productions in early child speech. Through the use of this process, the child is reducing the cluster to a form that it may produce, either as an effect of an inability to achieve the necessary articulatory coordination or as an effect of the child not having achieved an ability to perceive the difference between output forms.

The relative frequency of occurrence of process cluster reduction has been shown to be highest in initial attempted cluster productions, and to decrease with age. McLeod, van Doorn & Reed (2001b) presented a summary of the frequency of occurrence of cluster reduction in six longitudinal investigations of English-acquiring children between the ages 1:6 and 8:0. The data collected by McLeod et al. (2001b) is presented in Figure 2.1. As seen in Figure 2.1, the relative frequency of occurrence of the cluster reduction process decreased from 93% at 1:6 to 3% at 8:0.

An alternative to the cluster reduction process is the coalescence process (Grunwell 1987). Here the cluster may be reduced to a single consonant, incorporating features
Aspects of s+Plosive clusters

Figure 2.1: Frequency of reductions, in percent, found in the investigations reviewed in McLeod, van Doorn & Reed (2001b). The vertical lines indicate the range of frequencies found in the grouped results. No range is indicated for ages covered by a single investigation. The dotted line shows the progression of mean frequency of reductions across ages.

from both members of the target cluster. By the application of this process, the output form of \textquoteleft\textquoteleft/swim\textquoteright\textquoteleft\textquoteleft may become \textquoteleft\textquoteleft/rim\textquoteright\textquoteleft\textquoteleft (Dyson & Paden 1983). Thus, the combination of a fricative followed by a consonant produced partially at the labial Place of Articulation creates a labiodental fricative in the child’s output. Thus, through the application of this process, the child shows clear signs of being aware of both elements of the target consonant cluster.

At the time when complex syllable onsets are starting to be produced, structurally complex production may deviate from the quality of the target form in the phonetic specification of the sub-elements. This process is referred to as \textit{cluster simplification} and was described by McLeod et al. (2001b) as

“Cluster simplification occurs when two elements of the cluster are produced, but one or both of the two elements are produced in a non-adult manner”. (McLeod et al. 2001b, p 102)

McLeod et al. (2001b) note in their review of processes in consonant cluster development a trading relation between \textit{cluster reduction}, \textit{cluster simplification} and the production of adult-like consonant clusters. As the production frequency of \textit{cluster reduction} increases, an increase is generally observed in frequency of \textit{cluster simplification}. With time, the decrease in frequency of \textit{cluster simplification} will lead to adult-like production pattern for the cluster.

Before the child reaches the stage of adult-like production of the cluster, the processes of \textit{cluster reduction}, \textit{coalescence} and \textit{cluster simplification} cause the produced output form to differ from the adult target production. In the case of \textit{cluster simplification} and \textit{coalescence}, the output form has a structurally simple syllable onset. Thus, the productions made by the child towards consonant cluster target words may potentially be difficult to separate from output forms of targets with a structurally
simple syllable onset, creating homonym or pseudohomonym productions. With time, the output forms have to become to be increasingly distinct in terms of their phonetic content in order to achieve the adult-like contrasts of the ambient language.

2.2 Voicing in plosives

In the data driven model of voicing investigated in this Thesis, the nature of the voicing contrast in the speech of adult surrounding the child has to be considered.

2.2.1 Voicing in adult-directed speech

The acoustic quantification of phonological voicing has received much attention in previous phonetic investigations of plosive consonants. In an early study Lieberman, Delattre & Cooper (1958) noted that English voiceless plosives were consistently produced with a lag in time of onset of F₁ compared to that of F₂ and F₃. In voiced plosives, the voicing lag was less than voiceless plosives, if present at all. This lag in F₁ was termed “F₁ cut-back” and was shown to be dependent both of the Place of Articulation of the plosive and of the frontness of the vowel. In addition, Lieberman et al. (1958) noted that the onset frequency of F₁ also served as a strong cue to the voiced/voiceless distinction.

This observation was in part confirmed by Stevens & Klatt (1974) who proposed that a rapid spectral change in the frequency region of F₁ together with a temporal separation of F₁ onset and that of higher formants would serve as the dominating cues to the voicing distinction. In a later experiment investigating the perception of the phonological voicing feature in high vowels and in low vowels, they concluded that the temporal separation cue was dominant only for high vowels after velar plosives, where the F₁ transition is minimal. At the same time, Summerfield & Haggard (1974) reported contradictory findings. They found that the F₁ transition is larger, and used as stronger cue, in low vowels produced after velar plosives.

It was recognised in early research on voicing quantification that the primary cause of the observed temporal separation between the onset of F₁ compared to higher formants in adult speech is the difference in the time of f₀ onset. In voiced plosives, f₀ is observed soon after the release of the plosive; in voiceless plosives, F₂ and higher formants are exited by the aperiodic aspiration force but F₁ onset is delayed until the it is given acoustic energy by the onset of f₀. This lead Lisker & Abramson (1964) to argue that a single measure of the temporal delay of voicing onset relative to the release of the plosive adequately separated productions across a number of languages with 2–4 categories of voicing/aspiration contrasts. The term chosen for this acoustic measurement was Voice Onset Time (VOT) that was calculated as the time difference between the onset of voicing and the release of the plosive. A positive Voice Onset Time value signifies that the plosive has been released prior to the onset of voicing i.e. was produced with a voicing lag; a negative value signifies that the onset of voicing happened prior to the release (prevoiced).

The perceptual interpretation of the Voice Onset Time measurement was shown by Lisker & Abramson (1964) to be language dependent. In a language with two categories of plosives in initial position, there is a boundary value which separates voiced unaspirated from voiceless aspirated plosives both in production and perception. For
Aspects of s+Plosive clusters

In the languages English and Swedish, boundary values for voiced and voiceless plosives in initial position have been reported to be between $+20 \text{ ms}$ to $+40 \text{ ms}$ VOT depending on Place of Articulation (Lisker & Abramson 1964, Klatt 1975, Zlatin 1974).

Since Lisker & Abramson’s (1964) proposal, several factors have been shown to influence the perceptual interpretation of a produced Voice Onset Time value. For example, in recognition of the parallel identification of $F_1$ cutback as a voicing cue parameter, the onset frequency (Summerfield & Haggard 1974) and following transition (Stevens & Klatt 1974, Lisker, Lieberman, Erickson, Dechovitz & Mandler 1977) into a steady state in $F_1$ have been investigated as modifying cues to VOT. In a naturally produced early onset of $f_0$, corresponding to a voiced plosive, the initially rising $F_1$ of the plosive is given energy at an early stage and the onset frequency is low and the transition long. On the other hand, a late onset of $f_0$ would give little energy to $F_1$ compared to higher formants, and therefore results in both a higher perceived onset frequency of $F_1$ as well as a shorter transition duration. Through perception tests using synthesised stimuli, Lisker et al. (1977) was able to show that the 50% crossover point in voicing judgements measured in VOT could be pushed around within a 24 ms range through manipulation of the $F_1$ transition length and $F_1$ onset frequency.

Another feature that has been shown to influence the perception of voicing is the aspirated portion preceding voicing onset. Summerfield & Haggard (1974) presented results showing an effect of aspiration in all except one of the $\ddagger$ contexts and none of the $\ddagger$ contexts. For low vowels, voicing perception was found to be significantly affected only by Voice Onset Time. Winitz, LaRiviere & Herriman (1975) argued that aspiration amplitude should be regarded as the dominant acoustic cue to the voicing detection in English and that Voice Onset Time should be regarded as a secondary cue.

In a later investigation, Repp (1979) re-evaluated the results of Summerfield & Haggard (1974) and Winitz et al. (1975) and proposed that the influence of aspiration amplitude should be quantified as a quota of the aspiration amplitude and the amplitude of the following periodic portion of the signal. Using synthesised plosive stimuli with VOT ranges between 8 ms to 44 ms in which either the voiced or the aspirated portion of the stimulus had been amplified by -6 dB, 0 dB or +6 dB, Repp (1979) was able to show that the boundary between voiced and voiceless plosives may be shifted to the positive end of the scale (i.e. leading to more voiced judgements) if the periodic portion was amplified. If the aperiodic portion of the plosive was amplified relative to the periodic portion, the voicing boundary shifted towards the short-lag end of the scale, leading to more voiceless judgements. Repp (1979) proposed that a 1 dB increase in the aperiodic/periodic amplitude ratio would lead to a shortening of the VOT boundary by 0.43 ms.

Listener judgements of voicing quality in ambiguous stimuli and in pressed situations have been shown to be influenced by the manipulation of $f_0$ onset frequency (Haggard, Summerfield & Roberts 1981, Holt, Lotto & Khvendar 2001) and overall $f_0$ frequency (Whalen, Abramson, Lisker & Mody 1990, Whalen, Abramson, Lisker & Mody 1993). Voiceless plosives are usually produced with a high onset frequency and overall level (compared to the baseline set by prosodic properties of the surrounding utterance) compared to voiced plosives (Lisker 1986, Holt et al. 2001). This covariation of voicing and $f_0$ level has been shown to influence judgements of plosives with VOT values close to the voicing boundary (Whalen et al. 1990, Holt et al. 2001). Thus, in
cases where the listeners are put into a situation with increased uncertainty caused, for example, by increased stress, this secondary cue has been shown to have a greater impact on the labelling of plosive voicing quality that it otherwise would have.

2.2.2 Voicing in speech directed to children

Speech directed towards children differs acoustically and functionally from speech directed to adults. At the suprasegmental level, adults have been reported to use a higher \( f_0 \) level, producing a pitch level that is closer to the one produced by the child and, a more varied \( f_0 \) contour that may direct the child's attention towards verbal and social interaction (Fernald 1989).

Investigations of the acoustic properties of segments in words produced by adults to children are, as noted by (Baran et al. 1977, Faulkes, Docherty & Wat 2005) more limited than investigations of suprasegmenal phenomena. In an acoustic property study of six Swedish mothers, Sundberg & Lacerda (1999) collected Voice Onset Time measurements from speech directed towards 3 month old children and speech directed to the adult researcher. The results showed a significant main effect of the person addressed by the speech: both voiced and voiceless plosives were produced with a smaller VOT under the child-directed speech condition compared to the adult-directed condition. In addition, Sundberg & Lacerda (1999) noted a larger degree of variability in the adult-directed speech condition compared to the infant-directed condition for five out of six plosives. The exception was the voiced velar plosive which was produced with larger variance in child-directed speech than in the adult-directed speech. It thus appears that Swedish infants receive productions that may be tuned for easy perception and classification, but that do not emphasise the distinction between aspirated and unaspirated plosives.

In an investigation of speech directed towards older children, Malsheen (1980) measured VOT in plosives for a group of six mothers during a six-month period. The mothers were divided into groups by the age of the addressed child. At the onset of the study, the children in the youngest group were 0;6 and 0;8, in the middle group 1;3 and 1;4 and in the oldest group, 2;5 and 5;2. The results presented by Malsheen (1980) showed a non-significant difference in VOT between the adult-directed and child-directed conditions in the youngest and oldest group. A significant difference was, however, found between speaking conditions for the middle group. Malsheen (1980) noted that this groups differed from the other two groups in that the children addressed was just starting to produce their first words.

At that time, mothers were shown to produce significantly longer VOT in voiceless plosives, increasing perceptual distance between aspirated plosives and voiced plosives. Differentiated modes of production by the mother depending upon the linguistic stage of the child are, therefore, likely.

Sundberg & Lacerda (1999) proposed that obstruent properties, unlike prosodic properties and properties of sonorants, may be under-specified regarding their phonetic realisation in child-directed speech from the time of birth of the child. In the initial year, the phonetic distinctiveness will increase and reach a level of separation of contrasting forms well beyond that of adult-directed speech around the time when the child reaches one year. According to the model presented by Sundberg & Lacerda (1999), the manifestation of the contrasting forms gradually decreases towards that
of adult-directed speech during the later half part of the child’s second year of life. The results obtained from speech directed to the oldest group in Malsheen (1980) suggest that the VOT difference between child-directed speech and adult-directed speech reaches non-significance before the age of 2.5.

2.2.3 The child’s navigation in the multidimensional space of adult voicing

If phonological acquisition as a process is driven by perception rather than innate knowledge, the acquisition of a feature has to involve an extraction of candidate distinctive perceptual properties from the adult target. From these potential distinctive properties, the child then has to construct an hypothesis regarding production strategies for the segment to be perceived correctly by adults surrounding the child.

In light of the wealth of acoustic correlates of phonological voicing within a two-category language (Section 2.2.1), it is likely that the task of acquiring an adult-like phonological voicing contrast includes phases of experimentation with potential cues in order to find the dominant cues and their language specific boundaries. Furthermore, it involves a marking of other cues as secondary or possibly perceptually unimportant covariates.

Previous research on the development of the voicing contrast in word-initial plosives and plosive in word-initial clusters has focused on the VOT measure. For languages with two categories in voicing contrast, the developmental trends found in languages with a contrast between short lagged voicing and long lag voicing will be summarised in Section 2.2.4. In Section 2.2.5, within-category manipulations of perceptual distance using prevoicing are overviewed.

2.2.4 Progression from a unimodal to a bimodal VOT distribution

Several investigations into the initial stages of acquisition of voicing in plosives have concluded that there is a strong bias towards a short lag VOT (i.e. an onset of voicing which coincides roughly with the release of the plosive) across voicing categories (Eguchi & Hirsh 1969, Kent 1976, Kewley-Port & Preston 1974, Zlatin & Koenigsknecht 1976). Kewley-Port & Preston (1974) investigated initial plosives produced by three normally developing children analysed at ages spanning from 34 weeks to 101 weeks. The results showed that the subjects generally produced plosives in the adult short lag range throughout the investigated ages. They also revealed a progression in the individual child from the early productions, produced as a narrow, unimodal distribution, towards the latter produced, showing an increase in distribution width as the child got older. At the age of two years, the VOT distribution still had the bulk of productions in the short lag area, but there was a long tail in the distribution towards the long lag range.

The general tendency for children’s productions of primarily short lagged productions up until the age of 2–3 years has also been reported by Eguchi & Hirsh (1969), Macken & Barton (1978) and Catts & Kamhi (1984). At the age of three, the six children investigated by Gilbert (1977) produced /d/ with a mean value of +29 ms, a standard deviation of 19 ms and with a low frequency tail into the long lag VOT range (up to 100 ms lag). Productions of /t/ were produced with a mean VOT of +140 ms
and with a wide distribution (sd=63ms), extending over a range from 20 to 290 ms. A number of productions of phonologically voiced plosives were produced with VOT values more typical of the voiced plosives. Furthermore, some phonologically voiceless plosives were still produced with a short lag VOT. Therefore, previous research suggests that a production distinction of voicing is not fully established at the age of three years.

However, as development continues, the child starts producing distinct VOT distributions for aspirated and non-aspirated targets. Kewley-Port & Preston (1974) presented data showing more long lag VOT produced by their investigated subject at 3;6 than at 2;1 for labial, dental and velar plosives. For older children, data from Menyuk & Klatt (1975), Bond & Wilson (1980) and Barton & Macken (1980) show clear separation in the range of produced VOT at the end of the fifth year. In addition, the standard deviations and range statistics reported by Barton & Macken (1980) indicated a decrease in distribution width in the short lag range for voiced target plosives, indicating an increasing separation between the two distributions. Thus, children show a progression from a unimodal distribution centred at the short lag range to a bimodal distribution in VOT as more plosives are produced with a long lagged VOT. The distribution peaks are initially wide with substantial overlap. Yet, the overlap decreases with time, and at the age of five, the separation of the two peaks is comparable to that of adult speakers.

2.2.5 Prevoicing production in development

Although prevoicing is not distinctive at the phonological level, research that has examined the frequency of use and length of prevoicing have found developmental change for unaspirated plosives. Figure 2.2 presents the use of prevoicing in studies where these may be separated from the productions with VOT lag (Catts & Kamhi 1984, Macken & Barton 1978, Mack & Lieberman 1985, Zlatin & Koenigsknecht 1976). In Figure 2.2, the percentage use of prevoicing in relation to the total number of voiced productions is displayed against age.

Two observations can be made from Figure 2.2. First, there is no general trend of decrease in prevoicing use with age. The range of percentages observed at 16 months is similar to that observed at 28 months. Due to the lack of data from children between the age of 40 and 72 months, it is impossible to know whether children are using prevoicing as a distinctive cue or not at these ages. Second, the variance within each study is substantial. The relative use of prevoicing varied from 5% to 46% at the same age. The data from previous studies do not afford an analysis of the use of prevoicing as a function of age and the role of prevoicing in the development of the voicing contrast remains largely unexplored.

2.3 Place and manner of articulation

The acquisition of Place of Articulation and Manner of Articulation are a central to the acquisition of contrasts in singleton and cluster contexts. In this section, the properties of place and manner of articulation are presented. This section begins by presenting a developmental hierarchy based on the perceptual salience and robustness of the phonetic features of place and manner of articulation. Then, processes in child
speech that operate at the segmental feature level are reviewed in terms of their output forms and potential for homonym production and finally, the acoustic cues to place of articulation for plosives and fricatives are described.

2.3.1 Perceptual stability of phonological features in the ambient language

If a child’s phonology is shaped by the properties of the input, perceptual salience of a feature may be a factor in the acquisition of a production ability of this feature. Therefore, the salience and robustness of a feature in perception may potentially serve a role in the order of development.

Under ordinary listening conditions, the phonological features of place and manner have been shown in perceptual tests to be processed similarly. Miller & Eimas (1977) investigated the perceptual distinction between stops and nasals (as an instance of a manner distinction) and the labial and alveolar distinction (as an instance of a distinction in place). The results showed a near categorical perception for both features.
Although the within category discrimination performance was lower than the between category discrimination. Furthermore, Miller & Eimas (1977) were able to show a right-ear advantage in the processing of both the feature place and manner (See also Studdert-Kennedy & Shankweiler (1969)). Miller & Eimas concluded that the results obtained indicated a similar perceptual processing of the two articulatory features.

In constrained listening conditions, both place and manner have been shown to be perceptually stable. In an investigation of the perceived closeness of phonological features, Singh (1971) showed that nasality was the most perceptually stable feature across spectrally reduced conditions. Singh & Black (1966) furthermore provided evidence that the Place of Articulation feature was second only to nasality in perceptual strength.

It is, therefore, likely that the input given to the child from adult speakers will be perceived most robustly in terms of Manner of Articulation and then, somewhat less robustly, in terms of Place of Articulation. If perception of phonetic properties is a strong driving force behind both phonological development and production of adult-like speech, order of acquisition of adult-like production of a articulatory feature ought to follow the order of perceptual salience and robustness presented above.

### 2.3.2 Progression in phonological features at the phoneme level

Studies of the development of consonants have suggested a progression in the range of feature distinctions actively used in production increase from a limited set to a more elaborate set. At the phonological feature level, Ingram (1976) proposed a group of processes that affects classes of phonemes regardless of segmental context by substitutions at feature level. For example, it has been frequently noted that non-plosive consonants may in the early stages of development be produced perceptually similarly to an adult plosive (e.g. Ingram, 1976 and Grunwell, 1987). This process has been described as an application of a phonological stopping rule in the early stages of production. Examples provided by Ingram (1976) included *shoes* being pronounced as */ʃuː/ and *zebra* produced as */dʒebra/*. Ingram (1976) proposed a three-stage model of development of fricative speech sounds, stopping is applied to developing fricatives and affricates in the early stages of physical manifestation of the fricative sound.

In addition to the stopping rule, a phonological fronting rule, changing the place of consonants produced at a retracted place of articulation to a more fronted position such as labial or dental has been described (Ingram 1976). In the description of this process provided by (Grunwell 1987), velar places of articulation are changed to the more fronted alveolar (e.g. */kɡɑ/* → */tɹɑn/) or palatal (*/kɡɑ/* → */tʃɑn/) places (Grunwell 1987). The application of these phonological processes, may result in an output form with a perceptual quality that is very different from that of the adult target production. Situations may also arise in which the output form of a target word may be reduced to a form that matches the output form of another target word, producing a homonym. As the child progresses, the output forms produced by the child becomes increasingly distinct from each other.
2.3.3 Acoustic correlates of Place of Articulation in plosives

For Place of Articulation of plosives, a number of acoustic correlates of articulation of plosives have been suggested. In an early investigation into the acoustic correlates of plosives, Stevens & Blumstein (1978) considered the relative importance of the burst spectrum at the point of plosive release and the following formant transitions. Using stimuli with either information of formant transitions or a burst at the point of plosive release, Stevens & Blumstein (1978) argued that the stimuli with a burst provided better place identification than the formants-only stimuli. They therefore proposed that the gross shape of the spectrum of a 26 ms window of the burst serves as a primary cue in the distinction of plosive place of articulation and that the following formant transitions serve as supporting cues.

In addition, Stevens & Blumstein (1978) and Stevens (1980) proposed a classification scheme for plosive Place of Articulation based on the shape of the release burst spectra. Velar plosives were described as having narrow spectrum with a prominent mid-frequency peak. Plosives produced at the alveolar place of articulation were described as having a more evenly distributed spectral energy with a rise in energy with increasing frequency (labelled diffuse rising). Labial plosives were described as having an evenly distributed energy in the spectrum similar to that of alveolar plosives, but with either a flat or falling shape (diffuse flat or diffuse falling).

However, Kewley-Port (1983) argued that a quantification based on continuous running speech would serve as a better foundation than the static measure proposed by Stevens & Blumstein (1978) and Stevens (1980). Stevens & Blumstein’s (1978) classification scheme was therefore extended by Kewley-Port (1983), using a set of acoustic criteria describing changes in the spectrum over time. The criteria used were the tilt of the burst, the presence of a late onset of low-frequency energy (present in velar and some alveolar stops) and the presence of a mid-frequency peak (distinctive of velar plosives). Using visual displays of running spectra of the time from plosive release to onset of voicing, the participants in Kewley-Port’s (1983) study correctly identified 78–92% of the cases, depending on speaker. Thus, the classification scheme based to a large degree on the tilt of the burst were shown to be productive in visual judgement of Place of Articulation.

In a followup experiment, Kewley-Port, Pisoni & Studdert-Kennedy (1983) investigated the relationship between the context independent cue of burst spectrum shape and the dynamic and context dependent cues. In their first experiment, Kewley-Port et al. (1983) conducted a gating experiment where zero, one, three, five or seven pitch pulses of voiced plosives were presented to participants. Their results showed that even though CV syllables could be identified by listeners using only the release burst of labial and dental plosives, identification of velar plosives using the only the release burst was below chance level. In addition, the identification scores improved for all places of articulation if a longer portion of the following transition into the vowel was included in the stimulus.

Kewley-Port et al. (1983) argued that the success of subjects to identify Place of Articulation based on the release burst alone might be due to the presence of F1 transitions in the burst itself. In a second experiment, they presented synthetic CV syllables with either a static burst spectrum or a time-varying spectrum similar to a natural transition between consonant and vowel. Using these stimuli, Kewley-Port et al. (1983) were able to show a higher degree of correct place identification for the
time-varying stimuli compared to the stimuli with a static onset spectra. Kewley-Port et al. (1983) concluded, in contrast to the previous statement made by Stevens & Blumstein (1978), that the primary acoustic correlates of PLACE OF ARTICULATION is in the dynamic and context dependent spectral properties after the release, rather than the static and context independent properties of the burst spectrum. The perceptual effect of the burst spectrum identified by Stevens & Blumstein (1978) was argued to be a secondary cue that to a large degree depends on the presence of F₁ transitions.

The results of Kewley-Port et al. (1983) were in agreement with the results from an investigation of non-English speakers. Lahiri, Gewirth & Blumstein (1984) presented a cross-language investigation of the burst shapes obtained from speakers of Malayalam, French and English. Lahiri et al. (1984) reported an inability to differentiate labial and dental plosives based on Stevens & Blumstein's (1978) classification of burst shapes. Instead, Lahiri et al. (1984) proposed that measurements of spectral energy should be made at 1500 Hz and 3500 Hz both at the burst and at the onset of voicing. Using these acoustic measurements, Lahiri et al. (1984) argued that the relative tilt of the spectrum at the burst an at the onset of voicing (spectral tilt change) would afford an acoustic differentiation between labial and alveolar plosives.

Based on the quota between the spectral tilt at the release and the spectral tilt before the vowel, Lahiri et al. (1984) were able to correctly classify plosives in, on average, 87% of the cases for French, 91.6% of the plosives for Malayalam and 96% of the cases in English. Based on these results, Lahiri et al. (1984) argued that the proposed metric could serve as a productive quantification of the difference between plosives with spectral shapes described by Stevens & Blumstein (1978) as diffuse using gross dynamic properties of the plosive.

### 2.3.4 Acoustic correlates of Place of Articulation in fricatives

Forrest, Weismer, Milenkovic & Dougall (1988) suggested that classification of fricatives produced by adults may achieve a 92% or higher classification rate using the first (mean), third (skewness) and fourth (kurtosis) spectral moments. More recently, Jongman et al. (2000) were able to show significant main effects of spectral mean, spectral variance, skewness and kurtosis in the perception of fricatives. Measurements of spectral skewness were able to separate all places of articulation. For spectral mean and variance, with the exception of the bilabial–labiodental (\([\text{f,v}]-[\text{θ,ð}]\)) distinction, all place distinctions were separated by the acoustic cues. Kurtosis was able to distinguish between \([\text{f,v}]\) and \([\text{s,z}]\). These investigations have shown that the static cues of spectral moments are a solid cue to the place distinction in adult speech.

In a study examining dynamic cues to place in fricatives, Jongman et al. (2000) showed a significant effect of the onset frequency of F₂ but no significant effect of the calculated locus equation. Jongman et al. (2000) therefore argued that it is the onset frequency of F₂, not the following F₂ trajectory, that provides the significant effect on perception of place for the fricative. In contrast, Soli (1981) presented results indicating that the frequency of F₂ within the fricatives preceding a vowel varies both with place of articulation of the and with PLACE OF ARTICULATION of the vowel. Although the data presented by Soli (1981) show the largest effect due to the PLACE OF ARTICULATION of the fricative on the F₂ frequency manifested in the fricative, the
effect of vowel context presented indicated that the transition between fricative and vowel may serve as a significant cue to the place distinction.

From a developmental perspective, the cues of spectral moments and the onset frequency of formants have been shown to be weighted differently by children and adults. Nittrouer, Studdert-Kennedy & McGowan (1989) argued that a shift from a focus on the formant cues to the spectral cues from children to adults in the productions of fricatives. Nittrouer et al. (1989) proposed that this shift indicates a transition from a focus on articulatory gestures towards a more developed awareness of the individual segments. Similar results were obtained by Nittrouer & Miller (1997a, 1997b), and Nittrouer & Studdert-Kennedy (1987) from production experiments contrasting child and adult productions.

2.4 Conclusion

This chapter has overviewed previous research in contrast marking and acquisition in plosives and fricatives. This overview motivates the aspects selected for investigation in the acquisition of the Place of Articulation, Manner of Articulation, complexity and voicing in Swedish s+plosive clusters. The aspects selected for investigation using the Umeå Child Consonant Clusters Corpus (UC3 Corpus) that is presented in the Chapter 3 are VOT, Relative Aspiration Amplitude, $f_0$, spectral mean, spectral variance, skewness, kurtosis, spectral tilt change, $F_2$onset frequency and degree of $F_2$ frequency co-articulation.
Chapter 3

Data collection and the UC³ Corpus

One aim of this Thesis is to describe the production development of Swedish word-initial s+plosive clusters. Another is to compare the developmentally intermediate output forms of s+plosive clusters with the output form of homophonic target words. The set of target words selected to study these aspects of developmental speech needed to simulate the cluster reduction, coalescence, cluster simplification and stopping processes. Further, the segmental quality and quantity for the syllable nucleus and coda ought to be identical, and the target words familiar to young children.

The set of target words selected is presented in Table 3.1. The set of structurally simple homophonic target words were selected to provide adult versions of frequently occurring intermediate forms of s+plosive clusters, resulting from applications of the cluster reduction and coalescence processes. In addition, the target words were selected in order to afford investigations of fronting, stopping and velar harmony processes in the syllable onset segments.

Table 3.1: The corpus used in the study. The first column presents the target word containing the consonant cluster to be investigated in initial position. Consonants in the cluster are identified with subscripts 1 and 2. A target word with an aspirated version of the second consonant in the cluster (C₂) is shown in the second column. The third column contains target words with the voiced cognate of C₂ and the forth column shows a target word with an C₁ syllable onset.

<table>
<thead>
<tr>
<th>C₁C₂V(C)</th>
<th>C₂PV(C)</th>
<th>C₂V(C)</th>
<th>C₁V(C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>spik</td>
<td>[spœːk] /spœːk/</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The target word set consisted of four quadruples of the form $C_1C_2V(C)$, $C_2V(C)$ or $C_1V(C)$ and a familiarisation target word /spik/. Using this design, each set of target word quadruples afforded the study of cluster reduction, coalescence and cluster simplification by comparison of the non-adult production of the cluster with the production of a homophonic target cluster. For instance, the process of cluster reduction may be studied by investigating differences between the productions of the target word /skal/ with the output form of /gal/ and /sal/.

Other processes affecting the output form of the cluster may be investigated by comparison of the cluster form /skal/ and a target word with an unaspirated plosive /kal/. Aspiration applied to the unaspirated cluster plosive is an indication of a process other than cluster reduction. Thus, the target word corpus afforded investigation of the output forms of consonant cluster target words with possible substitutions.

3.1 Participants

This section presents the participant recruitment, briefing and identity encoding. The handling of personal and participant identifying information follows the ethical guidelines of Vetenskapsrådet (Vetenskapsrådet 2002).

3.1.1 Criteria

The following set of criteria was used in the selection of experiment participants.

1. Both parents/legal guardians had to have Swedish and only Swedish as their first language. This criterion was to ensure that the models the children were exposed to in the home environment were similar.

2. Both parents/legal guardians should speak a variety of Swedish that does not perceptually deviate from the Umeå variety of Swedish in its use of aspiration and voicing in word initial plosives. This criterion was remove dialect variation as a factor in the analysis of the acoustic data.

3. Participants should not be perceived as having any hearing deficiency or language delay at the onset of the study. This information was obtained through parental/legal guardian interview. This criterion aimed to restrict the set of participants to those with standard language development.

3.1.2 Participant recruitment

In order to have a broad socioeconomic and parental level educational background, four different participant recruitment procedures were adopted. The range of approaches aimed at encouraging parents/legal guardians outside of the university and regardless of an interest in linguistics to consider their children as possible participants for the study in spite of the commitment that participation in the project demanded.

Two of the recruitment approaches were directed at the general public. The first was a local newspaper article that described the project, the general form of the data gathering and provided contact details for interested parents (Lundberg 2001). This newspaper article did not provide any information relating to the objective of the study.
nor the target word set. Thirteen participants, that fulfilled the participant criteria, were recruited as a result of this newspaper article. The second approach that was directed randomly at the general public was the handing out of fliers giving information about the project and contact details at a large trade fair (Nolia) held in Umeå. Two participants, that fulfilled the participant criteria, were recruited in this way.

The third approach built upon the social networks of both university employees and the parents of recruited participants. Six participants, that fulfilled the participant criteria, were recruited via social networks. The fourth approach was to inform students taking linguistics courses about the project; one participant, that fulfilled the participant criteria, was recruited in this way. The parent of the participant was following a part-time course in Child Language Development.

In total, 25 parents/legal guardians indicated interest in the study. Two children were excluded as one or both of their parents did not have Swedish as their first language, and one was excluded as their mother was a native speaker of Finland-Swedish, a variant of Swedish which differ from Swedish spoken in Umeå in terms of application of aspiration. Thus, 22 participants were judged as meeting the set criteria and were recruited for the study. Table 3.2 overviews the participants providing information about their age at their first recording, their age at their final recording, their gender and how they were recruited.

### 3.1.3 Pre- and post investigation briefing

The parents/legal guardians of the participants were informed at the onset of the study that the researcher was not a licensed speech therapist and that the study should not in any way be associated with clinical examinations.

The guardians were also informed that participation in the investigation was voluntary and that they could stop participating in the data collection and withdraw from the study completely at any point. This information was provided in conjunction with the initial recording session. The guardians present were also informed that the child’s name would be coded in reports that would be written based on the collected speech recordings.

This information was repeated in written form after the final recording Session. The guardian present at the final session was then asked to sign a document confirming that they had received and understood this information and that they gave consent to the recordings of their child being used in research and teaching. The consent form was divided into two parts, one for research purposes and one for teaching purposes, thus enabling the guardian to consider these domains of use of the material separately.

### 3.1.4 Participant identity encoding process

After the entire set of recordings had been collected, the participants were given a unique identifier which was unrelated to the participants’ name. The identifier consisted of an initial ‘F’ for female participants and ‘M’ for male participants followed by a number. Thus, the participants were assigned identifiers M1 to M13 and F1 to F10. Neither the participants nor their legal guardians were provided with information regarding which identifier was assigned to which participant.
Table 3.2: Participant information. The columns indicate (from left to right) Participant age at the first recording session, age at the last recording session and how the participant was recruited. Ages in the second and third columns are in months.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Age at onset</th>
<th>Age at completion</th>
<th>Gender</th>
<th>Recruited</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>23</td>
<td>40</td>
<td>Female</td>
<td>Newspaper article</td>
</tr>
<tr>
<td>F2</td>
<td>23</td>
<td>33</td>
<td>Female</td>
<td>Newspaper article</td>
</tr>
<tr>
<td>F3</td>
<td>41</td>
<td>47</td>
<td>Female</td>
<td>Social network</td>
</tr>
<tr>
<td>F4</td>
<td>26</td>
<td>37</td>
<td>Female</td>
<td>Newspaper article</td>
</tr>
<tr>
<td>F5</td>
<td>23</td>
<td>35</td>
<td>Female</td>
<td>Newspaper article</td>
</tr>
<tr>
<td>F7</td>
<td>27</td>
<td>34</td>
<td>Female</td>
<td>Newspaper article</td>
</tr>
<tr>
<td>F8</td>
<td>18</td>
<td>30</td>
<td>Female</td>
<td>Trade fair</td>
</tr>
<tr>
<td>F9</td>
<td>32</td>
<td>43</td>
<td>Female</td>
<td>Trade fair</td>
</tr>
<tr>
<td>F10</td>
<td>25</td>
<td>35</td>
<td>Female</td>
<td>Newspaper article</td>
</tr>
<tr>
<td>M1</td>
<td>19</td>
<td>32</td>
<td>Male</td>
<td>Linguistics course</td>
</tr>
<tr>
<td>M2</td>
<td>21</td>
<td>33</td>
<td>Male</td>
<td>Newspaper article</td>
</tr>
<tr>
<td>M3</td>
<td>42</td>
<td>47</td>
<td>Male</td>
<td>Newspaper article</td>
</tr>
<tr>
<td>M4</td>
<td>22</td>
<td>36</td>
<td>Male</td>
<td>Social network</td>
</tr>
<tr>
<td>M5</td>
<td>40</td>
<td>51</td>
<td>Male</td>
<td>Newspaper article</td>
</tr>
<tr>
<td>M6</td>
<td>21</td>
<td>33</td>
<td>Male</td>
<td>Newspaper article</td>
</tr>
<tr>
<td>M7</td>
<td>20</td>
<td>30</td>
<td>Male</td>
<td>Newspaper article</td>
</tr>
<tr>
<td>M8</td>
<td>18</td>
<td>30</td>
<td>Male</td>
<td>Newspaper article</td>
</tr>
<tr>
<td>M9</td>
<td>26</td>
<td>36</td>
<td>Male</td>
<td>Newspaper article</td>
</tr>
<tr>
<td>M10</td>
<td>19</td>
<td>30</td>
<td>Male</td>
<td>Social network</td>
</tr>
<tr>
<td>M11</td>
<td>18</td>
<td>30</td>
<td>Male</td>
<td>Social network</td>
</tr>
<tr>
<td>M12</td>
<td>29</td>
<td>42</td>
<td>Male</td>
<td>Social network</td>
</tr>
<tr>
<td>M13</td>
<td>19</td>
<td>30</td>
<td>Male</td>
<td>Social network</td>
</tr>
</tbody>
</table>

3.1.5 Treatment of participant background information

The participants’ names as well as the names of the legal guardians and contact information were collected into a key table which was kept separate from the data. This key table was destroyed after the study was completed and the participants had received a copy of their recordings.

3.2 Data collection

3.2.1 Recording procedure

The recordings were conducted in a sound treated room at the Department of Philosophy and Linguistics, Umeå University, Sweden, using a mounted Brüel & Kjær Type 4012 microphone connected to a Panasonic SV-3700 DAT recorder. The sampling frequency of the DAT recordings were 48,000 Hz.

Elicitation of the target words were made by the child’s parent or accompanying guardian, with the researcher present in the studio when it was judged that it would not
influence the number of successful productions. Hand drawn, black and white picture prompts (approximately 10x10 cm) were used as a visual cue to the target word. The target word for each picture was given to the parent in the upper left corner of the picture prompts.

At the onset of the study, the parents/guardians were trained by the researcher in the general methodology that they should use when eliciting a production. The parents/guardians were informed that they should aim for spontaneous productions without a preceding adult model, as in the following examples.

- “Vad finns på bilden?”
  *What’s on the picture?*

- “Vad gör hunden?”
  *What does the dog do?*

In the case that this approach did not work, the parent was instructed to aim for a production which was not immediately preceded by the adult model, as in the following example.

- “Står hunden eller hoppar den?”
  *Is the dog standing or is he jumping?*

In cases where these strategies did not result in production of the target word, the parents/guardians were instructed to elicit a production in the form of a direct repetition as in the following example.

- “Såg skal!”
  *Say peal!*

The aim for each recording session was two elicitations of the sequence of 21 target words.

### 3.2.2 Data inclusion criteria

A number of criteria had to be fulfilled for a production of a target word to be included in the material to be analysed.

One, the nature of the intended word had to be known, implying that the production must have been made in response to a picture prompt or in discussion with the guardian which included enough detail for there to be no ambiguity as to nature of the intended word.

Two, the production should not co-occur in time with another noise sources that would influence the ability for the human transcriber to isolate the acoustic properties of the child’s productions. Therefore, all productions that coincided with an utterance by the guardian during elicitation or with other noise sources were removed from the material.

Three, the produced word had to be perceived by the transcriber as uttered in the tone of voice and speaking rate of casual, everyday speech. As a consequence, utterances that were produced by whispering, produced with a very weak voice or shouted were not included in the material. Furthermore, utterances that were perceived as prolonged or shortened were not included in the analysed material.
3.3 Markup

The productions to be investigated were extracted from the collected material using the markup procedure presented in section 3.3.1. The first step involved extraction of the utterances and marking them according to context. In the second step, the segmental contents of the utterances were transcribed phonetically. In the third step, the time of acoustic events were marked.

3.3.1 Markup procedure

The first stage of the markup procedure consisted of marking the child’s productions in the recorded session. For each marked word, a code sequence was assigned giving information regarding the degree of spontaneousness in the production, the context in which the production was made together with in which pass through the word set the production was made. The codes for spontaneity and degree of context used in the production extraction phase are presented in table 3.3 and table 3.4 respectively.

The second stage of the markup procedure consisted of finding the point of onset and offset of the segments in the child’s production and assigning a phonetic transcription of that segment based on an auditory analysis. The criteria used during segmentation are presented in Section 3.3.3 for each type of phonetic segment labelled. Furthermore, each syllable nucleus was marked as being primary stressed, secondary stressed or unstressed. The lexical accent of the produced word was marked on the nucleus of the primary stressed syllable.

In the third and final stage, the timing of acoustic events were marked for the syllable onset of the word. The acoustic events used for this investigation were onset of initial of the prevocalic consonant, onset of $f_0$, onset of $F_2$ and the time where $F_2$ reached a steady state. The markup criteria for these acoustic events are described in Section 3.3.3.

3.3.2 Markup configuration

The acoustical markup was conducted in the WaveSurfer transcription environment (Sjölander & Beskow 2000) and with components corresponding to an extended version of the recommendations of Buder (1996), i.e. a waveform, a wideband spectrogram, a narrow band spectrogram, a pitch contour as well as a spectrum plot.

The wideband spectrogram was created using a default analysis bandwidth of 400 Hz, a frequency range of 20 Hz to 22,050 Hz and a pre-emphasis factor of 0.97. In the case when mean $f_0$ for a given speaker exceeded 200 Hz, the analysis bandwidth was adjusted so that the analysis bandwidth exceeded the $f_0$ frequency. The spectrogram setup was designed to give information regarding formant structure by smearing the spectrogram enough for the harmonic structure to be less prominent.

Formant tracks were computed using the built-in ESPS tracking algorithm based on LPC analysis with 3 formants conducted on the signal downsampled to 12 kHz and using an LPC order of 11. These parameters have been evaluated by Lee et al. (1999) for automatically extraction of formant frequency tracks in child speech databases. Comparing the automatically computed formant values of 96 vocalic segments with manual estimations of the formants of children aged 5, 8 and 11 and adults, Lee et al. (1999) reported a mean deviation of 43.6 Hz for $F_1$ and 92.4 Hz for $F_2$. 

Table 3.3: Spontaneity labels used when extracting words productions from the original recording. For each label, an example elicitation is provided using the target word “spik”.

<table>
<thead>
<tr>
<th>Spontaneity</th>
<th>Label</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Direct Repetition</td>
<td>The production is made as a direct repetition of a preceding adult model. An example elicitation under this condition is “Say ‘spik’!”</td>
</tr>
<tr>
<td>2</td>
<td>Delayed Repetition</td>
<td>The target word is presented in the form of an adult model but there is a time lag, where other information is presented to the child, between the model and the child’s production. Example: “This is a ‘spik’. Daddy sometimes hits it with a hammer. Can you say that word?”</td>
</tr>
<tr>
<td>3</td>
<td>Sentence elicitation</td>
<td>The target word is elicited by a sentence with the target in non-final position. For example: “Is it a ‘spik’ or a ‘snake’?”</td>
</tr>
<tr>
<td>4</td>
<td>Partial Sentence</td>
<td>Elicitation is made using a non-complete sentence containing part of the target word. For instance, “Look! It’s a ‘s...’.”</td>
</tr>
<tr>
<td>5</td>
<td>Spontaneous</td>
<td>The target word is elicited without giving an adult model for the target word. For instance, “What’s on the picture?”</td>
</tr>
</tbody>
</table>

Table 3.4: Context labels used when extracting words productions from the original recording. Examples using the target word “skal” is provided.

<table>
<thead>
<tr>
<th>Context</th>
<th>Label</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Single Word</td>
<td>“skal”</td>
</tr>
<tr>
<td>2</td>
<td>Word with article</td>
<td>“skalet”</td>
</tr>
<tr>
<td>3</td>
<td>Compound words</td>
<td>“bananskal”</td>
</tr>
<tr>
<td>4</td>
<td>In sentence</td>
<td>“Det är ett skal”</td>
</tr>
<tr>
<td>5</td>
<td>In compound in sentence</td>
<td>“Det är ett bananskal”</td>
</tr>
</tbody>
</table>
These formant tracks were superimposed on the wideband spectrogram as a guide for formant movements. Where the formant automatically computed formant tracks and the judgement of the transcriber disagreed, the formant tracks were adjusted according to the analysis of the human transcriber.

The narrow band spectrogram was created using an analysis bandwidth of 75 Hz and a frequency range of 20 to 8000 Hz. Following Buder (1996), this setting was chosen in order to facilitate the localisation of formant tracks where the bandwidth chosen in the wideband spectrogram inhibited a clear view of adjacent formants. Furthermore, the setting provided information about a productions $f_0$ variations so that the analysis bandwidth of the wideband spectrogram could be correctly adjusted as per the description above.

The pitch track used was calculated using the same parameters as for the formant tracks. For the test sample evaluated in their study, Lee et al. (1999) reported a mean deviation of 7.6 Hz (with a standard deviation of 23.8 Hz) for $f_0$.

### 3.3.3 Markup criteria

The segmental quality was determined by the transcriber and noted in X-SAMPA notation. In order to increase transcription accuracy and consistency, recorded representations of the IPA speech sounds (International Phonetic Association 1999) were used as references in the judgement of phonetic quality. Diacritic descriptions of voicing quality, aspiration, breathiness of voice and vowel place compared to the traditional feature specification of the transcribed speech sound were added to the segmental transcriptions.

**Markup of vowels** Vowel onset was marked at the first zero crossing in the waveform corresponding to an onset of audible vocalic quality. Vowel offset was marked at the first zero crossing after the the quality of the vowel was no longer audible. Diphthongs were marked as one segment and transcribed using a concatenation of the transcriptions of the two vowel qualities.

**Markup of fricatives** Start time and stop time for fricatives were set at the first zero crossing at the onset and offset of a audible frication corresponding to an aperiodic portion of the waveform and a mid- to high-frequency component in the spectrogram.

**Markup of plosives** For initial plosives, time of plosive onset was marked at the first zero crossing in the waveform corresponding to an acoustic transient. For non-initial plosives, time of onset was marked at the onset of the silent interval indicating oral closure. The end point was marked at the time of the first zero crossing in the waveform before the onset of audibility of the quality of the following segment.

The acoustic correlates of voicing/aspiration and **Place of Articulation** were marked independently of the segmental boundaries. These features include $f_0$ onset time, plosive release, onset time of $f_0$, onset and offset time of post-release friction, onset time of $F_2$ and time when $F_2$ reached steady state. The markup criterion used for these acoustic events are presented below.
**Plosive release** The release of the plosive was marked at the first zero crossing corresponding to a transient in the waveform. The transient should have been preceded by a silent interval caused by articulatory closure. See figure 3.1 (left panel).

**Time of voice onset** The time of onset of F₀ was marked as the zero crossing at the onset of periodicity in the waveform which coincided with the approximate time of onset of the lowest frequency harmonic visible in the narrow band spectrogram. See figure 3.1 (right panel). Priority was given to the information provided by the waveform, as suggested by Francis, Ciocca & Yu (2003) who showed less influence of aspiration noise in identification of voicing onset from the waveform compared to spectrograms.

**Friction after plosive release** The section of the speech signal corresponding to the presence of a friction following a plosive release was marked starting at the first zero crossing before the start of an aperiodic section in the waveform. The end time of the friction was marked at the first zero crossing after the aperiodic portion in the waveform corresponding to the offset of an aperiodic source in the frequency region around F₂ and F₃ in a wideband spectrogram. An example markup is presented in Figure 3.2.

**Time and frequency of onset and steady-state of F₂** The onset time of F₂ was marked at the time when the formant structure was visible both in the wideband and in the narrow-band spectrogram. In difficult cases, the onset time of a formant was obtained by comparing spectral slices at different points in the speech signal. In addition to the time of F₂ onset, the time when F₂ reached a steady state was marked. In the case of diphthongs, the first plateau visible in the spectrogram was marked as the steady state. The frequency of a formant at the time marked as onset for the formant was extracted from automatically computed and manually controlled formant track estimations (see Section 3.3.2).

![Figure 3.1](image-url)  
*Figure 3.1: Segmentation criteria used for the markup of plosive release (left picture) and onset of voicing (right picture).*
Figure 3.2: Example markup of post release friction.

Figure 3.3: Histogram of age coverage. Bars indicate the number of participants which were recorded at the AGE, in months, indicated on the abscissa.
Figure 3.4: Histogram of the distribution of analysed utterances across Age (in months). The bars indicate the relative frequency of production at that age.

Figure 3.5: Histogram of target word frequency
3.4 Data organisation and data extraction method

The collected speech data was organised into a speech database both by an hierarchical directory structure and by the structure of file names. The directory structure organised recordings by PARTICIPANT and SESSION DATE. A sub-directory of a speaker specific main directory thus contained production made by a specific speaker in a particular session.

The information contained in the directory structure was repeated in the file name convention utilised. The first and second part of the filename contained information about the PARTICIPANT and SESSION DATE. Following SESSION DATE, information concerning the adults present with the child in the session was encoded. Following that, a three-digit sequence indicated the spontaneity of the utterance (using the integer codes of Table 3.3), context (using the integer codes of Table 3.4) and pass through the set of target words. In the last part of the file name, the productions position in the sequence of productions of the specific target word was indicated, followed by an orthographic transcription of the target word. The same naming convention was used in all files derived from the sound file (e.g. transcription files and computed formant and pitch contours). Thus, the naming convention assured an accurate placement of data gathered from the speech file or derived files throughout the data handling and analysis phase of the investigation.

From the hierarchical database, data points were extracted using custom scripts using the Annotation Graph Toolkit (Bird & Liberman 1999). Acoustic data was extracted using the Snack software library, described in Sjölander & Beskow (2000).

3.5 The UC³ Corpus

The data included into the UC³ Corpus were collected from 221 recording sessions. From these sessions, a total of 4732 productions of the target words displayed in Table 3.1 were accepted into the database and analysed. This section presents an overview of the distributions of these productions across participant ages.

The recording sessions were not evenly distributed across subject ages. The AGE distribution of the 221 recordings (Figure 3.3) show that the majority of recordings occurred when the children were between 1;6 and 3 years old. In the right tail of the distribution of recordings, the number of sessions decreases to between one and three.

This skewness in the distribution of recording sessions resulted in a skewness in target word productions. The barplot shown in Figure 3.4 presents the number of extracted productions of the target words at each age (in months). The skewness of the recording sessions was confirmed using a Shapiro-Wilk’s test of normality ($W=0.89, p<0.001$).

Similarly, the distribution shape of Figure 3.4 shows a positive skewness. Thus, the results presented in this Thesis provide a more accurate description of a larger population of children in the 1.6–3 year age range than later ages. Thus, the description of developmental trends within the investigated groups of children should be weighted by the lower frequency of productions at the older ages.

The distribution of productions across target words was also investigated for bias. Figure 3.5 shows the frequency of production for each target word in the form of a barplot. As may be observed in Figure 3.5, word frequencies in the smaller end of
the scale compared to frequencies in the larger end of the scale (skewness = 0.89). A Shapiro-Wilk’s test of normality confirmed that the distribution of frequencies was not normally distributed (W=0.87, p < 0.05). Thus, no bias due to production frequency of the target words was observed in the database. However, a bias due to age at sampling was observed.

3.6 Discussion

This chapter described the data collection, treatment of the participants and the markup of the gathered acoustic material that is collected in the UC³ Corpus. The nature of the database obtained using this investigation procedure was described using descriptive statistics, statistical tests of normality and tests of significant difference.

The results showed a bias due to age; the number of samples in the 1;6–3;0 year age interval was shown to be larger than the later ages. No statistically significant difference was, however, observed in the frequency of production of target words.

Based on the acoustic measurements obtained using the methodology described in this chapter, phonetically relevant derived measures were calculated as defined in Appendix A.
Part II

Voicing contrast in production
Chapter 4

Quantifying Development of Phonological Voicing using Measurements of Voice Onset Time

This chapter presents the Voice Onset Time (VOT) measurements collected from productions made towards the adult plosives [b], [p], [pʰ], [d], [t], [tʰ], [g], [k], [kʰ]. Group developmental trends in terms of frequency distribution across VOT CATEGORY category and the distribution of VOT length within each category are discussed.

In Section 4.1, the raw data is presented divided into categories based on voicing category and PLACE OF ARTICULATION of the target plosive in the syllable onset. In Section 4.2, a data categorisation scheme based on VOT -production length and participant age is presented. The categorisation serves as complement to the categorisation presented in Section 4.1. In Section 4.3 the distribution of VOT productions across the established categories are presented and examined for developmental trends, and the frequency of occurrence of prevoiced and long lag productions are investigated and discussed. In Section 4.4, VOT measurement lengths within each category are investigated in terms of distribution characteristics. In Section 4.5, differences in produced VOT due to the target voicing category and age of the child are tested statistically for significant group differences, and then in Section 4.6 the findings of the chapter are discussed.

4.1 Presentation of the VOT data set

The Voice Onset Time (VOT) data set in the UC3 Corpus contains the measurements extracted from target words with an initial plosive, or an s+ plosive consonant cluster in the syllable onset (/qal/, /kal/, /kal/, /bör/, /spör/, /pör/, /dör/, /sör/, /kö/, /bör/, /sör/ and /pör/). A large number of the productions made by the children neither included a plosive nor were produced in a way that afforded measurement of VOT according to the criteria established in Chapter 3.
A total of 2825 VOT measurements were obtained from the productions. The distribution of these productions across target plosive place and voicing category is shown in the form of a balloon plot in Figure 4.1. In this balloon plot, the relative frequency of each combination of the two categories is indicated by the size of the balloon and the frequency of production is presented by the integer in the centre of the cell.

Figure 4.2 illustrates the VOT data set in the form of a trellis plot. In this type of plot, the data is divided into subgroups by factors and the dependent variable is plotted against the independent variable for the data within the subgroup. In Figure 4.2, VOT measurements are plotted against the Age of the participants after having been divided into subsamples by the factors voicing type of the target (columns) and Place of Articulation (rows). The participant from whom the data point was collected is identified in the plot using the set of symbols listed in the figure key. In the speaker figure legend, male speakers are identified by an ‘M’ followed by an integer and female speakers are identified by ‘F’ before the integer.

In order to facilitate detailed investigation of the distribution of VOT measurements in relation to their perceptual properties as well as any developmental trends, the categorisations due to properties of the target plosive were supplemented with categories for VOT length and the Age of the participant.

<table>
<thead>
<tr>
<th></th>
<th>Voiced</th>
<th>Voiceless unaspirated</th>
<th>Voiceless aspirated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labial</td>
<td>385</td>
<td>680</td>
<td>428</td>
</tr>
<tr>
<td>Dental</td>
<td>208</td>
<td>277</td>
<td>246</td>
</tr>
<tr>
<td>Velar</td>
<td>202</td>
<td>209</td>
<td>190</td>
</tr>
</tbody>
</table>

Figure 4.1: Number of Voice Onset Time measurements by Place of Articulation (columns) and Voicing Category (rows) of the target plosive.

4.2 Data categorisation scheme

The data plot presented in Figure 4.2 gives information about the length of the produced VOT. However, it does not include a judgement regarding the perceptual quality of the production, nor whether the production agrees with adult model productions in the ambient language. A categorisation scheme based on VOT length categorised
Figure 4.2: Voice Onset Time measurements by Age of the participants. The data set is divided into panels by Place of Articulation (rows) and Voicing Category of the target plosive (columns). The Participant making the production is indicated by the plotting characters given in the figure legend. Female speakers are indicated by an ‘F’ and male speakers are indicated by an ‘M’.
by perceptual categories of adult speech was therefore used in the description of children's productions and the participants were grouped into AGE CATEGORIES in order to facilitate evaluation of the effect of the AGE variable.

4.2.1 Perceptually based VOT Category

In the investigation of the produced VOT, two boundaries were used: a boundary between prevoiced and short lag productions and a boundary between short lag and long lag productions. The boundary for the PREVOICED category was set at zero milliseconds (ms). The boundary between short lag and long lag region was set at 40 ms.

Selection of the boundary between the short lag and long lag VOT categories is problematic for speech gathered from children. As discussed in Section 2.2.1, the perceptual boundary between the short lag category, signalling no aspiration, and LONG LAG category, signalling aspiration, has been shown to be dependent on PLACE OF ARTICULATION of the plosive. In the data set, the PLACE OF ARTICULATION of the output form was frequently judged by the transcriber not to be identical to that of the adult model. In productions failing to meet the adult place of articulation goal, the question arises whether the VOT measurements should be divided into categories by the boundary appropriate for the target or for the output form produced. In the latter case, the selection of an appropriate boundary for the VOT category will depend on the cause of the place and VOT correlation.

If the correlation is caused by articulatory factors, the VOT value chosen for each production would have to be dependent on the place where the plosive is produced. In this scenario, the process of selecting an appropriate boundary between short and long VOT lag category is simple, as it can be made based on the output form of the child. However, Sawusch & Pisoni (1974) proposed that the VOT produced in adult speech may serve as cue for PLACE OF ARTICULATION. For instance, a VOT production in the 35–40 ms region may be a signal of an aspirated plosive produced at a fronted PLACE OF ARTICULATION since a velar plosive would have been produced further from the boundary in order to assure a distinction between aspirated and unaspirated voiceless plosives. The categorisation of VOT is, in this case, dependent on the child’s level of understanding for VOT as a cue for signalling PLACE OF ARTICULATION. Since the nature of the underlying representation of the plosive in the child’s phonology is not known, it is not possible to know whether the VOT measurement should be categorised by the place of the plosive perceived by the adult or the plosive attempted in the production.

The investigation of the VOT produced by children presented in this thesis focuses on the transition from the unmarked SHORT LAG production observed by Kewley-Port & Preston (1974) for both aspirated and unaspirated plosives towards LONG LAG productions for aspirated target plosives. The presentation of the results, therefore, focuses on the productions in the PREVOICED and LONG LAG categories. As a more restrictive classification into these categories is more informative of the progression of the investigated children than a less restrictive categorisation, the most restrictive boundary of the 40 ms velar boundary was chosen as the boundary for the classification into categories by VOT. This ensures that the risk of reporting signs of development due to miss-classification is minimised.
4.2.2 Division by Age Category

The small number of productions made in each session by each child presents a problem when analysing the data. Small sample sizes in each session make estimates of sample statistics such as mean, median and variance less reliable. This makes it difficult to interpret the data in terms of broad developmental trends in produced VOT length and VOT variability due to age. The productions were, therefore, grouped into six categories based on the age of the child. The productions were divided into categories spanning 6 months: \(18 \leq \text{age} < 24 \text{ months}, 24 \leq \text{age} < 30 \text{ months}, 30 \leq \text{age} < 36 \text{ months}, 36 \leq \text{age} < 42 \text{ months}, 42 \leq \text{age} < 48 \text{ months} \) and \(48 \leq \text{age} < 54 \text{ months} \).

This approach created reasonable sample sizes without hindering the analysis of the developmental trend by creating too few groups. Together, the categorisation by target Voicing Category Age Category and VOT Categories were used in the description of the group data distribution.

4.3 Distribution of VOT across Voicing Categories

This section presents an analysis of the frequency of production within each cell created by the factors target Voicing Category, VOT length category and age category. The analysis is presented by a stepwise division of the data set into increasingly smaller subsamples.

Figure 4.3 presents the distribution of VOT measurements across voicing and VOT categories in the form of a mosaic plot. In this plot, the rectangles display the relative frequency of productions in the prevoiced, short lag and long lag categories (rows) and for each target Voicing Category (columns). In this initial categorisation of the data set, a strong preference towards a short lagged VOT production is observed. This preference extends to both voiceless unaspirated and voiced target plosives. This conforms with Kewley-Port & Preston’s (1974) proposal suggesting a predominance of short lagged VOT. Even though there is also a strong presence of short lagged VOT in the productions of aspirated plosives, the statistical mode of production for this target is, however, in the long lag VOT category. Since productions are made both in the proposed default category and in the adult-like category, the results suggest a transition from one production pattern to another within the investigated time frame (1;6–4;6).

For prevoiced productions, the agreement between the predictions made by Kewley-Port & Preston (1974) and the group distributions observed here is not as evident. As shown in the distribution for productions towards unaspirated targets, prevoicing occurs frequently, even though Kewley-Port & Preston (1974) argued that they involve more complex articulator coordination than for short lag productions. The presence of prevoiced productions is, however, in agreement with the work on children’s voicing development by Catts & Kamhi (1984) and Macken & Barton (1978). The categorisation of the data presented in Figure 4.3 indicates a partial deviation from the proposed default category towards both the prevoiced and the long lag categories, depending on the Voicing Category of the target. It is proposed that this observed bimodality in the data is part of a developmental trend.
### Figure 4.3: Distribution of VOT Categories across target Voicing Category for the entire data set. Row visualizes the relative frequency of prevoicing, short lag and long lag for each of the target plosive categories Voiced, voiceless unaspirated and voiceless aspirated.

<table>
<thead>
<tr>
<th></th>
<th>Voiced</th>
<th>Voiceless unaspirated</th>
<th>Voiceless aspirated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevoiced</td>
<td>159</td>
<td>158</td>
<td>28</td>
</tr>
<tr>
<td>Short lag</td>
<td>590</td>
<td>821</td>
<td>309</td>
</tr>
<tr>
<td>Long lag</td>
<td>78</td>
<td>232</td>
<td>549</td>
</tr>
</tbody>
</table>

### 4.3.1 Progression from short lag to long lag VOT

The statistical mode of the VOT productions for the aspirated plosive target type presented in Figure 4.3 lies within the long lag category. The sub plots of figures 4.4 and 4.5 show the distribution of Voice Onset Time productions across VOT length categories and target voicing type in one balloon plots per six-month Age interval. The frequencies of occurrence between ages 18–36 months are shown in Figure 4.4 and the frequency distributions for ages 36–54 months in Figure 4.5. The frequency value for each category combination is presented in the centre of each cell. The size of the balloon indicates the relative size of the frequencies.

The balloon plots in Figures 4.4 and 4.5 indicate that the distribution of VOT presented in Figure 4.3 may not be used as a model of the entire Age category. An age dependency in the distribution across cells is observed. At the earliest ages, the uppermost plot of Figure 4.4, a clear preference is found for short lag VOT across all target voicing categories. In the plot displaying data collected between 24–30 months, aspirated plosive targets are produced with a long lag VOT almost as frequently as with a short lag. For unaspirated targets, productions are most frequently produced in the short lag category. However, according to the frequencies presented in Table 4.1, the relative frequency of unaspirated voiceless targets produced with a long lag VOT more than doubled between 18-14 months and 24–30 months.

From the presentation of the production distribution across the short lag and long lag categories, a clear progression is observed. Productions are made in the short lag category in the initial Age Category; thereafter, increasingly more frequent productions in the adult-like category are observed. Thus, the frequency counts in cells created by the factor Age Category and VOT Category provide a clear picture...
### Age = 18 – 24 months, N = 400

<table>
<thead>
<tr>
<th></th>
<th>Voiced</th>
<th>Voiceless unaspirated</th>
<th>Voiceless aspirated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevoiced</td>
<td>9</td>
<td>30</td>
<td>8</td>
</tr>
<tr>
<td>Short lag</td>
<td>100</td>
<td>116</td>
<td>87</td>
</tr>
<tr>
<td>Long lag</td>
<td>19</td>
<td>16</td>
<td>25</td>
</tr>
</tbody>
</table>

### Age = 24 – 30 months, N = 851

<table>
<thead>
<tr>
<th></th>
<th>Voiced</th>
<th>Voiceless unaspirated</th>
<th>Voiceless aspirated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevoiced</td>
<td>52</td>
<td>36</td>
<td>6</td>
</tr>
<tr>
<td>Short lag</td>
<td>158</td>
<td>264</td>
<td>99</td>
</tr>
<tr>
<td>Long lag</td>
<td>30</td>
<td>70</td>
<td>138</td>
</tr>
</tbody>
</table>

### Age = 30 – 36 months, N = 830

<table>
<thead>
<tr>
<th></th>
<th>Voiced</th>
<th>Voiceless unaspirated</th>
<th>Voiceless aspirated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevoiced</td>
<td>44</td>
<td>49</td>
<td>9</td>
</tr>
<tr>
<td>Short lag</td>
<td>176</td>
<td>244</td>
<td>51</td>
</tr>
<tr>
<td>Long lag</td>
<td>20</td>
<td>47</td>
<td>191</td>
</tr>
</tbody>
</table>

Figure 4.4: Distribution of VOT Categories across target voicing and aspiration quality at Ages 18–24 months (top), 24–30 months, (middle) and 30–36 months (bottom). Number of productions for each target plosive Voicing Category and VOT Category is indicated by a frequency count in the centre of the cell. Balloons in the centre of the cell provides a visual impression of the magnitude differences between the cells.
### Age = 36 – 42 months, N = 352

<table>
<thead>
<tr>
<th>Prevoiced</th>
<th>Voiced</th>
<th>Voiceless unaspirated</th>
<th>Voiceless aspirated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15</td>
<td>22</td>
<td>2</td>
</tr>
<tr>
<td>Short lag</td>
<td>69</td>
<td>101</td>
<td>26</td>
</tr>
<tr>
<td>Long lag</td>
<td>11</td>
<td>25</td>
<td>81</td>
</tr>
</tbody>
</table>

### Age = 42 – 48 months, N = 303

<table>
<thead>
<tr>
<th>Prevoiced</th>
<th>Voiced</th>
<th>Voiceless unaspirated</th>
<th>Voiceless aspirated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>22</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Short lag</td>
<td>52</td>
<td>49</td>
<td>24</td>
</tr>
<tr>
<td>Long lag</td>
<td>5</td>
<td>66</td>
<td>76</td>
</tr>
</tbody>
</table>

### Age = 48 – 54 months, N = 187

<table>
<thead>
<tr>
<th>Prevoiced</th>
<th>Voiced</th>
<th>Voiceless unaspirated</th>
<th>Voiceless aspirated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>17</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>Short lag</td>
<td>35</td>
<td>47</td>
<td>22</td>
</tr>
<tr>
<td>Long lag</td>
<td>3</td>
<td>8</td>
<td>40</td>
</tr>
</tbody>
</table>

Figure 4.5: Distribution of VOT categories cross target voicing and aspiration quality at Ages 36–42 months (top), 42–48 months, (middle) and 48–54 months (bottom). Number of productions for each target plosive Voicing Category and VOT Category is indicated by a frequency count in the centre of the cell. Balloons in the centre of the cell provides a visual impression of the magnitude differences between the cells.
Table 4.1: Percentage of productions of each of the VOT CATEGORIES prevoiced, short lag and long lag divided into categories by target Voicing Category and the Age Category at which the productions were made.

<table>
<thead>
<tr>
<th>Age category</th>
<th>VOT category</th>
<th>Voiced unaspirated</th>
<th>Voiceless unaspirated</th>
<th>Voiceless aspirated</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 - 24</td>
<td>Prevoiced</td>
<td>7.6</td>
<td>18.5</td>
<td>6.6</td>
</tr>
<tr>
<td></td>
<td>short lag</td>
<td>84.7</td>
<td>71.6</td>
<td>72.5</td>
</tr>
<tr>
<td></td>
<td>long lag</td>
<td>7.6</td>
<td>9.8</td>
<td>20.8</td>
</tr>
<tr>
<td>24 - 30</td>
<td>Prevoiced</td>
<td>21.7</td>
<td>9.7</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>short lag</td>
<td>65.8</td>
<td>71.3</td>
<td>41.0</td>
</tr>
<tr>
<td></td>
<td>long lag</td>
<td>12.5</td>
<td>18.9</td>
<td>56.4</td>
</tr>
<tr>
<td>30 - 36</td>
<td>Prevoiced</td>
<td>18.3</td>
<td>14.1</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>short lag</td>
<td>73.3</td>
<td>71.9</td>
<td>20.3</td>
</tr>
<tr>
<td></td>
<td>long lag</td>
<td>8.3</td>
<td>13.8</td>
<td>76.1</td>
</tr>
<tr>
<td>36 - 42</td>
<td>Prevoiced</td>
<td>15.8</td>
<td>14.9</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>short lag</td>
<td>72.6</td>
<td>68.2</td>
<td>23.8</td>
</tr>
<tr>
<td></td>
<td>long lag</td>
<td>11.5</td>
<td>16.9</td>
<td>74.3</td>
</tr>
<tr>
<td>42 - 48</td>
<td>Prevoiced</td>
<td>27.8</td>
<td>6.5</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>short lag</td>
<td>65.8</td>
<td>39.8</td>
<td>23.7</td>
</tr>
<tr>
<td></td>
<td>long lag</td>
<td>6.3</td>
<td>53.6</td>
<td>75.2</td>
</tr>
<tr>
<td>48 - 54</td>
<td>Prevoiced</td>
<td>30.9</td>
<td>20.3</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>short lag</td>
<td>63.6</td>
<td>68.1</td>
<td>34.9</td>
</tr>
<tr>
<td></td>
<td>long lag</td>
<td>5.4</td>
<td>11.5</td>
<td>63.5</td>
</tr>
</tbody>
</table>

of the progression from default forms towards the adult-like norm. It should, however, be noted that the observed progression towards the adult output form may only be considered a sign of both phonological and articulatory development if it can be shown that this pattern is not found for unaspirated targets.

4.3.2 Place of Articulation and Age dependency in the manifestation of voiceless unaspirated plosives

The progression observed in aspirated plosives from the short lag category to the long lag category does not generally apply to unaspirated targets; most of the unaspirated targets were produced in the short lag category. However, there is a brief period of increase in long lag VOT production between 42–48 months. In this age period, a majority (53.7%) of the voiceless unaspirated productions are manifested with a VOT within the Long lag category. The rate of production of voiceless unaspirated targets at 42–48 months is more similar to aspirated targets produced at the same age than unaspirated targets produced before or after this age category. This increase in long lag productions, that violates the adult norm, appears to be dependent on Place of Articulation.

Inspection of the data plot in Figure 4.3 suggests that the change towards long lag productions occurs primarily in labial targets. The confusion of aspirated and unaspirated targets observed only in labial plosives could be due to the adult versions of the
un aspirated and aspirated labial plosives being less perceptually distinct than plosives produced at other places of articulation. In the data presented by Koenig (2000), 5 year old children produced longer average VOT for \( [t^{h}] \) than for \( [p^{h}] \), suggesting that labial plosives form an acoustically closer aspirated-unaspirated pair. This would make these target forms likely candidates for confusion in aspiration due to the lack of perceptual distinctiveness. However, the age at which this similarity in the use of aspiration occurred is limited. The confusion is neither observed in the productions made in this study’s older age groups nor in Koenig’s (2000) study of the aspiration of 5-year old children. Thus, if the productions observed are due to perceptual properties, one could argue that the confusion should be more pronounced in the earliest stages where a long lag production is attempted. However, in the study presented in this Thesis, the confusion arises 18 months after the stage where more than 50% of the productions of aspirated plosives are made in the long lag category. Thus, an interpretation of the data in terms of acoustical distinctiveness is less plausible.

It has been shown that VOT data from labial plosives creates a period of perceptual confusion in the output forms of the children in the UC3 Corpus. However, it is clear that the application of the long lag production pattern is not limited to aspirated plosives. Thus, the data discussed here serves as an example of an over-generalisation of a production pattern to output forms where it does not agree with the adult norm.

### 4.3.3 On the frequency of prevoicing production across target voicing categories

The literature on prevoicing development presents a picture of individual children either decreasing their prevoicing production frequency throughout development, or showing a peak in prevoicing production in intermediate sessions. However, the nature of the investigations make it difficult to assess whether this developmental trend is valid for a larger group of children, since the number of children participating in the investigations were very small and the sampling intervals were far apart. Due to the nature of the data, it is difficult to detect detailed developmental trends in the use of prevoicing.

In this section, the progression in prevoicing occurrence for the investigated group of Swedish children is be presented. The data set is collected with shorter inter-sampling intervals compared to previous studies, providing increased time resolution in the description, and from a larger group of participants than used by Macken & Barton (1978) and Catts & Kamhi (1984), which increases the likelihood that the data presented provides a good description of pre-voicing use. The description presented in this Thesis, therefore has a greater descriptive power than previous investigations.

The data presented in raw frequencies in Figures 4.4 and 4.5 and in the form of relative frequencies in Table 4.1 indicate a progression in the use of prevoicing that is dependent upon target plosive voicing category (voiced, voiceless unaspirated and aspirated).

For aspirated targets, productions are made in the prevoiced VOT category more frequently, relative to the number of productions made, at the earliest age categories. In older age categories, productions in the prevoiced VOT category become less frequent.

For voiceless unaspirated plosives, no clear trend is observed. At ages 18–24 months, 18.5% of the productions are made in the prevoiced category. Between 24
Voice Onset Time

and 48 months the rate varies between 6.5% and 14.9% of the total number of productions. In the oldest age category, this increased to 20.3%, which only marginally differs from the rate observed in the first age category. Thus, the prevoicing data for voiceless unaspirated plosives does not show signs of a single developmental direction in the frequency of prevoicing use; the production frequency indicates a developmental curve similar to a concave up parabola.

For voiced plosives, however, a single developmental direction is observed. In the age categories following 42 months, there is an increase in relative production frequency of prevoicing from 15.8% at 42 months to 27.8% at 42–48 months and 30.9% at 48–54 months. Thus, prevoicing occurs more frequently in the productions made by older children than the younger children. This progression is not observed in productions of voiceless target plosives. Thus, the acoustic distance between productions of voiced plosives compared to voiceless plosives increased in the final sessions.

This pattern of increase in prevoiced VOT productions for only voiced plosives and the simultaneous increase in long lag productions for aspirated target (see Section 4.3.1) suggests that these increases may be part of a production strategy aimed at separating the output forms of rivaling voicing categories. Hence, the VOT measurements indicate a gradual refinement in the manifestation of the contrast between the rivaling target forms.

Later in development, the number of long lag productions of unaspirated targets is reduced and aspirated targets are kept constant. At the approximately the same point in development, the use of prevoicing occurred frequently relative to the total number of productions in all unaspirated target forms. Between the ages of 42 and 54 months, the frequency of prevoiced productions increased in voiced plosives, reaching 27.9–30.9% of the total number of production with a measurable VOT. An increase in prevoicing production is also observed in the last age category for voiceless unaspirated plosives. Thus, prevoicing is used as a further reinforcement of the contrast between unaspirated target forms and aspirated target forms. Furthermore, the frequency of prevoicing use separates productions of voiced target plosives from voiceless targets and indicates a careful reorganisation of the manifestation of rivaling output forms in order to achieve perceptually distinct manifestations of discovered contrasts. The data, thus, indicate that at 42 months, there is a difference in the production of unaspirated plosives depending on their phonological voicing quality.

4.4 Investigating VOT length across development

A summary of the VOT data in the UC3 Corpus is given in Table 4.2 that presents sample size, mean and median values for each subsample created by the factors VOT category, participant age category and target plosive voicing type. An ‘NA’ label marks cells where mean and median values were not calculated as the cell contains only a single sample.

4.4.1 Means and medians within cells

The mean and median values for productions in the long lag category are larger in the 24–36 months category than the 18–24 months category. Similar mean and median
values are kept throughout the 30–36 months category for all voicing types. For voiceless plosives, this level is sustained for the 36–42 months category, while a decrease is observed for voiced plosives. In the 42–48 months age category, a decrease in mean and median values is observed across all plosive voicing types. In the 48–54 months category, mean and median levels are comparable, or slightly lower, than the values for the 18–24 months category. Thus, a U-shaped developmental curve is observed.

For productions in the short lag VOT interval, the mean and median VOT values change minimally for unaspirated targets across the age categories investigated. For aspirated targets, however, there is a progression from a small, positive VOT to larger VOT values. This pattern of increase in mean and median VOT indicates that even though the productions were made in the adult long-lag category, there is a progression in the distribution towards the long lag interval. This indicates a progression towards adult-like properties in aspirated targets that is not observed in attempted productions of unaspirated plosives.

For productions in the prevoiced VOT category, the results presented in Table 4.2 show a decrease in mean and median VOT for voiceless unaspirated targets over the investigated period. No clear progression is seen in the voiced and voiceless aspirated productions. The small number of samples in the prevoiced productions of aspirated plosives should be considered when interpreting the cell mean as an approximation of the true mean for a larger population since it increases the confidence interval of the mean estimate.

The means and median values discussed above do, however, provide information relating to the central tendency of the data sample in each cell. Previous investigations of the development in motor skills (Eguchi & Hirsh 1969, Kent 1976, Koenig 2000, Koenig 2001, Karlsson, Landgren & Sullivan 2005) have noted that the variability in the distribution of productions may vary with age. Koenig (2001) found that distribution skew of VOT measurements may not be approximated using normal distribution, and in (Koenig 2000) she reported skew of the distribution to be a significant factor in the determination of statistical test of significance of group differences and therefore choose a non-parametric Kruskal-Wallis test for testing of group differences. More recently, Karlsson, Zetterholm & Sullivan (2004) showed that the skew of the distribution across target plosive type may not be assumed to be identical and that the variance of VOT measurements varies with age of the participant.

The shape of the distribution of VOT for a specific target plosive voicing type and VOT category may therefore provide additional information about the nature of the productions from a developmental perspective. Furthermore, the distribution characteristics will be the determining factor when choosing an appropriate statistical testing for factor effects. For these reasons, a detailed investigation into the distribution shape within each sub-sample of the data was conducted.

4.4.2 Investigating the distribution of VOT measurements

As a complement to the measurements of means and medians presented in Table 4.2, Table 4.3 presents the standard deviation of each cell created by dividing the data into groups by the factors plosive voicing type, the age category of the child and VOT category category. For each cell, values describing the distribution shape are displayed within parenthesis; the left value indicates the skewness and the right value indicate
excess kurtosis. For skewness scores, negative values indicate a tail in the distribution towards the negative end of the scale and positive values indicate a tail towards the positive end. For excess kurtosis, positive scores indicate a more pointed distribution compared to the normal distribution and negative scores indicate a distribution flatter than the normal distribution.

Cells with small sample sizes are regarded as less informative of group developmental trends. The standard deviations in these cells are considered not to be sufficiently stable and therefore weaker indications of the true behaviour. Samples with less than ten data points are displayed in italics. Cells containing a single data point or less are marked with NA and do not form part of the analysis presented in this Thesis.

The results of a cell-by-cell application of a Shapiro-Wilk’s test of normality are shown in Table 4.3 in the form of a sequence of asterisks following the standard deviation scores. A single asterisk in a cell indicates a deviance from the normal distribution at the 0.05 level, two asterisks indicate a deviance at the 0.01 level and three asterisks indicate a deviance at the 0.001 level.

4.4.3 Interpretation of the progression in distribution shape

The distributional statistics displayed in Table 4.3 show a diverse pattern of development. In the long lag category, aspirated targets are initially produced with a near flat distribution centred close to the VOT boundary. Later in development, the distribution is centred around higher VOT values and the shape becomes more pointed (Figure 4.6).

In productions of unaspirated plosives with VOT in the long lag category, the distributional statistics indicate a positive skewness in all cells. However, no clear developmental trend is observed in skewness or kurtosis scores either for voiced target plosives or for voiceless aspirated plosives. There is only weak evidence of a decline in kurtosis and skewness with age. In addition, a gradual decrease in variability can be observed across age category. Thus, even though the number of productions in this VOT category did not always decrease, indicating a progression towards the adult target, the productions become increasingly centred around the distribution mean. Therefore, the distribution characteristics presented in Table 4.3 show signs of a gradual adjustment of production parameters on a microscopic scale.

For the short lag category, the presence of a developmental pattern is dependent on the aspiration category of the target. The distribution of VOT in unaspirated plosives produced in the default short lag category does not change over the age range under investigation. For unaspirated targets, no progression is observed in the standard deviation for any of the target plosive voicing types and the skewness measures remain. This is seen as an indication of the bias for an onset of voicing that is close in time to the release of the plosive proposed by Kewley-Port & Preston (1974).

For aspirated targets produced in the short lag category, the skewness scores are initially positive, but changes to a zero or slightly negative scores in the two final age categories. Thus, the data indicates a shift in the distribution towards more productions in the long lag category even though the production failed to cross the adult perceptual boundary. Furthermore, a shift is observed in kurtosis for voiced plosives from a flat distribution to one that is more pointed than the normal distribution. Thus, the production of aspirated targets progressed from a distribution that is similar to the
un aspirated targets to one that is more focused around the posi tive end of the VOT category, and is therefore more adult-like.

For plosives produced with prevoicing concerning distribution shape, there is no clear developmental pattern. However, excluding cells with less than ten data points, a tentative progression from larger to smaller standard deviations for voiced target plosives can be observed. Only one distribution, the period between 24–30 months, was found to be near zero and slightly positively skewed. The remaining distributions are negatively skewed, indicating a preference in production of prevoicing closer to the origin.

The standard deviations of productions towards a voiceless unaspirated plosive vary substantially with the largest value in the 30–36 month period. As for the voiced targets, the distributions for voiceless unaspirated plosives produced with prevoicing are negatively skewed, indicating a preference of productions of VOT close to zero.

The description of the distribution of VOT measurements has provided a detailed description of the progression towards a distinction between aspirated and unaspirated plosives. It has also provided information concerning prevoicing production; as children grow older, pre-voiced tokens are produced with decreasing prevoicing length.
Table 4.2: Sample size, mean and median values for the produced VOT, divided into cells by Age Category and target plosive Voicing Category. NA indicates that variance could not be computed since only a single production was made in this category. Italics indicate less than ten values in the cell.

<table>
<thead>
<tr>
<th>Age</th>
<th>Voiced</th>
<th>Voiceless unaspirated</th>
<th>Voiceless aspirated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevoiced</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 - 24</td>
<td>N=9, $\bar{x}$=49.8576, M=32.498</td>
<td>N=20, $\bar{x}$=59.9618, M=56.189</td>
<td>N=9, $\bar{x}$=50.6187, M=40.998</td>
</tr>
<tr>
<td>24 - 30</td>
<td>N=37, $\bar{x}$=71.7309, M=67.462</td>
<td>N=27, $\bar{x}$=37.4574, M=17.591</td>
<td>N=6, $\bar{x}$=29.8258, M=33.4035</td>
</tr>
<tr>
<td>30 - 36</td>
<td>N=40, $\bar{x}$=62.2177, M=54.4085</td>
<td>N=36, $\bar{x}$=58.4634, M=44.687</td>
<td>N=6, $\bar{x}$=63.4428, M=71.9105</td>
</tr>
<tr>
<td>36 - 42</td>
<td>N=12, $\bar{x}$=48.5497, M=33.87</td>
<td>N=21, $\bar{x}$=42.9555, M=27.422</td>
<td>N=2, $\bar{x}$=34.8415, M=34.8415</td>
</tr>
<tr>
<td>42 - 48</td>
<td>N=20, $\bar{x}$=63.6599, M=63.8695</td>
<td>N=6, $\bar{x}$=26.4182, M=23.44</td>
<td>N=1, NA</td>
</tr>
<tr>
<td>48 - 54</td>
<td>N=15, $\bar{x}$=37.3829, M=22.533</td>
<td>N=14, $\bar{x}$=28.4316, M=22.3765</td>
<td>N=1, NA</td>
</tr>
<tr>
<td>Short lag</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 - 24</td>
<td>N=100, $\bar{x}$=12.2865, M=11.023</td>
<td>N=116, $\bar{x}$=14.5005, M=11.622</td>
<td>N=87, $\bar{x}$=14.426, M=12.299</td>
</tr>
<tr>
<td>24 - 30</td>
<td>N=158, $\bar{x}$=15.0199, M=13.829</td>
<td>N=264, $\bar{x}$=17.0922, M=15.1125</td>
<td>N=99, $\bar{x}$=19.5664, M=18.878</td>
</tr>
<tr>
<td>30 - 36</td>
<td>N=176, $\bar{x}$=14.3435, M=12.24</td>
<td>N=244, $\bar{x}$=15.1075, M=12.9965</td>
<td>N=51, $\bar{x}$=20.2749, M=10.152</td>
</tr>
<tr>
<td>36 - 42</td>
<td>N=69, $\bar{x}$=14.5564, M=12.841</td>
<td>N=101, $\bar{x}$=17.2269, M=15.97</td>
<td>N=26, $\bar{x}$=21.7218, M=21.557</td>
</tr>
<tr>
<td>42 - 48</td>
<td>N=52, $\bar{x}$=12.7978, M=11.481</td>
<td>N=49, $\bar{x}$=17.5826, M=18.392</td>
<td>N=24, $\bar{x}$=26.6735, M=29.791</td>
</tr>
<tr>
<td>48 - 54</td>
<td>N=35, $\bar{x}$=11.8664, M=11.121</td>
<td>N=47, $\bar{x}$=12.99, M=10.739</td>
<td>N=22, $\bar{x}$=23.8772, M=23.1295</td>
</tr>
<tr>
<td>Long lag</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 - 24</td>
<td>N=9, $\bar{x}$=60.9963, M=59.445</td>
<td>N=15, $\bar{x}$=59.8115, M=54.102</td>
<td>N=24, $\bar{x}$=82.2264, M=66.1875</td>
</tr>
<tr>
<td>24 - 30</td>
<td>N=28, $\bar{x}$=73.0881, M=65.7995</td>
<td>N=66, $\bar{x}$=85.1987, M=74.6535</td>
<td>N=130, $\bar{x}$=93.3124, M=84.7905</td>
</tr>
<tr>
<td>30 - 36</td>
<td>N=17, $\bar{x}$=72.756, M=67.536</td>
<td>N=44, $\bar{x}$=75.3497, M=65.0255</td>
<td>N=183, $\bar{x}$=98.9422, M=97.117</td>
</tr>
<tr>
<td>36 - 42</td>
<td>N=10, $\bar{x}$=50.7167, M=48.6045</td>
<td>N=25, $\bar{x}$=80.112, M=75.89</td>
<td>N=79, $\bar{x}$=102.1638, M=94.249</td>
</tr>
<tr>
<td>42 - 48</td>
<td>N=5, $\bar{x}$=65.1756, M=60.507</td>
<td>N=63, $\bar{x}$=77.6797, M=71.505</td>
<td>N=74, $\bar{x}$=87.6022, M=78.732</td>
</tr>
<tr>
<td>48 - 54</td>
<td>N=3, $\bar{x}$=55.1293, M=46.733</td>
<td>N=8, $\bar{x}$=72.2946, M=53.954</td>
<td>N=40, $\bar{x}$=71.6314, M=66.5275</td>
</tr>
</tbody>
</table>
Table 4.3: Sample size and standard deviation for the produced VOT, divided into cells by Age Category and target plosive Voicing Category. NA indicates that variance could not be computed since only a single production was made in this category. Italics indicate less than ten values in the cell. Values within parenthesis indicate the skewness and kurtosis of the VOT distribution in that cell. Asterisks indicate significance levels resulting from a Shapiro-Wilk’s test of normality.

<table>
<thead>
<tr>
<th>Age</th>
<th>Voiced</th>
<th>Voiceless unaspirated</th>
<th>Voiceless aspirated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevoiced</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 - 24</td>
<td>N=9, s=46.616 (-0.425; -1.665)</td>
<td>N=20, s=40.742 (-0.425; -0.649)</td>
<td>N=9, s=48.416 (-0.48; -1.048)</td>
</tr>
<tr>
<td>24 - 30</td>
<td>N=37, s=50.572 (0.072; -1.424)**</td>
<td>N=27, s=43.406 (-1.089; -0.454)**</td>
<td>N=6, s=24.204 (0.117; -2.142)</td>
</tr>
<tr>
<td>30 - 36</td>
<td>N=40, s=44.444 (-0.207; -1.182)*</td>
<td>N=36, s=51.1 (-0.316; -1.524)**</td>
<td>N=6, s=32.122 (0.47; -1.501)</td>
</tr>
<tr>
<td>36 - 42</td>
<td>N=12, s=40.86 (-0.94; -0.342)</td>
<td>N=21, s=40.604 (-0.918; -0.408)**</td>
<td>N=2, NA</td>
</tr>
<tr>
<td>42 - 48</td>
<td>N=20, s=46.301 (-0.286; -1.188)</td>
<td>N=6, s=21.689 (-0.288; -1.704)</td>
<td>N=1, NA</td>
</tr>
<tr>
<td>48 - 54</td>
<td>N=15, s=38.016 (-0.772; -0.788)*</td>
<td>N=14, s=27.695 (-1.077; 0.04)*</td>
<td>N=1, NA</td>
</tr>
<tr>
<td>Short lag</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 - 24</td>
<td>N=100, s=7.572 (0.899; 0.439)***</td>
<td>N=116, s=9.704 (0.851; 0.294)**</td>
<td>N=87, s=8.552 (0.682; 0.406)**</td>
</tr>
<tr>
<td>24 - 30</td>
<td>N=158, s=8.347 (0.57; 0.312)***</td>
<td>N=264, s=9.852 (0.4; -0.887)**</td>
<td>N=99, s=11.145 (0.165; 1.227)**</td>
</tr>
<tr>
<td>30 - 36</td>
<td>N=176, s=9.451 (0.825; -0.08)***</td>
<td>N=244, s=9.018 (0.608; -0.265)**</td>
<td>N=51, s=9.441 (0.361; -0.763)</td>
</tr>
<tr>
<td>36 - 42</td>
<td>N=69, s=9.492 (0.814; -0.183)**</td>
<td>N=101, s=10.107 (0.471; -0.72)**</td>
<td>N=26, s=10.912 (0.021; -1.362)</td>
</tr>
<tr>
<td>42 - 48</td>
<td>N=52, s=7.825 (0.967; -0.106)**</td>
<td>N=49, s=9.569 (0.184; -0.902)</td>
<td>N=24, s=10.684 (-0.964; 0.072)*</td>
</tr>
<tr>
<td>48 - 54</td>
<td>N=35, s=7.733 (1.173; -0.283)**</td>
<td>N=47, s=9.463 (1.086; 0.317)**</td>
<td>N=22, s=10.051 (-0.051; -1.488)</td>
</tr>
<tr>
<td>Long lag</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 - 24</td>
<td>N=9, s=16.731 (0.074; -1.729)***</td>
<td>N=15, s=21.479 (1.636; 2.473)**</td>
<td>N=24, s=11.665 (0.961; -0.68)**</td>
</tr>
<tr>
<td>24 - 30</td>
<td>N=28, s=29.564 (0.757; -0.695)**</td>
<td>N=66, s=37.452 (1.014; 0.103)**</td>
<td>N=130, s=38.273 (0.716; -0.325)**</td>
</tr>
<tr>
<td>30 - 36</td>
<td>N=17, s=23.021 (0.293; -1.302)</td>
<td>N=44, s=35.04 (1.206; 0.918)**</td>
<td>N=183, s=36.099 (0.591; -0.214)**</td>
</tr>
<tr>
<td>36 - 42</td>
<td>N=10, s=10.579 (1.113; -0.329)</td>
<td>N=25, s=31.488 (0.318; 1.248)</td>
<td>N=79, s=38.047 (0.606; -0.305)**</td>
</tr>
<tr>
<td>42 - 48</td>
<td>N=5, s=17.908 (0.717; -1.286)</td>
<td>N=63, s=26.868 (0.876; 0.022)***</td>
<td>N=74, s=30.57 (0.667; -0.587)**</td>
</tr>
<tr>
<td>48 - 54</td>
<td>N=3, NA</td>
<td>N=8, s=30.375 (0.736; -1.417)*</td>
<td>N=40, s=26.246 (1.056; 0.403)**</td>
</tr>
</tbody>
</table>
4.5 Investigating group differences

The descriptive statistics presented in Section 4.4, in particular the statistics presented in Table 4.3, show that the data set cannot be adequately represented using a Gaussian distribution. This observation is in agreement with the findings of Koenig (2000), Koenig (2001) and Karlsson et al. (2004) that VOT measurement distributions tend to be skewed relative to the Gaussian distribution. However, the non-parametric test of statistical differences between groups used in these studies does not afford the ability to account for the effect of repeated measurements made on the same speaker at different ages. This effect must be accounted for in order to ascertain the effects of the independent variables. Consequently, the VOT data presented in this chapter was fitted using a Mixed Effects model using the nime statistical package (Pinheiro & Bates 2002, Pinheiro, Bates, DebRoy & Sarkar 2004). In this model, AGE and VOICING CATEGORY were treated as independent variables. A histogram of the residuals of this model is displayed in Figure 4.7. Placed on top of the histogram, a curve with the same mean and standard deviation indicates the shape of a normal distribution. The visual representation of the shape of the distribution of residuals and the calculated kurtosis (Kurtosis = 2.35) indicates a more pointed distribution than would be expected if the distribution was approximately normal. Thus, the model investigated is a bias estimate of the effect of the independent variables on the dependent variable VOT.

![Figure 4.7: Histogram of residuals for the mixed-effects model of VOT by AGE and target plosive VOICING CATEGORY.](image-url)
However, Figure 4.7 shows that the distribution of the residuals is more pointed than a normal distribution, does not include a heavy tail and is symmetric around the distribution mean. Thus, the direction of the model bias is towards producing test statistics associated with a higher probability for an effect than would be produced by an unbiased model. Hence, the true probability of group differences is smaller than or equal to the obtained probability score. This characteristic of the model increases the chance of making a Type II error and decreases the chance of a Type I error.

Bearing in mind the bias towards Type II error of analysis, the model from which the residuals of Figure 4.7 were obtained was analysed for group effects using an ANOVA. The results showed a significant main effect of **VOT Category** \(F(2,1154)=605.5, p<0.001\) and **Voicing Category** \(F(2,1154)=129.6, p<0.001\). However, **Age** was found not to be significant \(F(1,1154)=0.17, p=0.67, \text{n.s.}\). In addition, no significant difference was found for the spontaneity of the utterance \(F(4,1154)=0.87, p=0.48, \text{n.s.}\). Thus, the different elicitation strategies did not influence the resulting VOT produced.

In addition to the main effects, interaction effects of **Age**, **VOT Category** and **Voicing Category** were investigated using the same ANOVA procedure as for the main effects. The results showed a significant main effect to the interaction between **VOT Category** and **Voicing Category** \(F(1,1154)=7.8, p<0.01\). No significant effect was, however, found for the interaction between **Age** and **Voicing Category** \(F(2,1154)=2.52, p=0.11, \text{n.s.}\) or for the interaction between **Age** and **VOT Category** \(F(2,1154)=0.09, p=0.90, \text{n.s.}\). Thus, the shift with age from short lag to long lag VOT in productions of aspirated plosives in the UC³ Corpus was not sufficiently strong to provide a statistical effect.

### 4.6 Discussion

From research on voicing cues in adult speech, VOT measurement has been argued to provide an effective quantification of the perception of voicing (Lisker & Abramson 1964, Lisker & Abramson 1967, Kluender, Lotto & Jenison 1995, Lotto & Kluender 2002, Steinschneider, Volkov, Noh, Garell & Howard 1999) that may be used across languages in differing ways (Lisker & Abramson 1964, Abramson & Lisker 1973, Cho & Ladefoged 1999). Thus, children acquiring a language have to learn a language-specific use of the VOT cue as a part of the acquisition process.

Possibly in an effort to facilitate the acquisition of VOT as a voicing cue, VOT has been shown to be produced differently when directed towards children in their second year of life compared to when directed towards adults. Data from the two speakers investigated by Malsheen (1980) indicated an increased separation between voiced and voiceless unaspirated plosives at the time when the children are starting to produce their first words (1;3-1;4 in the study by Malsheen (1980)). Speech directed towards older children (2;5 and 5;2) was, however, shown by Malsheen (1980) to be produced with a perceptual separation more similar to that which may be observed in adult-directed speech.

In contrast, the output produced by the children has been shown to include a perceptually unclear distinction between the plosive voicing categories in English. A strong preference for short lag VOT, independent of the target plosive voicing and aspiration quality, has been observed in initial productions made by children (Catts & Kamhi 1984, Eguchi & Hirsh 1969, Kewley-Port & Preston 1974, Macken & Barton 1978, Zlatin &
Voice Onset Time

Koenigsknecht 1976). Data from longitudinal investigations of children’s VOT productions (Kewley-Port & Preston 1974, Zlatin & Koenigsknecht 1976) has provided evidence of a gradual progression from a unimodal VOT distribution where target plosives are manifested similarly, to a bimodal distribution where aspirated target plosives are increasingly produced with a long lag VOT.

The data presented in this chapter has provided evidence supporting previous claims of a short lag VOT predominance (Kewley-Port & Preston 1974), especially at the earliest ages. Frequency counts of productions within the prevoiced, short lag and long lag categories (Figures 4.4–4.5 and Table 4.1) indicate a gradual shift from the short lag VOT category towards a relative increase in long lag productions of aspirated target plosives. A similar trend is observed in the distribution shapes within each category; productions of aspirated plosives progress in a gradual manner towards the long lag range with age (Figure 4.6). Thus, the data agrees well with the data provided by investigations of children acquiring English in the way the produced VOT progresses towards increasingly adult-like levels in aspirated plosives.

In parallel to the development from a predominance of short lag VOT to a long lag VOT in aspirated plosives, the data presented in this chapter provides evidence of a simultaneous progression towards an increased use of prevoicing. As seen in Table 4.1, the relative frequency of prevoiced productions for voiced and voiceless unaspirated target plosives increases with age. Thus, the productions of unaspirated plosives are increasingly being separated from aspirated plosives by an increased use of prevoicing, complementing the move towards long lag VOT. Hence, the collected VOT data provides evidence of a careful re-organisation of their phonological system that gradually increases the perceptual distance between the target forms.

The development towards acoustically separated productions is viewed as a transition that is strongly governed by the process of development in the child’s underlying representation. Two factors provide evidence for this view. First, a progression similar to that which was observed for aspirated targets is initially also present in the productions made towards an voiceless aspirated target. According to Kewley-Port & Preston (1974), long lag VOT should be more difficult to produce than short lag VOT. It is, therefore, not plausible to propose that this progression is due a shift in articulatory strategy motivated by increased articulatory ease. Thus, the failure to produce long lag VOT only in productions where it agrees with adult target indicates a period in development where the difference between the two output forms is not manifested in the child’s underlying specification of the plosive.

Second, there is a progression in the manifestations of prevoicing in voiced target plosives from strong prevoicing towards less marked values (Tables 4.2–4.3). Similarly to the case of long lag VOT, prevoicing was argued by Kewley-Port & Preston (1974) to entail an increased articulatory effort compared to short lag VOT. This would make, in the absence of structural constraints, extreme manifestations of prevoicing a non-preferred output form in developing children. In this case, the structural constraint may come from the need to separate the output forms of voiced targets from that of aspirated targets. With development, children move away from the extreme manifestations of prevoicing towards an onset of voicing closer to the release of the onset, thus reducing the articulatory effort needed, while maintaining the achieved contrasts.

Based on the observations about VOT development away from proposed unmarked production in the short lag range, it is proposed that the child’s increase in the un-
derstanding of the contrasts available in the system to be acquired is a strong factor determining the output form produced. It is also proposed that children may show signs of erroneous applications of acoustic cues being explored, indicating a failure to apply discovered features across the board, solely in contexts where it agrees with the adult target. This chapter has provided strong evidence of exploration of the acoustic cue of VOT and periods of confusion.
Use of aspiration to signal voicelessness

Developmental studies of aspiration have to account for variation in the loudness of the tokens. This is achieved by using relative aspiration amplitude that is plosive aspiration amplitude relative to the amplitude after the onset of phonation. To calculate relative amplitude aspiration amplitude a point of aspiration onset, a point of aspiration offset and a time of voice onset are needed. The relative aspiration amplitude was calculated by normalizing the mean amplitude of aspiration by mean amplitude of an equal-length section of the following voiced section of the syllable. It was possible to extract relative aspiration amplitude from 1522 of the 4338 tokens in the UC3 Corpus. The number of tokens from which these values could be extracted was fewer than the number of voiceless tokens from which it was possible to extract VOT data.

Three factors contributed to this lower number. One, there were few attempts at the production of aspiration in the earlier ages. Two, as was noted by Karlsson, Sullivan, van Doorn & Czigler (2003), a child’s initial attempts at producing aspiration often have a low amplitude and that this may result in an increased co-occurrence of voicing and the aperiodic component. This co-occurrence can make accurate measurement of the aspiration amplitude and the amplitude after the onset of phonation infeasible. In such cases, no measurements were extracted. Three, as a result of the markup criteria used in the main part of the corpus; fricative portions of the signal were included in the segment labelled as aspiration as well as the aperiodic portion of the signal perceived by the transcriber as aspiration. For this part of the study, only the aperiodic portions where the time of onset could be determined from the acoustic waveform were included in the analysis. Measurements were not obtained from stimuli with a gradual rise in the amplitude of the fricative noise, since the mean amplitude obtained would be highly influenced by the arbitrary selection of boundaries.

The distribution of these productions across the Age, target Voicing Category and Place of Articulation is displayed in Table 5.2. The per cent frequency of occurrence of a measurable relative aspiration amplitude is displayed in table 5.3. Together, tables 5.2 and 5.3 show that the relative frequency of productions, where
amplitude measurements of fricatives were obtained, increases with age in the aspirated targets. This increase occurs in targets of all places of articulation.

The trend in increase is similar to the increase observed in VOT in figures 4.5 and 4.4. The correlation between the frequency of VOT measurements and relative amplitude of aspiration obtained in each cell created by the factors age and Place of Articulation was 0.88. A Pearson’s product-moment correlation applied to the frequency scores showed that this correlation was significant at the 0.05 level (t=4.9, p<0.001). Thus, the presence of a measurable amplitude was largely determined by the presence of a measurable release, providing an abrupt onset of frication noise, and a clear point of voicing onset.

For voiced and voiceless unaspirated target plosives, the increase observed for aspirated plosives does not occur in dental and labial targets. In contrast, velar targets showed a the per cent occurrence increase after the age of 36 months even in unaspirated targets that is not seen in the other places of articulation. However, the small sample sizes in these cells make a comparison across Place of Articulation in terms of prevalence of measurable aspiration difficult.

Based on the frequency of production of measurable aspiration, one may conclude that the productions are made by children in an increasingly differentiated manner depending on the aspiration quality of the target plosive. If the rise in measurable amplitude is indeed caused by a rise in aspirated sections produced rather than an decrease in loss of measurements for one of the other reasons discussed above, the data serves as a complement to the frequency data for VOT presented in Section 4.3, showing an increase number of in adult-like productions throughout development.

In order to establish whether the aspiration data obtained may be viewed as a signs of development in the voicing contrast, paralleling the VOT data presented in 4, the relative amplitude obtained for the aspirated portion was measured. The results from these measurements will be discussed in the following section.

5.1 Investigating the Measurable Relative Aspiration Amplitude

A change in the frequency distribution of measurable Relative Aspiration Amplitude does not, by itself, mean that Relative Aspiration Amplitude is being used to indicate voicelessness. It is necessary to consider the ratio values, their development and their distribution in order to confirm that the change found in the frequency distribution of measurable Relative Aspiration Amplitude is, indeed, part of the development of the voicing contrast.

The nature of the obtained measurements of Relative Aspiration Amplitude is investigated in terms of descriptive statistics in Table 5.4. The distribution within each data cell created by Voicing Category, Age Category and Place of Articulation was investigated in terms of mean and standard deviation of the cell and tested for deviance from the normal distribution using a Shapiro-Wilk’s test of normality. The results from the statistical testing of normality are indicated by asterisks. There is no great variation in standard deviation across age categories and that there distribution within a cell deviates from normality in only 10 out of the 54 cells.
Table 5.1: Frequency of productions of measurable Relative Aspiration Amplitude divided by Voicing Category and Place of Articulation.

<table>
<thead>
<tr>
<th>Place</th>
<th>Voiced</th>
<th>Voiceless unaspirated</th>
<th>Voiceless aspirated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Velar</td>
<td>70</td>
<td>106</td>
<td>134</td>
</tr>
<tr>
<td>Dental</td>
<td>64</td>
<td>136</td>
<td>189</td>
</tr>
<tr>
<td>Labial</td>
<td>83</td>
<td>253</td>
<td>273</td>
</tr>
</tbody>
</table>

Table 5.2: Number of productions of measurable relative amplitude divided by Voicing Category, Place of Articulation and Age Category.

<table>
<thead>
<tr>
<th>Age</th>
<th>Voiced</th>
<th>Voiceless unaspirated</th>
<th>Voiceless aspirated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Velar</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 - 24</td>
<td>8</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>24 - 30</td>
<td>18</td>
<td>35</td>
<td>44</td>
</tr>
<tr>
<td>30 - 36</td>
<td>19</td>
<td>20</td>
<td>35</td>
</tr>
<tr>
<td>36 - 42</td>
<td>15</td>
<td>20</td>
<td>17</td>
</tr>
<tr>
<td>42 - 48</td>
<td>9</td>
<td>15</td>
<td>17</td>
</tr>
<tr>
<td>48 - 54</td>
<td>1</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Dental</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 - 24</td>
<td>10</td>
<td>28</td>
<td>26</td>
</tr>
<tr>
<td>24 - 30</td>
<td>26</td>
<td>42</td>
<td>57</td>
</tr>
<tr>
<td>30 - 36</td>
<td>15</td>
<td>32</td>
<td>50</td>
</tr>
<tr>
<td>36 - 42</td>
<td>6</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>42 - 48</td>
<td>6</td>
<td>10</td>
<td>19</td>
</tr>
<tr>
<td>48 - 54</td>
<td>1</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>Labial</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 - 24</td>
<td>14</td>
<td>40</td>
<td>29</td>
</tr>
<tr>
<td>24 - 30</td>
<td>25</td>
<td>84</td>
<td>69</td>
</tr>
<tr>
<td>30 - 36</td>
<td>19</td>
<td>45</td>
<td>82</td>
</tr>
<tr>
<td>36 - 42</td>
<td>11</td>
<td>43</td>
<td>29</td>
</tr>
<tr>
<td>42 - 48</td>
<td>10</td>
<td>37</td>
<td>51</td>
</tr>
<tr>
<td>48 - 54</td>
<td>4</td>
<td>4</td>
<td>13</td>
</tr>
</tbody>
</table>
Table 5.3: Relative frequency of productions (in per cent) of measurable relative amplitude divided by Voicing Category, Place of Articulation and Age Category.

<table>
<thead>
<tr>
<th>Age</th>
<th>Voiced</th>
<th>Voiceless unaspirated</th>
<th>Voiceless aspirated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Velar</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 - 24</td>
<td>18</td>
<td>29</td>
<td>47</td>
</tr>
<tr>
<td>24 - 30</td>
<td>26</td>
<td>43</td>
<td>62</td>
</tr>
<tr>
<td>30 - 36</td>
<td>32</td>
<td>30</td>
<td>55</td>
</tr>
<tr>
<td>36 - 42</td>
<td>60</td>
<td>74</td>
<td>77</td>
</tr>
<tr>
<td>42 - 48</td>
<td>41</td>
<td>65</td>
<td>74</td>
</tr>
<tr>
<td>48 - 54</td>
<td>8</td>
<td>44</td>
<td>71</td>
</tr>
<tr>
<td>Dental</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 - 24</td>
<td>21</td>
<td>46</td>
<td>48</td>
</tr>
<tr>
<td>24 - 30</td>
<td>32</td>
<td>42</td>
<td>64</td>
</tr>
<tr>
<td>30 - 36</td>
<td>26</td>
<td>39</td>
<td>69</td>
</tr>
<tr>
<td>36 - 42</td>
<td>32</td>
<td>65</td>
<td>74</td>
</tr>
<tr>
<td>42 - 48</td>
<td>24</td>
<td>30</td>
<td>61</td>
</tr>
<tr>
<td>48 - 54</td>
<td>10</td>
<td>33</td>
<td>86</td>
</tr>
<tr>
<td>Labial</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 - 24</td>
<td>19</td>
<td>28</td>
<td>35</td>
</tr>
<tr>
<td>24 - 30</td>
<td>18</td>
<td>32</td>
<td>43</td>
</tr>
<tr>
<td>30 - 36</td>
<td>15</td>
<td>21</td>
<td>55</td>
</tr>
<tr>
<td>36 - 42</td>
<td>22</td>
<td>47</td>
<td>58</td>
</tr>
<tr>
<td>42 - 48</td>
<td>20</td>
<td>45</td>
<td>80</td>
</tr>
<tr>
<td>48 - 54</td>
<td>24</td>
<td>15</td>
<td>81</td>
</tr>
</tbody>
</table>

Figure 5.1: Histogram of residuals for the mixed-effects model of Relative Aspiration Amplitude by Age and target plosive Voicing Category. A normal distribution is shown on top of the histogram as a visual comparison.
Table 5.4: Frequency, mean and standard deviation of the Relative Aspiration Amplitude (in dB), divided into cells by Age Category and target plosive Voicing Category. NAs indicates that variance could not be computed since only a single production was made in this category. Numbers in italics indicate less than ten values in the cell. Values within parenthesis indicate the skew of the VOT distribution in that cell. Asterisks indicate significance levels resulting from a Shapiro-Wilk's test of normality.

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>Voice</th>
<th>Voiceless unaspirated</th>
<th>Voiceless aspirated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Velar</td>
<td>18 - 24</td>
<td>N=8, x̄= -8.554, s=5.307</td>
<td>N=12, x̄= -11.126, s=3.55</td>
<td>N=16, x̄= -9.985, s=7.091</td>
</tr>
<tr>
<td></td>
<td>24 - 30</td>
<td>N=18, x̄= -5.227, s=8.748</td>
<td>N=35, x̄= -9.262, s=7.465</td>
<td>N=44, x̄= -8.631, s=8.034</td>
</tr>
<tr>
<td></td>
<td>30 - 36</td>
<td>N=19, x̄= -5.132, s=8.717</td>
<td>N=20, x̄= -6.627, s=9.215</td>
<td>N=35, x̄= -10.653, s=10.299*</td>
</tr>
<tr>
<td></td>
<td>36 - 42</td>
<td>N=15, x̄= -3.415, s=7.797</td>
<td>N=20, x̄= -8.657, s=7.102</td>
<td>N=17, x̄= -11.095, s=8.225</td>
</tr>
<tr>
<td></td>
<td>42 - 48</td>
<td>N=9, x̄= -6.978, s=7.431</td>
<td>N=15, x̄= -12.44, s=6.686</td>
<td>N=17, x̄= -13.41, s=4.387</td>
</tr>
<tr>
<td></td>
<td>48 - 54</td>
<td>N=1, NA</td>
<td>N=4, x̄= -11.551, s=7.086</td>
<td>N=5, x̄= -13.077, s=5.493**</td>
</tr>
<tr>
<td>Dental</td>
<td>18 - 24</td>
<td>N=10, x̄= -12.683, s=7.085</td>
<td>N=28, x̄= -10.344, s=7.36</td>
<td>N=26, x̄= -13.436, s=6.36</td>
</tr>
<tr>
<td></td>
<td>24 - 30</td>
<td>N=26, x̄= -10.041, s=6.907</td>
<td>N=42, x̄= -11.323, s=8.04</td>
<td>N=57, x̄= -11.642, s=7.853**</td>
</tr>
<tr>
<td></td>
<td>36 - 42</td>
<td>N=6, x̄= -8.438, s=4.503</td>
<td>N=20, x̄= -10.298, s=5.465</td>
<td>N=25, x̄= -12.639, s=7.695</td>
</tr>
<tr>
<td></td>
<td>42 - 48</td>
<td>N=6, x̄= -8.634, s=2.867</td>
<td>N=10, x̄= -12.43, s=4.737</td>
<td>N=19, x̄= -10.489, s=5.645**</td>
</tr>
<tr>
<td></td>
<td>48 - 54</td>
<td>N=1, NA</td>
<td>N=4, x̄= -14.646, s=8.168</td>
<td>N=12, x̄= -9.29, s=6.571</td>
</tr>
<tr>
<td>Labial</td>
<td>18 - 24</td>
<td>N=14, x̄= -12.347, s=9.091</td>
<td>N=40, x̄= -7.776, s=10.159</td>
<td>N=29, x̄= -10.807, s=9.86</td>
</tr>
<tr>
<td></td>
<td>24 - 30</td>
<td>N=25, x̄= -6.387, s=8.046</td>
<td>N=84, x̄= -9.261, s=9.942</td>
<td>N=60, x̄= -11.802, s=10.755**</td>
</tr>
<tr>
<td></td>
<td>30 - 36</td>
<td>N=19, x̄= -12.64, s=8.28</td>
<td>N=45, x̄= -12.138, s=8.843</td>
<td>N=82, x̄= -15.875, s=7.4**</td>
</tr>
<tr>
<td></td>
<td>36 - 42</td>
<td>N=11, x̄= -6.989, s=8.454</td>
<td>N=43, x̄= -9.464, s=10.038</td>
<td>N=29, x̄= -13.193, s=11.208*</td>
</tr>
<tr>
<td></td>
<td>42 - 48</td>
<td>N=10, x̄= -5.36, s=5.477</td>
<td>N=37, x̄= -11.674, s=7.445*</td>
<td>N=51, x̄= -15.399, s=7.931*</td>
</tr>
<tr>
<td></td>
<td>48 - 54</td>
<td>N=4, x̄= -5.997, s=6.867</td>
<td>N=4, x̄= -5.054, s=9.906</td>
<td>N=13, x̄= -8.19, s=8.25</td>
</tr>
</tbody>
</table>
5.2 Investigation of group differences

The data set of collected measurements of relative aspiration amplitude were analysed in terms of main and interaction effects of age of the subject and aspiration category of the target plosive.

A repeated measures analysis of variance was calculated on a fixed effects linear model with speaker as the grouping factor. The model was fitted and analysed using the \texttt{nlme} library (Pinheiro & Bates 2002, Pinheiro et al. 2004) for the R statistical software package. The residuals of the fitted model, displayed in Figure 5.1, were shown not to deviate significantly from a normal distribution using a Shapiro-Wilk test of normality (W=0.999, p=0.82, n.s.). Thus, this mixed effects model was used for the subsequent statistical testing of effects in the data set.

The data set was analysed for main and interaction effects using an ANOVA on the fitted mixed effects model. The analysis included the effect of target Voicing Category and Age on the use of relative aspiration amplitude in the produced plosives.

The results from the showed a main effect of speaker age ($F_{(1,1282)}=436.9, p<0.001$). Table 5.4 indicates that aspirated plosives are produced with a stronger aspirated portion in the initial ages compared to when the children are older. A similar, but less prominent, trend may be observed in voiceless unaspirated plosives, but not in voiced plosives.

In addition to the age effect, the voicing type of the target was shown to provide a significant main effect ($F_{(2,1282)}=21.5, p<0.001$). From the display of descriptive statistics presented in Table 5.4 it observed that aspirated plosives are produced with smaller mean relative aspiration amplitude than both unaspirated categories. Similar results were found in an application of the same model on the absolute aspiration amplitude data ($F_{(2,1452)}=5.7, p<0.01$).

The interaction between age and voicing type was shown to be not significant ($F_{(2,1282)}=2.3, p=0.096, n.s.$). Thus, difference in the use of high relative aspiration amplitude for aspirated versus unaspirated plosives observed in Table 5.4 are considered not significant.

5.3 Discussion

The frequency counts presented for relative aspiration amplitude in the UC³ Corpus indicate that the occurrence of measurable aspiration amplitude increases with age. This progression is observed across places of articulation for aspirated plosive targets, but for unaspirated plosive target. This selective progression is in agreement with the pattern found in the UC³ Corpus' VOT data discussed in Chapter 4. A strong correlation between measurable Relative Aspiration Amplitude and long lag VOT is evident in the UC³ Corpus data and suggests that the child’s ability to produce measurable Relative Aspiration Amplitude develops in parallel with the child’s increase in use of long lag VOT.

The statistical tests applied provide evidence of a general change in Relative Aspiration Amplitude due to age. The descriptive statistics show that Relative Aspiration Amplitude is greatest in the youngest age group and lowest in the oldest age group. The productions made in the early sessions are extreme versions of the production of aspiration that are later refined as the child’s speech develops. This
Aspiration Amplitude

development in Relative Aspiration Amplitude can be explained both in terms of articulatory motor-control development and by speech production theories such as the H & H theory proposed by Lindblom (1990).

The articulatory explanation is that the production of a strong fricative segment is easier than a weak fricative segment. In the child’s case, if the lungs are fully filled at the start of production and the child is not able to regulate the airflow, the resultant output will be dissimilar to the adult model. It is possible that the child can perceive the difference between the produced aspiration and the perceived aspiration, yet is unable to reduce it, until its articulatory ability increases. Thus, as articulatory motor-control improves, the productions should become more adult-like and a linear trend of development towards lower aspiration amplitudes, possible accompanied with a decrease in variability, would be expected.

Although the expected pattern of decrease in Relative Aspiration Amplitude is found for velar and labial voiceless plosives (See Table 5.4, it is not found for the dental plosives of the same voicing category. Thus, a decrease in Relative Aspiration Amplitude does not occur across the board for all aspirated plosives. Additionally, there are no clear signs of a decreasing variability. In the samples of comparable size, standard deviation does not decrease with age within a category (Table 5.4). In some instances, the variance is largest in the middle or later age groups. Thus, the UC³ Corpus does not support an explanation based on articulatory development (alone).

The alternative explanation based on a developmental application of the H & H theory (Lindblom 1990), viewing the produced speech as initially an extreme (hyper) version of the adult model. Thus, it is expected that feature Relative Aspiration Amplitude will initially be larger than the adult model in order to ensure a maximum likelihood that the production will be perceived by the adult listener as the intended aspirated segment. Then, through an interaction loop between the speaker and the listener, the child would become aware of the exaggerated and non-economic nature of their previous productions and try to find a way of optimally reducing effort and of maintaining sufficient contrast. In this way, the output form is altered through a process of re-evaluation of the cues for aspiration that is driven by observations of the adult model productions.

The developmental path in such a model is highly dependent upon the adult model productions presented to the child, and the model presented as part of a recording session dialogue could have impacted upon the recorded productions. In the collection of the UC³ Corpus, it proved impossible to always have the same care-giver present for a child’s set of recordings. This increased the variability of adult productions made to the child in the recording setting. As variability in input is viewed as leading to variability in production, this model does not predict a steady decrease in variability of the aspirations produced. This model therefore provides a better explanation to the data on Relative Aspiration Amplitude extracted from the UC³ Corpus than an explanation based on articulatory development alone. It permits the re-evaluation and re-organisation of the child’s understanding of the use of Relative Aspiration Amplitude in signalling of the voicing contrast as an integral part of child’s acquisition of this contrast.

The analysis of measurable Relative Aspiration Amplitude showed that unaspirated plosives targets were produced with stronger aspiration amplitudes than aspirated plosive targets. This was true for both the absolute and relative aspiration amplitudes.
This does not lend itself easily to interpretation. The results are opposite of what would be expected if aspiration amplitude was actively being used by the child as a tool to signal voicelessness. One possible explanation is that children explore the acoustic construction of unaspirated plosives in detail before they explore aspirated plosives to the same degree. That the voiceless targets has stronger aspiration amplitudes than the voiced targets could then be explained as a result of the familiarity with the distinctive perceptual cues that results in a more confident and forceful production by the child.

5.4 Conclusions drawn from the use of relative amplitude

The data presented in this chapter indicate a development in the aspiration amplitude as the child grows older. The initial productions are made with a more extensive force than the productions made in later sessions.

The data, however, does not provide evidence for the active use of relative aspiration amplitude as a solitary cue for the signalling the voicing contrast. Statistical testing provided a significant main effect of target voicing type. It was found, however, upon close inspection that it was portions of aspiration wrongfully applied to voiced plosives that received additional amplitude compared to the aspirated targets that were the cause of this main effect. Further, no significant interaction effect between age and target voicing type, which would be expected if the main effect found for voicing target type was part of a strategy for the voicing contrast, was found.

The progression found due to age is, it can, therefore, be concluded, a result of the production of aspiration in an increasingly economical manner and part of the general process of development production skills, and not a sign of the reorganisation of the child’s understanding of the voicing contrast. The analysis of relative aspiration amplitude presented in this chapter has shown that Relative Aspiration Amplitude is not used by the children in the UC3 Corpus as the primary cue of voicing. It is possible, however, that Relative Aspiration Amplitude may serve as a complement to the VOT cue discussed in Chapter 4. The interaction of these cues in the productions made by the children in the UC3 Corpus is discussed in Chapter 6.
Chapter 6

Linguistic uses of voicing cues complementing Voice Onset Time

In the previous chapters of Part II of the Thesis, potential main voicing and aspiration cues have been investigated in terms of developmental trends and contrastive use. Evidence has been provided for a distinctive usage of VOT, but not for relative aspiration amplitude, as a main cue to the voicing and aspiration distinction.

In this chapter, the use of auxiliary cues in combination with VOT will be explored. First, the use of relative aspiration in ambiguous cases in terms of VOT is examined. Second, the main and auxiliary use of $f_0$ as a voicing cue is investigated for developmental trends and distinctive use across target and produced plosives. Finally, a synthesis of the results presented in Chapters 4–6 is provided in terms of distinctive use and developmental trends.

6.1 Interaction of aspiration force and VOT

In Chapter 5, it was concluded that aspiration amplitude was not used by the investigated children as a main acoustic cue for signalling an aspirated contrast. Thus, the possibility suggested by Winitz et al. (1975) that aspiration amplitude is a dominant cue to voicelessness was not found to be applicable for the developmental data in the UC3 Corpus.

However, the effect of aspiration amplitude proposed by Repp (1979) was not covered in Chapter 5. Repp (1979) proposed that the relative amplitude of aspiration may push the VOT boundary towards zero, causing more VOT values to be perceived as voiceless or voiceless aspirated. In this section, the interaction of aspiration amplitude and VOT will be examined.

Placing the interaction between VOT and Relative Aspiration Amplitude in a developmental perspective, an hypothesis may be formed that strong aspiration amplitude may be used by children in ambiguous VOT productions. When failing to cross the adult VOT boundary, the child may be experimenting with aspiration amplitude instead. If relative aspiration amplitude is used as an auxiliary cue to voicing, it can
be expected that relative aspiration amplitude of a production is dependent of the distance between the boundary between short lag and long lag VOT and the produced VOT.

Testing this hypothesis, a mixed-effects model was applied to data set, investigating the use of aspiration amplitude as a function of the absolute distance of the VOT value of the production from the VOT boundary, Age of the participant and Voicing Category of the target plosive. The Participant factor was included as a fixed effect in the model in order to account for the fact that the data had been obtained using repeated measures of the same participants.

The ANOVA performed using this model showed a significant main effect of the absolute value of the distance in time between the produced VOT and the boundary between the short lag and long lag VOT Categories ($F_{(1,1459)}=16.9, p<0.001$). A Kendall’s rank correlation showed a negative correlation between VOT distance and relative aspiration amplitude ($\tau=-0.32, p<0.001$). Thus, the use of forceful aspiration increased in ambiguous cases in terms of VOT. No significant effect of Age was found ($F_{(1,1459)}=0.09, p=0.76, n.s$). Thus, there is no significant developmental effect in the use of aspiration amplitude in pushing the VOT boundary forward.

The data set did not allow for the integration of age, VOT distance and Age and Voicing Category to be tested. Thus, it was not possible to test whether relative amplitude was used in attempted productions of plosives of a specific Voicing Category.

Repp (1979) postulated that an increase of 1 dB in relative aspiration amplitude would decrease the VOT boundary by 0.43 ms, increasing the likelihood of the plosive being perceived as a long lag plosive. This suggestion is investigated in Figure 6.1. This balloon plot shows the increase in perceived long lag productions caused by a shifted VOT boundary. The data set is divided by Age (rows) and Voicing Category (columns). The average amplitude ratio of the productions made in the same session was used as a baseline from which amplitude increase was calculated.

In agreement with the observations made in Chapter 5, but contrary to what an hypothesis of linguistic use would propose, Figure 6.1 indicates that linguistic use of aspiration amplitude occurs most frequently in the voiceless unaspirated target plosives produced in the early ages. A trend of decrease in Relative Aspiration Amplitude with age is observed.

The data from voiceless aspirated target plosives does not show a developmental trend in aspiration use. According to Figure 6.1, relative aspiration amplitude is less frequently used as an auxiliary cue to voicelessness in productions of voiceless aspirated targets than the voiceless unaspirated targets discussed above. Further, the age trend is less pronounced; the number of productions pushed into the long lag range VOT by the perceptual effect of aspiration is approximately the same in the 18–24 and 42–48 age categories.

The differentiated use of relative aspiration amplitude across target plosive Voicing Category was, in Chapter 5, suggested to be caused by a higher familiarity with the target plosive or the target word, increasing the assertiveness in the production. In order to further evaluate this hypothesis, the amplitude of the entire utterance produced by the child was extracted for testing. A mixed-effects model investigating the effects of Age and Voicing Category was applied to the extracted Utterance Amplitude. Participant was included as a fixed effect due to the longitudinal design of the study.
The results showed a significant effect of both Age (\(F_{(1,1607)}=9.0, p<0.01\)) and Voicing Category (\(F_{(2,1607)}=8.3, p<0.001\)). The data set did not allow interaction effects to be investigated.

In order to further investigate the effect of overall amplitude of the utterance, Utterance Amplitude was modelled by mean Relative Aspiration Amplitude and VOT using a mixed effect model with Participant as a fixed effect. The results showed that the Relative Aspiration Amplitude did not cause a significant effect on the amplitude used in the utterance (\(F_{(1,1607)}=3.2, p=0.07, n.s.\)). The significance effect remained when using Session Date as the fixed effect (\(F_{(1,1607)}=1.9, p=0.16, n.s.\)).

Thus, the use of increased Relative Aspiration Amplitude at the VOT boundary may not be explained by a more forceful production of the entire utterance.

A significant effect on mean Utterance Amplitude was, however, found for VOT (\(F_{(1,1607)}=21.3, p<0.001\)). This effect remained when incorporating the Session Date as a fixed effect (\(F_{(1,1607)}=19.8, p<0.001\)). The correlation between produced VOT and produced amplitude of the utterance was shown to be negative (\(\tau=-0.4\)), indicating that utterances containing a plosive with a small VOT were produced with a significantly higher amplitude.

The higher amplitude in productions with smaller VOT may be taken as an indication of an increased confidence in the production. However, if this confidence is caused by an increased familiarity with the target it would be expected that spontaneous utterances would show a significantly higher amplitude. The effect of spontaneity of the production on utterance amplitude was therefore investigated using a mixed-effects model. The results showed a significant effect of spontaneity: utterances produced spontaneously had a higher mean amplitude than the rest of the productions (\(F_{(4,1831)}=2.77, p<0.05\)). Thus, the higher amplitude observed in productions with a short VOT lag is argued to be consistent with an increased confidence in the attempted target.

6.2 The use of \(f_0\) onset frequency for signalling of a voicing distinction

In addition to the time of voicing onset, quantified in the VOT measure, \(f_0\) has also been proposed to influence the perception of plosives by its frequency. Early accounts argued that the entire frequency contour of \(f_0\) influenced voicing perception (Lisker 1986, Haggard, Ambler & Callow 1970). However more recent accounts have argued for an effect of onset frequency of \(f_0\), not the \(f_0\) contour, (Fujimura 1971, Haggard et al. 1981, Stevens 2000, McCrea & Morris 2005, Ohde 1985) or an effect of both cues (Whalen et al. 1993, Whalen et al. 1990) on the perception of voicing quality. The perceptual influence of \(f_0\) reported by these investigations is that voiced plosives have been observed to have a low \(f_0\) onset frequency and aspirated plosives have been observed to have higher \(f_0\) onset frequency.

An interplay between VOT and \(f_0\) onset frequency signalling a VOT distinction has also been proposed (Fujimura 1971, Holt et al. 2001). As to the nature of this interplay, Holt et al. (2001) stated that
Figure 6.1: Balloon plot of the increase in productions speculated to be perceived as long lag VOT due to an incorporation of the effect of relative aspiration amplitude. Data is divided by Age Category and Voicing Category.

“...that low f₀ contributes to the presence of low frequency energy during and near the consonant, thus enhancing the perception of voicing.”
(Holt et al. 2001, p. 765)

Thus, the presence of a low frequency f₀ would, according to Holt et al. (2001), give stronger prominence to the onset of voicing due to spectral separation in the acoustic energy. Through a perception experiment on Japanese Quail birds using artificial stimuli where VOT and f₀ onset frequency either co-varied or were composed randomly, Holt et al. (2001) were able to provide results indicating that “[...f₀ does not exert an obligatory influence on categorization of consonants as [VOICE][...]]” (Holt et al. 2001, p.764). They therefore argued for a learned covariance between VOT and f₀.

The question then arises when this co-variation is learnt by children. Ohde (1985) investigated the development of VOT and f₀ as a cue to the voicing and aspiration distinction for children aged 8–9 years. The results presented by Ohde (1985) showed that by this point in development, the VOT distinction had reached a state of high agreement with VOT in adult speech, both in mean VOT and in variability. The f₀ cue, however, showed significant differences in mean frequencies and variability across children and adults. Thus, the use development of f₀ as a voicing cue was shown to be delayed compared to VOT.
Table 6.1: Descriptive statistics for $f_0$ at voicing onset across Voicing Category of the plosive target and Age Category. The statistics presented are cell size, mean, median and standard deviation.

<table>
<thead>
<tr>
<th>Voiced Age Category</th>
<th>Voiceless aspirated Age Category</th>
<th>Voiceless unaspirated Age Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 - 24</td>
<td>N=170, $\bar{x}$=202.7, $M$=203, $s$=62.1</td>
<td>N=167, $\bar{x}$=206.7, $M$=207, $s$=57.8</td>
</tr>
<tr>
<td>24 - 30</td>
<td>N=316, $\bar{x}$=199.3, $M$=199, $s$=64.6</td>
<td>N=343, $\bar{x}$=204, $M$=204, $s$=62.4</td>
</tr>
<tr>
<td>30 - 36</td>
<td>N=210, $\bar{x}$=203.3, $M$=203, $s$=61.6</td>
<td>N=235, $\bar{x}$=193.3, $M$=193, $s$=69.2</td>
</tr>
<tr>
<td>36 - 42</td>
<td>N=70, $\bar{x}$=191.3, $M$=191, $s$=70.1</td>
<td>N=98, $\bar{x}$=201.5, $M$=201, $s$=64.4</td>
</tr>
<tr>
<td>42 - 48</td>
<td>N=87, $\bar{x}$=160.7, $M$=161, $s$=47.4</td>
<td>N=118, $\bar{x}$=178.2, $M$=178, $s$=40.8</td>
</tr>
<tr>
<td>48 - 54</td>
<td>N=22, $\bar{x}$=160.7, $M$=161, $s$=39</td>
<td>N=25, $\bar{x}$=199.7, $M$=200, $s$=55.4</td>
</tr>
</tbody>
</table>

Table 6.2: Descriptive statistics for $f_0$ at voicing onset across produced segment’s Voicing Category and Age Category. The statistics presented are cell size, mean, median and standard deviation.

<table>
<thead>
<tr>
<th>Voiced Unaspirated Age Category</th>
<th>Voiceless Aspirated Age Category</th>
<th>Voiceless Unaspirated Age Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 - 24</td>
<td>N=340, $\bar{x}$=200.5, $M$=200, $s$=60</td>
<td>N=47, $\bar{x}$=202, $M$=202, $s$=53</td>
</tr>
<tr>
<td>24 - 30</td>
<td>N=572, $\bar{x}$=195.3, $M$=195, $s$=67</td>
<td>N=161, $\bar{x}$=190, $M$=190, $s$=65.3</td>
</tr>
<tr>
<td>30 - 36</td>
<td>N=329, $\bar{x}$=200.7, $M$=201, $s$=66</td>
<td>N=120, $\bar{x}$=188.2, $M$=188, $s$=66.2</td>
</tr>
<tr>
<td>36 - 42</td>
<td>N=110, $\bar{x}$=193.7, $M$=194, $s$=61.2</td>
<td>N=67, $\bar{x}$=197.3, $M$=197, $s$=55.3</td>
</tr>
<tr>
<td>42 - 48</td>
<td>N=92, $\bar{x}$=158.8, $M$=159, $s$=50</td>
<td>N=104, $\bar{x}$=180.5, $M$=180, $s$=47.4</td>
</tr>
<tr>
<td>48 - 54</td>
<td>N=20, $\bar{x}$=162.2, $M$=162, $s$=41.9</td>
<td>N=16, $\bar{x}$=196.7, $M$=197, $s$=64.5</td>
</tr>
</tbody>
</table>
Table 6.3: Descriptive statistics for $f_0$ at voicing onset across produced segment’s VOT Category and Age Category. The statistics presented are cell size, mean, median and standard deviation.

<table>
<thead>
<tr>
<th>VOT Category</th>
<th>Age Category</th>
<th>Long lag</th>
<th>Prevoiced</th>
<th>Short lag</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 - 24</td>
<td>N=67, $\bar{x}$=193.3, M=193, s=66.1</td>
<td>N=22, $\bar{x}$=185.7, M=186, s=81.7</td>
<td>N=201, $\bar{x}$=204.7, M=205, s=59.6</td>
<td></td>
</tr>
<tr>
<td>24 - 30</td>
<td>N=257, $\bar{x}$=196.7, M=197, s=66.6</td>
<td>N=85, $\bar{x}$=188, M=188, s=67.1</td>
<td>N=335, $\bar{x}$=209, M=209, s=61.2</td>
<td></td>
</tr>
<tr>
<td>30 - 36</td>
<td>N=229, $\bar{x}$=188, M=188, s=69.3</td>
<td>N=57, $\bar{x}$=199, M=199, s=63.4</td>
<td>N=212, $\bar{x}$=205.5, M=206, s=64.7</td>
<td></td>
</tr>
<tr>
<td>36 - 42</td>
<td>N=93, $\bar{x}$=207.7, M=208, s=42.4</td>
<td>N=17, $\bar{x}$=203.3, M=203, s=73.4</td>
<td>N=126, $\bar{x}$=190.7, M=191, s=56.2</td>
<td></td>
</tr>
<tr>
<td>42 - 48</td>
<td>N=122, $\bar{x}$=180.7, M=181, s=48.9</td>
<td>N=27, $\bar{x}$=169.7, M=170, s=47.2</td>
<td>N=80, $\bar{x}$=169, M=170, s=50</td>
<td></td>
</tr>
<tr>
<td>48 - 54</td>
<td>N=23, $\bar{x}$=188, M=188, s=42.4</td>
<td>N=8, $\bar{x}$=171.5, M=172, s=34.1</td>
<td>N=32, $\bar{x}$=177.3, M=177, s=47.9</td>
<td></td>
</tr>
</tbody>
</table>
The results of Ohde (1985) suggest that younger children should not show signs of linguistic use of $f_0$ as a voicing cue. In order to investigate this proposal, the distribution characteristics of $f_0$ at the different Age Category investigated are presented in Table 6.1 divided by Voicing Category of the target plosive, Table 6.2 divided by Voicing Category of the produced plosive. The division of $f_0$ impact by perceived Voicing Category and Table 6.3 divided by produced VOT Category. These descriptive statistics show no difference across Voicing Category before the age of 42 months. After the age of 42 months, mean $f_0$ onset frequency is reduced in both voiced and voiceless unaspirated plosives, but not in aspirated plosives. The pattern of development is similar in Table 6.1–6.3.

In order to test for statistical significance of the differences observed above as well as any use of $f_0$ in ambiguous cases similar to what was observed for relative aspiration amplitude, an ANOVA was applied to a mixed-effects model with Participant as a fixed effect. The dependent variable $f_0$ onset frequency was modelled by VOT, distance from the VOT boundary, Voicing Category of the target, Voicing Category of the produced plosive and Age. The results showed no significant main effects of VOT ($F_{(1,1105)}=0.1, p=0.75, n.s.$), distance from the VOT boundary ($F_{(1,1105)}=0.45, p=0.5, n.s.$), Voicing Category of the target ($F_{(1,1105)}=0.19, p=0.82, n.s.$), Voicing Category of the produced plosive ($F_{(1,1105)}=0.001, p=0.99, n.s.$) and Age ($F_{(1,1105)}=0.63, p=0.42, n.s.$) on the $f_0$ onset frequency used.

Thus, the statistical testing failed to show a significant difference in the use of $f_0$ as a voicing cue due to the nature of the target, the nature of other properties of the production or due to development. It is therefore concluded, in accordance with the proposal made by (Ohde 1985), that $f_0$ is not used by children to produce a contrast between aspirated and unaspirated plosives.

6.3 Discussion

In the previous chapters of this part of the Thesis, acoustic correlates, proposed in the literature to be main cues of the voicing or aspiration distinction were analysed in terms of developmental trends and distribution across Voicing Category of the target. Failure to show a main effect of Relative Aspiration Amplitude in Chapter 5 raises the question as to whether this cue functions as an auxiliary cue to aspiration, helping productions to be perceived as more adult-like.

This chapter investigated the possible auxiliary usage of aspiration amplitude in the signalling of aspirated rather than unaspirated plosives. The results showed that Relative Aspiration Amplitude was, indeed, used to a significantly larger extent when the produced VOT was close to the boundary between short lag and long lag VOT. Thus, it is proposed that Relative Aspiration Amplitude is used by the children as an alternative way of causing the plosive to be perceived as voiceless aspirated.

No statistically significant age effect was found in the use of relative aspiration amplitude. However, using Repp’s (1979) postulated relationship between VOT and relative amplitude of aspiration, an estimation of the developmental trend in relative amplitude usage was obtained. This estimation indicated that a change in production classification due to the impact of aspiration amplitude was largest in productions made before the age of 30 months. Thus, it is proposed that Relative Aspiration Amplitude may be overly used by the children in the early productions, and that
the subsequent productions are moderated by further input from the adult’s model productions towards an increased reliance on the VOT cue and a decrease distinctive usage of the auxiliary aspiration amplitude cue.

It is proposed that this is a learned rather than an articulatory effect due to the shift from one production pattern that may be argued to be costly in articulatory energy, i.e. high aspiration amplitude, towards another production pattern that has been argued to be complex coordination of articulators, i.e. long lag VOT (Kewley-Port & Preston 1974). If this was an articulatory effect, it would be expected that mean Relative Aspiration Amplitude would decrease with age across the board, not just in ambiguous productions in terms of VOT. No clear evidence of such a decreasing trend was, however, found (See Chapter 5).

Statistical testing of the interaction between VOT distance from boundary and Voicing Category was not possible. However, the estimation of the interaction effect by Repp’s (1979) data provided an approximation of the distribution of VOT distance and Voicing Category interaction (Figure 6.1). In this display of the interaction, it was observed that attempted productions of voiceless unaspirated target plosives were more likely to be pushed into the perceptual space of aspirated plosives than both voiced unaspirated and voiceless aspirated plosives. This result is confusing, since it would create decrease in adult-like productions, without any apparent articulatory reason.

One explanation explored was that the target form of voiceless unaspirated plosive would be more accurately specified in the child’s underlying representation, which would cause the output form to be more familiar and receive more emphasis in the production. One factor that might cause such an effect is the agreement of the default VOT value in children’s early productions of plosives (Kewley-Port & Preston 1974, Zlatin 1974, Zlatin & Koenigsknecht 1976) and the target form.

The proposed effect of a more confident production was investigated by investigating the mean amplitude of the utterance. The significant effects found included VOT and spontaneity of the production. Utterances with a short VOT were produced with a higher mean amplitude than productions with long lag VOT. Further, spontaneous productions were produced with a higher mean amplitude. The results are therefore viewed as consistent with the hypothesis of a higher familiarity with the target word causing the observed distribution of Relative Aspiration Amplitude across target Voicing Categories.

In addition to the use of Relative Aspiration Amplitude, this chapter also investigated the use of \( f_0 \) as a main and auxiliary acoustic cue to aspiration. It has been suggested in the literature (Lisker 1986, Haggard et al. 1970, Fujimura 1971, Haggard et al. 1981, Stevens 2000, McCrea & Morris 2005, Olde 1985, Whalen et al. 1993, Whalen et al. 1990) that the acoustic cue of \( f_0 \) onset frequency may either vary with perceived voicing or covary with VOT as an auxiliary voicing cue. The statistical testing failed to show a main effect of any of the investigated variables on \( f_0 \) onset frequency. An analysis of distribution of \( f_0 \) across age and voicing category of the target plosive as well of the produced plosive showed a decrease in mean and median for unaspirated plosives in the last year investigated. For aspirated plosives, no clear signs of a decrease in \( f_0 \) onset frequency with age was observed.

If \( f_0 \) onset frequency was actively being used by the children as an acoustic cue to voicing, an \( f_0 \) decrease would be expected to occur in the division by voicing category
of the target plosive, but not in the division by perceived voicing quality of produced plosive. Thus, the similarity in the distributions of \( f_0 \) in the three data division schemas indicate that the lowering of \( f_0 \) observed in the last investigated year is viewed as evidence for an increased overall similarity between the plosives produced and the target plosive. Thus, the \( f_0 \) development observed in the last year investigated is consistent with a co-variation between \( f_0 \) and VOT, and does therefore not constitute evidence for a developmental effect.

It is therefore concluded, in agreement with the predictions made from Ohde’s (1985) report, children are not using \( f_0 \) as a primary or an auxiliary cue to the voicing distinction between 18 and 54 months. Thus, the combined results from this chapter and Chapters 4 and 5 is that the primary cue to voicing explored by the children at this stage of development is the timing of plosive release relative to the onset of phonation. Aspiration amplitude may serve as an auxiliary, but not main, cue to the distinction, and \( f_0 \) is not used by these children as a method to produce a contrast in voicing.
Linguistic uses of voicing cues
Part III

Structural complexity
Chapter 7

Acquisition of structurally complex syllable onsets

An integral part of producing a consonant cluster is the ability to produce two consonants without a vowel in between them. This chapter investigates the ability of children to produce complex syllable onsets. In Section 7.1 the development of complexity in the target words is investigated using frequency counts of structurally complex onset production. In Section 7.2, the duration of syllable onset is investigated in terms of development in median values and variability across age, and in Section 7.3 the observations made in Sections 7.1 and 7.2 are discussed.

7.1 Structural complexity

The concept of structural complexity is a theoretically difficult issue. This section provides an overview of the judgements made in the classifications of onsets and presents the progression in production of structurally complex syllable onsets by the children in the UC3 Corpus.

In the simple case, the syllable may be seen as structurally complex if it contains two or more consonants. Thus, [sp] is considered a structurally complex syllable onset and [s] is considered to be a simple onset. This definition, however, is troublesome when the child’s production is perceived as, for instance, an affricate. Although affricates are perceived as different compared to a plosive+fricative sequence, it may not be assumed that the child has acquired an understanding of this distinction at the time of production.

As the focus of attention, in this Thesis, is the development in the production of consonants with differing characteristics, a complex syllable onset is considered to have been observed when the syllable onset is perceived by the researcher as being composed of two, or more, speech sounds produced at differing places of articulation or with differing MANNER OF ARTICULATION. Thus, in this Thesis all sequences of plosive+fricative are considered complex onsets.
The following section presents the production of complex syllable onsets by comparing the productions made of target words with an initial consonant cluster with target forms in the same target form quadruple (See Table 3.1).

### 7.1.1 Frequency of occurrence

It was observed in Chapter 4 that an increase in production ability may be generalised by the child to target words in which the acquired feature does not agree with the adult model. In this Thesis, therefore, the acquisition of a complex syllable onset is judged to have *commenced* when clusters are produced in cluster target words, and production is defined as having *stabilised* when target words with complex syllable onsets are produced with a complex syllable onset, and target words with a simple onset are produced with a structurally simple onset.

Figures 7.1–7.4 presents the percentage of productions of complex syllable onsets for each of the quadruples containing the target words /Spek/, /skal/, /stc/ and /spcr/. It can be observed in Figures 7.1–7.4 that target words with a complex syllable are reduced to singleton onsets in the initial ages. In these ages, the percentage of production of complex onsets are comparably independent of the attempted target word.

In the older age groups, the relative frequency of productions of complex onsets increased; a similar increase may also be observed for the target words with a simple syllable onset (see /skal/ and /sal/ panels of Figure 7.2). Thus, at this stage of development, the group of children investigated may be seen as having achieved a basic understanding of the production of complex syllable onsets, but not as having achieved stability in the application the production pattern.

Figures 7.1–7.4 include all productions of complex syllable onsets, including productions with syllable onsets that exceed the complexity of the two-consonant target clusters. An analysis of the frequency of production of overly complex syllable clusters is provided in the Section 7.1.2.

### 7.1.2 False positives

Even though the target words elicited included at most two consonants in sequence, the consonant clusters produced by the investigated children are a at times produced as a sequence of more than two consonants. The distribution of these overshoot productions across target word and age is presented in Figures 7.5 and 7.6.

The production of target overshoot in syllable onsets complexity shows signs of being age dependent. Figure 7.5 shows that the frequency of overshoot productions dropped at 31 months. An analysis of the distributions of these productions across **Participant** and **Age** revealed a strong participant effect in the distribution of the overshoot production across **Age**. Most of the productions made with overly complex syllable onsets before the age of 31 months were made by speaker M11; only seven out of the observed 86 productions made before the age of 31 months were not made by this speaker. The seven remaining productions were made by speakers F2 (3 productions) and M13 (4 productions). None of these speakers produced any overly complex syllable onsets after the age of 30 months. The productions observed after 30 months (see Figure 7.5) were produced by the speakers M4 and M5: M4 produced three overly complex syllable onsets at the 31 and 36 months and M5 produced the remaining
Figure 7.1: Bar chart showing relative frequency of complex onsets against Age (in months) for target words /sak/, /pak/, /bak/ and /spak/.

Figure 7.2: Bar chart showing relative frequency of complex onsets against Age (in months) for target words /sal/, /kal/, /gal/ and /skal/.
Figure 7.3: Bar chart showing relative frequency of complex onsets against Age (in months) for target words /so/, /to/, /do/ and /sto/.

Figure 7.4: Bar chart showing relative frequency of complex onsets against Age (in months) for target words /sor/, /por/, /bor/ and /spor/.
overly complex syllable onsets observed in the UC³ Corpus after 30 months. The data, thus, indicate a speaker dependence in the frequency of production of overly complex syllable onsets, but no general trend in terms of developmental direction.

The overly complex productions were also investigated in terms of their distribution across target words. Figure 7.6 presents the frequency of production of syllable onsets with more than two consonants. Two observations can be made concerning the distribution of overly complex productions across target words attempted First, it can be observed from Figure 7.6 that target words with a structurally complex syllable onset are more likely to be produced with an overtly complex syllable onset. Second, a bias can be observed that is due to the target word quadruple. Productions of the skal quadruple (/skəl/, /kəl/, /gəl/ and /səl/) are produced more frequently with more than three consonants in the syllable onset than the other quadruples. A \( \chi^2 \) test applied to the frequency counts shown in Figure 7.6, grouped by quadruple, showed that the differences observed are significant (\( \chi^2=47, p<0.001 \)). Since the quadruple is the only one containing a velar plosive, the significant results of the applied \( \chi^2 \) test may indicate a place dependency in the confusion of complex output forms. A larger data set is, however, needed to verify the accuracy of this hypothesis.

This section has investigated the distribution of syllable onsets that were produced in a more complex manner than the target words with a complex syllable onset demand. These productions represent productions that fail to meet the adult norm. The next section focuses on the progression of both less complex and more complex than demanded productions towards two-consonant clusters in solely the appropriate target words.

![Figure 7.5: Frequency of too long sequences in the onset by Age of the Participant.](image)
7.1.3 Production stability in the manifestation of complex syllable onsets

The previous section provided evidence for consonant clusters being reduced to singleton consonants in early stages of development. Further, evidence has been provided for the existence of a stage in development where children can produce syllable onsets of greater complexity than the adult target complex syllable onset.

In this Thesis two criteria are used to test for stable progression in the acquisition of the production of complex syllable onsets. First, the relative production of two-consonant clusters should show an increase towards 100% of the total number of productions. Second, the relative production frequency of a two-consonant syllable onset should be reduced in non-cluster target words. Together these criteria assure a detection of a stable progression in the acquisition of the production of complex syllable onsets as it maximizes the contrast in syllable Onset Complexity between the target words of differing type within the quadruples used in this Thesis.

Figures 7.7–7.10 present bar plot visualizations of the progression in of the production of complex syllable onsets for each of the investigated quadruples. Each bar shows the average rate of production of two-consonant onset clusters for the ages following the position of the bar. For example, the bar shown at 40 months represents the average percentage of productions with a cluster between 40 and 51 months. The panels show the progression for each of the target words within each quadruple. Within each panel, the 0%, 10%, 50% and 80% achievement levels are shown using horizontal lines. Figures 7.7–7.10 indicate, that within each quadruple the clustered target word is separable from the rest of the target words; target words with consonant clusters as their onset are initially produced at a level around 50% whereas target words with
simple onsets at around the 10% mark. Hence, the clustered onset target words are
produced with two-consonant onset clusters more frequently and from an earlier stage
during acquisition of appropriate production of complex syllable onsets than singleton
onset target words.

![Figure 7.7: Progressive relative frequency of complex onsets with two components for

target words /sak/, /pak/, /bak/ and /spak/

The establishment of an age of acquisition is dependent on the production rate
chosen as the milestone. Table 7.1 presents the ages at which the productions of two-
consonant clusters reach the 70%, 80%, 90% and 100% mark of all productions of
target words with a complex syllable onset. Depending upon the chosen milestone,
stable adult-like production of structural complexity is achieved in target words with a
CC onset between 34 and 50 months. However, if the 100% mark is chosen only three
of the four CC-onsets are achieved; the CC-onset [st] in the target word /stʃ/ does not
reach the 100% mark.

In order to investigate the cause of this failure of achievement of a two-consonant
cluster onset for the target word /stʃ/, the frequency of occurrence of a reduction to
a simple onset was investigated for all complex target words. The results showed that
the target word /stʃ/ was reduced to a single plosive more frequently than the other
cionant cluster target words. This difference was shown to be significant by a χ²
 test applied to the frequency of reductions for the CC target words (χ²(2)=6.9, p<0.05).
Thus, the failure to achieve a stable two-consonant cluster in this target word is viewed
as being caused by it being more likely to be subjected to the cluster reduction process,
i.e. reduction of the complex syllable onset [st] to [t] or [ʃ].

The increase in complex syllable onsets of target words, where these are in agree-
ment with the adult production is, however, not sufficient to determine the age of
acquisition of a production ability for consonant cluster complexity. An additional
Figure 7.8: Progressive relative frequency of complex onsets with two components for target words /sal/, /kal/, /gal/ and /skal/.

Figure 7.9: Progressive relative frequency of complex onsets with two components for target words /so/, /to/, /do/ and /sto/.
Table 7.1: Age of stable production of two-element onsets in sC target words. Each row presents the Age (in months) at which the rate of following productions increased above the level indicated in the left-most column for the CC target words (columns).

<table>
<thead>
<tr>
<th>% CC</th>
<th>skal</th>
<th>spak</th>
<th>spor</th>
<th>sto</th>
</tr>
</thead>
<tbody>
<tr>
<td>50%</td>
<td>20</td>
<td>21</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>60%</td>
<td>28</td>
<td>30</td>
<td>37</td>
<td>27</td>
</tr>
<tr>
<td>70%</td>
<td>35</td>
<td>36</td>
<td>43</td>
<td>34</td>
</tr>
<tr>
<td>80%</td>
<td>41</td>
<td>43</td>
<td>44</td>
<td>42</td>
</tr>
<tr>
<td>90%</td>
<td>44</td>
<td>48</td>
<td>48</td>
<td>45</td>
</tr>
<tr>
<td>100%</td>
<td>45</td>
<td>48</td>
<td>49</td>
<td>NA</td>
</tr>
</tbody>
</table>

Table 7.2: Age of reduction of CC onset productions in target words with C syllable onset structure. The Age (in months) at which the relative frequency of production of a complex onset was reduced to the 10, 5 and 0 percent of the total number of productions is indicated for each target words.

<table>
<thead>
<tr>
<th>% CC</th>
<th>sal</th>
<th>gal</th>
<th>kal</th>
<th>sak</th>
<th>bak</th>
<th>pak</th>
<th>sor</th>
<th>bor</th>
<th>por</th>
<th>so</th>
<th>do</th>
<th>to</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>44</td>
<td>48</td>
<td>42</td>
<td>48</td>
<td>35</td>
<td>48</td>
<td>48</td>
<td>36</td>
<td>45</td>
<td>49</td>
<td>31</td>
<td>47</td>
</tr>
<tr>
<td>5%</td>
<td>44</td>
<td>48</td>
<td>36</td>
<td>48</td>
<td>23</td>
<td>48</td>
<td>48</td>
<td>23</td>
<td>45</td>
<td>49</td>
<td>21</td>
<td>47</td>
</tr>
<tr>
<td>10%</td>
<td>43</td>
<td>48</td>
<td>31</td>
<td>48</td>
<td>18</td>
<td>48</td>
<td>48</td>
<td>18</td>
<td>18</td>
<td>49</td>
<td>18</td>
<td>18</td>
</tr>
</tbody>
</table>
condition is that a decrease in CC productions in target words with simple onsets should be observed. Table 7.2 presents the ages at which a rate of CC onsets that were below or equal to 10%, 5% and 0% for the target words /buk/, /bor/, /do/, /kal/, /puk/, /por/, /suk/, /sal/, /so/, /sor/ and /ko/ was achieved. From Table 7.2 it can be seen that the reduction in consonant cluster production in singleton target words is more scattered than the age at which the production of appropriate complex onsets is achieved (See Table 7.1). Production of complex syllable onsets for singleton syllable onsets in the target words reaches the 10% level between 18 and 49 months, the 5% level between 21 and 49 months and 0% between 31 and 49 months. The group data, thus, reveal a variability that covers a large part of the age range investigated irrespective of whether the 10%, 5% or 0% milestone was chosen.

The acquisition of a pattern of producing simple onsets for simple onset targets happens before, or at the same time as, the acquisition of a stable two-consonant cluster for three out of the four target words with complex onsets. For /skal/, the singleton control word /gal/ is not sufficiently reduced in cluster production frequency at the time when clusters are being produced with an adult-like complexity. A stable manifestation of the complex syllable onset for /skal/ will, therefore, be considered to have occurred once production of a singleton onset for /gal/ has stabilised.

In this section, the development of a contrast in production between simple and complex syllable onsets was investigated using analysis of production frequencies. The achievement of a two-consonant cluster in productions of target words where this agrees with the adult norm was contrasted with the productions of target words with a simple onset. In this paradigm of cluster complexity investigation, development is said to occur when the number of productions of complex syllable onsets increases in cluster target words and decreases in target words with a structurally simple syllable onset.

The quantification of complexity in terms of number of consonants in the syllable onset is, however, dependent on the definition of structural complexity and on the perceptual closeness of the produced speech segment to an adult category. For instance, an intended [st] transformed by metathesis may be acoustically similar both to, for instance, an affricate and possibly also an attempted production of aspirated plosive, depending on the acoustic manifestation of the aperiodic portion. Thus, categorisation of the produced segment may potentially different depending upon the judge and on individual variation in the boundaries for the dominating acoustic cues.

In order to reduce listener dependency, a complementary investigation of the development of syllable onset duration was conducted. The results from this complementary investigation are presented in the following section before the results of the two investigations are synthesised.

7.2 Duration of the produced syllable onset

Syllable onset duration may serve as an alternative quantification of the complexity of the syllable onset for the age range investigated in this Thesis. In a pilot investigation of a subset of the data investigated in this Thesis (productions made by eight of the female participants), Karlsson (2004) reported a significant difference in syllable onset duration due to complexity of the target.

Previous research has provided evidence for a developmental effect on the duration of the syllable onsets produced in attempted productions of a consonant cluster. In an
early series of investigations, Hawkins (1979) investigated the duration of consonants in a cluster environment produced by children 4–7 years old and adults. The data presented by Hawkins (1979) showed that the durations of consonant clusters produced by children are smaller than in consonant clusters productions by adults. For the subset of the UC³ Corpus, presented in Karlsson (2004) this result was confirmed; a significant main effect of AGE was found.

This section examines the contrastive use of duration in productions towards complex syllable onsets compared to productions towards simple syllable onsets. Variability in the productions of the target works in terms of developmental effects is also examined.

7.2.1 Statistical testing of Age and produced onset complexity effects on onset duration

There is a strong, but not perfect, connection between produced complexity and duration in the investigated syllable onsets. Figure 7.11 displays box-and-whiskers plots of the durations of syllable onsets containing one (bottom panel), two (middle panel) and three (top panel) consonants.

The analysis approach of Karlsson (2004) was repeated for the entire data set investigated in this Thesis. In addition, to the Karlsson (2004) analysis, the interaction effect of AGE and produced ONSET COMPLEXITY was included into the analysis of the entire data set. The effects were investigated using a mixed-effects model, with repeated measurements of the same speakers, using the \texttt{nlinem} package (Pinheiro & Bates 2002, Pinheiro et al. 2004). As for the subsample investigated by Karlsson (2004), a main effect was found for AGE ($F_{(1,4707)}=48.8, p<0.001$) and produced ONSET COMPLEXITY ($F_{(1,4707)}=95.9, p<0.001$). However, the interaction of AGE and ONSET COMPLEXITY, which was not investigated by Karlsson (2004), was shown to not provide a significant effect ($F_{(1,4707)}=2.27, p<0.131, n.s.$).

Based on this evidence, it may be concluded that the duration of a produced syllable onset is significantly affected by the structural complexity of the syllable onset. Syllable onsets produced as a single onset differ significantly in duration from complex syllable onsets. Thus, durational data may serve as a detailed acoustic quantification of structural complexity, with the added ability to investigate the development of production variance.

The following section provides a contrastive investigation into the duration of syllable onsets across target words within quadruples of potential homonyms. An analysis of the progression is presented in terms of median and variance values.

7.2.2 Comparison of syllable onset durations

In the contrastive investigation of production development presented in this Thesis, differences in the duration of simple and complex syllable onsets are viewed as potential signs of a developing manifestation of complexity. The onset durations produced in attempted productions of the target words are, therefore, grouped according to the contrasting quadruples.

Figures 7.12–7.15 present the onset durations obtained in the form of a Box-and-whiskers plot for each month, in cells by target word. The durations presented within
The contrasted quadruples reveal similar patterns. The target word with a complex syllable onset is at the early ages produced with a short duration and small variability. As development progresses, the median duration of the cluster, indicated by the dot inside the box, increases towards a global maximum during the second part of the child’s second year of life. After this maximum value is reached, a decrease in median duration of the cluster is observed. In addition to the decrease in median duration, a decrease in variability is observed in the production of the target words /skal/ (Figure 7.12 and /spor/ (Figure 7.14). The cluster target words display the greatest variation among the target words of each quadruple.

The duration increase in the onset is not as pronounced in target words with a simple onset. Target words with an initial [s] target onset show the second greatest increase in median duration of the words in the quadruple. Median duration decreases after achieving this local maximum in the target word /so/ (Figure 7.15). The variability in the productions made at any particular age is, however, much less than in the cluster target productions.

For the two remaining target word categories, words with an initial aspirated plosive and words with a voiced unaspirated plosive, development is observed in the duration of the onset in two different ways. For aspirated plosives, there is a tendency of an increase...
Figure 7.12: Box-and-whisker plots of the duration of the syllable onsets produced for the target words /sal/, /kal/, /gal/ and /skal/.

Figure 7.13: Box-and-whisker plots of the duration of the syllable onsets produced for the target words /sak/, /pak/, /bak/ and /spak/. 
in median duration with age. The degree of increase is target word dependent. A large increase in duration of the onset is observed for the target word /to/ (Figure 7.15); smaller increases are observed for the target words /knl/ and /pak/ (Figures 7.12 and 7.13).

For target words with an initial voiced plosive, an increase in duration is not observed with age. For the target word /ga1/, instances of very long durations are observed between 22–30 months that result in large variability. One cause of this increase in duration may be observed in Figures 7.6 and 7.8; the target word /ga1/ is frequently produced as a long cluster. Further, /ga1/ stabilised as a non-cluster 6 months after /knl/ (Table 7.2). During this period, the production of /ga1/ and /skal/ are not significantly different ($F_{(1,261)}=2.64, p=0.105, n.s.$). A decrease in both median duration and variability are observed after this period of development. A significant difference is, however, present after this period ($F_{(1,287)}=43.9, p<0.001$). Thus, both measured structural complexity and duration measurements indicate a period of confusion in the productions of /ga1/ and /skal/ between 22 and 30 months.

### 7.3 Discussion

This chapter has investigated the acquisition of structurally complex clusters and considered when a child has the ability to produce complex syllable onsets in a stable manner. Stability in complex onset production is quantified in this chapter using two different strategies. First, a complex cluster is considered to have been achieved when productions of two-consonant target clusters are produced with exactly two consonants and the simple target onsets within the quadruple are produced as a single consonant. Second, a complex syllable onset is considered to have been acquired stability when a systematic increase in syllable onset duration is observed.

Complex syllable onsets do not only occur in productions of target words with a complex syllable onset. Figures 7.1 – 7.4 show that complex onsets occur frequently in the control words with simple syllable onsets, and that a complex onset is most likely in simple [s] onset words. The greater likelihood that [s] is produced as a complex syllable onset is further manifested in the later stabilisation of these target syllable onsets as simple syllable onset (See: Figures 7.7 – 7.10). Thus, it can be concluded that the children investigated in this Thesis display greater confusion between [s] target words and the [sC] target word than between [sC] and the control words containing a single plosive. Further, it can also be concluded that the substantial increase in syllable onset duration observed for [s] target words is not solely caused by the longer duration of the target phoneme [s].

The increased confusion of the [s] target word for cluster target words may have perceptual reasons. It is possible that the initial [s] of the adult model is more easily confused with the cluster due to the identical initial element in the adult model productions. If the child has not acquired a robust segmentation of the consonant cluster, confusion may arise as to which adult model form is a cluster, creating a common pattern of consonant cluster production for both the [s] and [sC] forms.

The age of acquisition of a stable production of complexity is dependent on the milestone selected. The results presented in Table 7.1 indicate that a stable production of complex syllable onset is achieved at the 80% level after the age of 41 months and the 100% level after the age of 44 months. A stable production contrast is, for some
Figure 7.14: Box-and-whisker plots of the duration of the syllable onsets produced for the target words /sor/, /por/, /bor/ and /spor/.

Figure 7.15: Box-and-whisker plots of the duration of the syllable onsets produced for the target words /so/, /to/, /do/ and /sto/. 
target words, be achieved after that, due to confusion in the production of simple-onset control words (Table 7.2). For instance, productions of /gal/ reach the 0% level of complex syllable onset productions after the 100% production level is reached for the cluster word in the same quadruple (/skal/). Thus, although the increasing complexity of productions of complex targets are the dominant factor in the determination of achievement rates in the structural complexity contrast, failure to reduce productions of structurally simple targets may influence the achievement age.
Part IV

Acquisition of manner and place of articulation
Acquisition of manner of articulation

As was shown in the previous chapter, word-initial clusters may be reduced to single consonants, in the early stages of the acquisition of consonant clusters. Children’s productions of singletons, in turn, are subject to processes that cause the manner of the production to differ from the adult model. Grunwell (1987) described two of these processes as stopping (e.g., /s/⇒[t]) and lenition (e.g., /f/⇒[z]). Thus, the Manner of Articulation of a plosive may change to a weaker form such as a fricative, and a fricative may be produced as a plosive. With development, the child gradually acquires a more adult-like production of the elements in word-initial clusters. The manners of articulation used by the child in the production of word-initial clusters are therefore investigated in this Chapter in terms of developmental trends and contrastive use.

First, the output forms produced as singleton consonants are contrasted within a target word quadruple (Section 8.1). That is, the Manner of Articulation of productions with adult-like syllable onset complexity in CV(C) target words will be contrasted with reduced forms of the attempted sCV(C) word productions.

Second, the Manner of Articulation of the first and second consonant in complex productions is investigated. Productions with adult-like complexity are contrasted with complex productions where the adult model consists of a single consonant (Section 8.2). In addition, the Manner of Articulation for the entire sequence is investigated.

Third, the stabilisation of an adult-like Manner of Articulation is examined contrastively within a target word quadruple (Section 8.3). The findings regarding the acquisition of a contrast in Manner of Articulation are discussed and interpreted in Section 8.4.

8.1 Manner of articulation in productions of simple syllable onsets

It was shown in Chapter 7 that attempted productions of both simple and complex target are produced as single consonants. In this section, the target words produced by the children are contrasted in terms of used Manner of Articulation.
Figures 8.1–8.4 present, for each quaduple, a dotplot of the frequency of occurrence (on the abscissa) against participant Age (on the ordinate) in panels by target word. The used MANNERS OF ARTICULATION are indicated by the plotting character shown in the figure legend. Tables 10.1–10.4 present, for each quaduple, the distribution of productions in percent across MANNER OF ARTICULATION for each target word.

Figure 8.1: Relative frequency of occurrence for each MANNER OF ARTICULATION in simple productions of the /sal/, /kal/, /gal/ and /skal/ target words.

Figure 8.1 and Table 10.1 present the data for the /skal/ quadruple. For the target word /sal/ fricative and plosive productions coexist across most of the age range investigated. Table 8.1 shows that 44% of the productions were fricatives and 53% were plosives. An increase in fricative productions and a decrease in plosive productions is observed with age as the number of adult-like productions in terms of MANNER OF ARTICULATION for this target word increased. For the remaining three target words within the quadruple, plosive productions dominate. Table 8.1 shows that between 98.1% and 98.6% of the syllable onsets for the /skal/, /kal/ and /gal/ targets were produced as plosives and that fricatives were produced most frequently for the target word /kal/ with the singleton aspirated plosive syllable onset [kʰ].
Table 8.1: Distribution of productions across MANNER OF ARTICULATION for each target word within the quadruple /sal/, /kal/, /gal/ and /skal/. The total number of productions were 750. Values are in percent.

<table>
<thead>
<tr>
<th></th>
<th>Approximant</th>
<th>Fricative</th>
<th>Lateral</th>
<th>Nasal</th>
<th>Plosive</th>
</tr>
</thead>
<tbody>
<tr>
<td>gal</td>
<td>0.5</td>
<td>0.0</td>
<td>0.5</td>
<td>0.5</td>
<td>98.6</td>
</tr>
<tr>
<td>kal</td>
<td>0.0</td>
<td>2.9</td>
<td>0.5</td>
<td>1.0</td>
<td>95.6</td>
</tr>
<tr>
<td>sal</td>
<td>0.6</td>
<td>44.0</td>
<td>2.4</td>
<td>0.0</td>
<td>53.0</td>
</tr>
<tr>
<td>skal</td>
<td>0.0</td>
<td>0.0</td>
<td>0.6</td>
<td>1.2</td>
<td>98.1</td>
</tr>
</tbody>
</table>

Figure 8.2: Relative frequency of occurrence for each MANNER OF ARTICULATION in simple productions of the /sak/, /pak/, /bak/ and /spak/ target words.
Table 8.2: Distribution of productions across MANNER OF ARTICULATION for each
target word within the quadruple /bak/, /pak/, /sak/ and /spak/. The total number
of productions were 854. Values are in percent.

<table>
<thead>
<tr>
<th>Approximant</th>
<th>Fricative</th>
<th>Lateral</th>
<th>Nasal</th>
<th>Plosive</th>
</tr>
</thead>
<tbody>
<tr>
<td>bak</td>
<td>0.4</td>
<td>3.9</td>
<td>0.0</td>
<td>0.4</td>
</tr>
<tr>
<td>pak</td>
<td>0.7</td>
<td>3.6</td>
<td>0.0</td>
<td>0.4</td>
</tr>
<tr>
<td>sak</td>
<td>1.1</td>
<td>50.6</td>
<td>0.6</td>
<td>0.0</td>
</tr>
<tr>
<td>spak</td>
<td>0.6</td>
<td>4.7</td>
<td>0.6</td>
<td>0.6</td>
</tr>
</tbody>
</table>

For the spak quadruple (Figure 8.2 and Table 8.2) a pattern of production similar
to the skal quadruple is observed. The syllable onset of the target word /skal/ is
produced as a plosive in 44.7% of the productions and as a fricative in 50.6% of the
productions. Fricative productions for the target words /spak/, /pak/ and /bak/ oc-
cur more frequently than for the corresponding target words in the /skal/ quadruple.
Between 3.6% and 4.7% of the productions of these target words were produced with a
fricative, with the highest relative frequency being observed for /spak/. However, this
larger relative frequency of production should be considered a potential artifact of the
smaller number of productions of /spak/ with a single-consonant onset (i.e. an effect
of the truncation imposed in this analysis). The number of fricative productions was
eight for /skal/, nine for /bak/ and ten for /pak/. The relative frequency for alter-
native manners of articulations was below 1% of the total number of productions and
represents one or two productions. The fricative productions are, therefore, regarded
as too few to be considered part of a production pattern for the target words.

Table 8.3: Distribution of productions across MANNER OF ARTICULATION for each
target word within the quadruple /sor/, /por/, /bor/ and /spor/. The total number
of productions were 1036. Values are in percent.

<table>
<thead>
<tr>
<th>Approximant</th>
<th>Fricative</th>
<th>Lateral</th>
<th>Plosive</th>
</tr>
</thead>
<tbody>
<tr>
<td>bor</td>
<td>0.8</td>
<td>2.9</td>
<td>0.0</td>
</tr>
<tr>
<td>por</td>
<td>1.2</td>
<td>7.4</td>
<td>0.0</td>
</tr>
<tr>
<td>sor</td>
<td>3.4</td>
<td>47.3</td>
<td>0.5</td>
</tr>
<tr>
<td>spor</td>
<td>0.4</td>
<td>5.6</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Parts of the pattern of production observed in the /skal/ and /spak/ quadruples
are also observed for the /sor/, (Figure 8.3 and Table 8.3) and the /sco/ quadruple
(Figure 8.4 and Table 8.4). Approximately half of the consonants are produced as
plosives, and the other half are mainly produced as fricatives. The distribution of
these manners of articulation reveals a dominance of plosive productions in the early
ages with fricative production increasing with age. In both quadruples, the relative
frequency of productions of fricative simple onsets in non-adult like positions is largest
in the aspirated target onsets (/[pʰ]/ and /[tʰ]/. The target words with consonant cluster
onsets were produced with singleton fricative syllable onsets less often than the target
words with single aspirated plosive target onsets. Voiced consonant targets was only
replaced by a fricative in the target word /bor/ (Table 8.3). The distribution of fricative productions in non adult-like onset positions in the /spɔr/ and /stɔ/ quadruples agrees well with the trend observed in the /spək/ and /skəl/ quadruples. The aspirated plosive target onsets are the ones most frequently substituted by a fricative.

The data presented in this section has provided evidence for a strong presence of plosive productions in single-consonant output onsets. A similar production trend was also observed both in target onsets that include a plosive and those that include a fricative as one of the element of a cluster. Over time, the rate of fricative production increased in target words with a single fricative or aspirated plosive in the adult syllable onset. Fricative productions in attempted productions of s+plosive target clusters were less frequent. The observed low relative frequency of fricative productions in s+plosive target clusters could, however, be an artifact of the truncation performed when observing only reduced forms of the consonant cluster. Therefore, an analysis of the Manner of Articulation in two-consonant productions by the children will be provided in the following section.
Figure 8.4: Relative frequency of occurrence for each MANNER OF ARTICULATION in simple productions of the /so/, /to/, /do/ and /sto/ target words.

<table>
<thead>
<tr>
<th></th>
<th>Fricative</th>
<th>Nasal</th>
<th>Plosive</th>
</tr>
</thead>
<tbody>
<tr>
<td>do</td>
<td>0.0</td>
<td>0.4</td>
<td>99.6</td>
</tr>
<tr>
<td>so</td>
<td>44.8</td>
<td>0.0</td>
<td>55.2</td>
</tr>
<tr>
<td>sto</td>
<td>1.9</td>
<td>0.5</td>
<td>97.6</td>
</tr>
<tr>
<td>to</td>
<td>2.2</td>
<td>0.0</td>
<td>97.8</td>
</tr>
</tbody>
</table>

Table 8.4: Distribution of productions across MANNER OF ARTICULATION for each target word within the quadruple /so/, /to/, /do/ and /sto/. The total number of productions were 961. Values are in percent.
8.2 Manner of articulation in complex syllable onsets

The previous section provided a contrastive analysis of the differences in MANNER OF ARTICULATION for productions with a single consonant syllable onset. However, the analysis excluded productions with complex syllable onsets, which for the /skñ/, /spɔk/, /spɔɔ/ and /stɔ/ excluded the most adult-like productions. In this section, a contrastive comparison of the adult-like and not adult-like productions made with two consonants in the syllable onset is, therefore, presented. Before, the MANNER OF ARTICULATION for complex productions for adult-model simple-syllable onsets are considered, MANNER OF ARTICULATION for complex productions where they agree with the adult model are considered.

Table 8.5 presents the relative frequency for combinations of MANNER OF ARTICULATION for the first and second consonant in the syllable onset. A strong preference was found for the adult-like fricative+plosive production pattern. Alternative production patterns were observed. These included Fricative+Nasal productions observed in 2% of the productions (Table 8.5), however over a very limited time frame (Figure 8.5); Plosive+Fricative productions, resulting form a metathesis of the target consonant cluster, occurring in 1% of the productions; and close cases of metathesis, resulting in Plosive+Nasal productions, occurring in an additional 1% of the cases. Thus, it can be concluded that placement of plosive first in the sequence, increasing the cluster’s agreement with the sonority hierarchy, occurred in only 2% of the cases (8 out of 384 productions). The distribution of these production types against participant age is shown in Figure 8.5. Fricative+Plosive productions occur in a high relative frequency in structurally complex productions across the entire investigated age range. The alternative production patterns occur over the restricted time-frame of between 22 and 30 months of age. After 30 months, the correct MANNER OF ARTICULATION for the entire cluster occurs in 100% of the structurally complex.

Table 8.5: Summary table of the relative frequency of the combination of MANNER OF ARTICULATION for the first and second consonant in clusters produced for structurally simple target onsets. Values are in percent of the total number of two-consonant productions for a complex target.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approximant + Plosive</td>
<td>&lt;1</td>
</tr>
<tr>
<td>Fricative + Lateral</td>
<td>&lt;1</td>
</tr>
<tr>
<td>Fricative + Nasal</td>
<td>2</td>
</tr>
<tr>
<td>Fricative + Plosive</td>
<td>94</td>
</tr>
<tr>
<td>Nasal + Fricative</td>
<td>&lt;1</td>
</tr>
<tr>
<td>Plosive + Approximant</td>
<td>1</td>
</tr>
<tr>
<td>Plosive + Fricative</td>
<td>1</td>
</tr>
</tbody>
</table>

As stated earlier, structurally complex productions also occur in productions of target words with a structurally simple syllable onset. Table 8.6 presents the relative
Figure 8.5: Manner of Articulation for two-consonant onsets produced for complex target words. The total number of productions are 271.
frequency of combination of manner of articulations found in these non adult-like productions. Similar to Table 8.5, the descriptive statistics presented in Table 8.6 show an overall preference for Fricative+Plosive productions. However, this preference is observed only for target words with an initial fricative or an initial voiceless aspirated plosive. The production frequencies for voiced, unaspirated plosives are more evenly spread across manifested combinations.

The relative frequency of occurrence of the adult-like Fricative+Plosive sequence is similar in the productions of target words with an initial fricative or an initial voiceless aspirated plosive. However, Figures 8.6 and 8.7 present evidence that there are differences in the distribution of these productions across age. In target words with an [s] onset, the productions of Fricative+Plosive clusters occur mainly between ages 29–40 months (Figure 8.6). For voiceless plosive onsets, on the other hand, Fricative+Plosive sequences dominate the productions from 33 months onwards (Figure 8.7). Thus, overly complex productions of [s] and [pʰ] are observed to stabilise as Fricative+Plosive sequences paralleling the productions of target words with complex syllable onsets (Figure 8.5), but with a greatly lower frequency of occurrence. The overly complex productions made after 33 months are therefore not seen as signs of a robust discrimination between the cluster target word and the [s] and [pʰ] target words.

The perceptual confusion in the representation of target words with simple syllable onsets is evident in the large variety of output forms manifested. A comparison of Table 8.5 and Table 8.6 provides evidence that productions with an initial plosive occur relatively more frequently in the overly complex productions than in productions of s+plosive target words. Productions of Plosive+Fricative sequences occur in 24% of the voiceless aspirated targets onsets and 30% of the [s] target onsets. Plosive+Lateral productions occur in 12% of the voiceless aspirated targets onsets and 16% of the [s] target onsets. In contrast, Plosive+Lateral and Plosive+Fricative sequences occur relatively infrequently in productions where this could be interpreted as a metathesis of the target onset (Table 8.5). Thus, the overly complex syllable onsets produced are more likely to be changed into a form which agrees with the sonority hierarchy.

Table 8.6: Summary table of the relative frequency of the combination of MANNER OF ARTICULATION for the first and second consonant in clusters produced for structurally simple target onsets.

<table>
<thead>
<tr>
<th></th>
<th>[s] onset</th>
<th>Voiced onset</th>
<th>Voiceless onset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fricative + Lateral</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Fricative + Nasal</td>
<td>7</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Fricative + Plosive</td>
<td>43</td>
<td>20</td>
<td>41</td>
</tr>
<tr>
<td>Nasal + Approximant</td>
<td>0</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>Nasal + Plosive</td>
<td>0</td>
<td>20</td>
<td>6</td>
</tr>
<tr>
<td>Plosive + Approximant</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Plosive + Fricative</td>
<td>30</td>
<td>10</td>
<td>24</td>
</tr>
<tr>
<td>Plosive + Lateral</td>
<td>16</td>
<td>20</td>
<td>12</td>
</tr>
</tbody>
</table>
Figure 8.6: Manner of Articulation for two-consonant onsets produced for the /sak/, /sal/, /so/ and /sor/ target words. The total number of productions are 44.

Figure 8.7: Manner of Articulation for two-consonant onsets produced for the /pæk/, /pæk/, /tɔ/ and /tɔr/ target words. The total number of productions are 17.
8.3 Achievement of a stable production of manner

In the previous sections of this chapter, structurally simple and complex productions made of the simple and complex target syllable onsets have been investigated in terms of age and target word dependencies in the output forms. The focus has, therefore, been on the presentation of the diversity of the productions for a specific target word at a given age. In this section, the focus shifts to the acquisition of a stable production of adult-like Manner of Articulation.

In Figures 8.9–8.12, the progressive relative frequency of a correct manifestation of Manner of Articulation, for the entire syllable onset, is presented. The bar displayed at each age indicates the percentage of productions following this point in development which agree with the adult model sequence. Thus, in the quadruple displayed in Figure 8.9, the progression indicated for the target word /sæl/ indicates the agreement with a Fricative+Plosive sequence and the progression for /sæl/ indicates the agreement with a single Fricative syllable onset.

It can be observed in Figures 8.9–8.12 that the voiced plosive onsets are the first in each quadruple to acquire stable production of Manner of Articulation. Voiceless aspirated plosives also stabilize at a high relative frequency of production early in development but are confused more often than voiced plosives, lowering the progressive relative frequency values. Fricative onsets are initially not produced consistently as fricatives, but increase in production stability with age. As seen in Figures 8.1–8.4 and Figure 8.6, the observed difference between the plosive targets and the fricative targets are mainly caused by a high rate of stopping of the fricative and production of consonant clusters for the target fricative syllable onsets.
Figure 8.9: Progressive relative frequency of adult-like MANNER OF ARTICULATION for all elements in the onset for target words /sal/, /kal/, /gal/ and /skal/.

Figure 8.10: Progressive relative frequency of adult-like MANNER OF ARTICULATION for all elements in the onset for target words /sak/, /pak/, /bak/ and /spak/.
Figure 8.11: Progressive relative frequency of adult-like MANNER OF ARTICULATION for all elements in the onset for target words /sor/, /por/, /bor/ and /spor/.

Figure 8.12: Progressive relative frequency of adult-like MANNER OF ARTICULATION for all elements in the onset for target words /so/, /to/, /do/ and /sto/.
Table 8.7: Age of stable production of adult-like Manner of Articulation of target words. Each column presents the Age (in months) at which the rate of following productions increased above the level indicated in the column headers.

<table>
<thead>
<tr>
<th>Target word</th>
<th>50%</th>
<th>60%</th>
<th>70%</th>
<th>80%</th>
<th>90%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>stɔ</td>
<td>42</td>
<td>44</td>
<td>48</td>
<td>48</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>sʊ</td>
<td>29</td>
<td>37</td>
<td>50</td>
<td>50</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>do</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>30</td>
</tr>
<tr>
<td>to</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>47</td>
</tr>
<tr>
<td>spak</td>
<td>43</td>
<td>48</td>
<td>48</td>
<td>48</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>sak</td>
<td>22</td>
<td>29</td>
<td>40</td>
<td>48</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>bak</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>43</td>
</tr>
<tr>
<td>pak</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>24</td>
<td>48</td>
</tr>
<tr>
<td>skal</td>
<td>48</td>
<td>49</td>
<td>49</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>sal</td>
<td>30</td>
<td>37</td>
<td>42</td>
<td>43</td>
<td>49</td>
<td>49</td>
</tr>
<tr>
<td>gal</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>48</td>
</tr>
<tr>
<td>kal</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>48</td>
</tr>
<tr>
<td>spor</td>
<td>41</td>
<td>44</td>
<td>48</td>
<td>48</td>
<td>49</td>
<td>49</td>
</tr>
<tr>
<td>sor</td>
<td>34</td>
<td>41</td>
<td>41</td>
<td>48</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>bor</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>48</td>
</tr>
<tr>
<td>por</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>26</td>
<td>44</td>
</tr>
</tbody>
</table>

Based on the data underlying Figures 8.9–8.12, milestone production rates were extracted. A milestone was defined as the first achievement of the indicated progressive relative frequency. The milestones are presented in Table 8.7.

The most rapid acquisition in terms of Manner of Articulation is observed in the plosive onsets. This is not surprising, considering the prevalence of this Manner of Articulation in Tables 8.1–8.4, as well as the frequently noted early acquisition of plosives (e.g., Ingram (1976) and Grunwell (1987)). In addition, a trend for aspirated plosives to be acquired after voiced plosives is observed in two of the quadruples (the /dɔː/-/tɔː/ and /bʊk/-/pʊk/ distinctions). In contrast, the distinction between the /pɔr/ and /bɔr/ are acquired at the same rate. For /bɔr/ and /pɔr/, the 90% milestone was reached by voiced plosive before the aspirated plosive, but the 100% progressive production rate was first reached by the aspirated plosive.

From Table 8.7 it can also be observed that target words with a Fricative or Fricative+Plosive onset are acquired at a progressive relative frequency higher than 80% outside of the time frame investigated in this Thesis. One exception exists, /sʊk/, that reached 80% at 40 months (3;4 years), eight months before the other Fricative and Fricative+Plosive target words. However, the 90% progressive production rate of /sʊk/ is not reached at the age of 48 months. At that time the progression of /sʊk/ is aligned with the other Fricative or Fricative+Plosive target words. Thus, /sʊk/ productions initially have a steeper acquisition curve, but they align with the other target words with a Fricative in the onset in the later stages of development.

Stable production of manner in cluster onset target words is acquired later than in single plosive and single fricative syllable onsets. For /spak/ and /spɔr/, the progressive
Table 8.8: Delay in stable production of adult-like MANNER OF ARTICULATION compared to production of complex syllable onsets. Each column presents the Age (in months) at which the rate of following productions increased above the level indicated in the row labels.

<table>
<thead>
<tr>
<th>% CC</th>
<th>skal</th>
<th>spak</th>
<th>spor</th>
<th>sto</th>
</tr>
</thead>
<tbody>
<tr>
<td>50%</td>
<td>28</td>
<td>22</td>
<td>16</td>
<td>22</td>
</tr>
<tr>
<td>60%</td>
<td>21</td>
<td>18</td>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>70%</td>
<td>14</td>
<td>12</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>80%</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100%</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Relative frequency reaches 100% of the productions at the age of 48 and 49 months respectively. Productions of /st/ do not reach the 90% level in the investigated age period, and /sk/ does not reach the 80% production rate.

The delay in acquisition of manner for complex target onsets relative to the acquisition of manifestation of complexity varies depending on the milestone selected. Table 8.8 presents the difference, in months, between the child achieving a milestone in complexity (Table 7.1) and the time of achievement of the same milestone in manner (Table 8.7). As Table 8.8 shows, the delay in MANNER OF ARTICULATION at the 50% milestone is between 16–28 months, at the 60% milestone between 7–21 months, at the 70% milestone between 5–14 months and at the 80% milestone between 4–6 months. After the 80% milestone, the developmental curves for initial consonant cluster complexity and manner of articulation aligned at the same level of stability.

8.4 Discussion

Previous investigations have shown the feature of MANNER OF ARTICULATION to be perceptually stable; results from investigations of the perception of phonetic features in constrained listening conditions by Singh (1971) and Singh & Black (1966) provided evidence that perception of MANNER OF ARTICULATION is more stable compared to PLACE OF ARTICULATION. In a developmental model where constraints in perception are strong factors underlying the process of acquisition, the perceptual robustness of MANNER OF ARTICULATION is proposed to afford an early adult-like manifestation of this feature.

In agreement with previous research on consonant acquisition in productions, plosive production was shown to be acquired earliest. Plosives were, however, not produced only in the contexts where it agreed with the adult model. A strong presence of fricatives being produced as plosives was also observed across all target words within a contrasted quadruple.

Fricatives were, however, present in the productions. For target words with an initial single fricative, a reduction in the relative frequency of plosive productions with age is observed, paralleled with an increase in fricative and other continuant productions. For
Manner of Articulation

fricative targets, fricative and plosive productions dominate, with a relative frequency of between 44% and 50.6% for fricatives, and between 47.7% and 55.2% for plosives across the investigated age range. Approximant, lateral and nasal productions account for between 0% and 3.9% of the total number of productions (Tables 8.1–8.4). Thus, the data presented here provides strong evidence for a dominance of the stopping process in the reduction of singleton fricative onsets.

A preference for production of plosives is also observed in structurally simple manifestations of s+plosive target onsets. Thus, a prevalence of reduction of s+plosive consonant clusters towards the plosive was found rather than the fricative observed in previous investigations of early manifestations of s+plosive clusters. Reductions to a single fricative do, however, occur in up to 5.6% of the productions of s+plosive target onsets (Tables 8.1–8.4), with no apparent trend across age (Figures 8.1–8.4). Thus, the data presented in this chapter do not provide any evidence that fricative productions of s+plosive targets form part of a developmental trend.

Simple onset fricative productions do not only occur in productions towards a simple syllable fricative onset; Tables 8.1–8.4 show a relative frequency of occurrence of fricatives in productions towards aspirated plosives that is comparable to, or larger than, fricatives in productions towards s+plosive clusters (2.2–7.4%).

The substitution of plosives for fricatives is surprising, since there is no production reason for the easier production pattern of plosives, manifested by the strong presence of the stopping process, to be substituted by a process that produces the more difficult fricative production when the target segment is a plosive. An explanation may lie in both the child’s and the adult observer’s perception. As shown in Chapter 4–6, the period investigated in this Thesis involves an increase in aspiration production. Fricative consonants like aspiration are identified acoustically by an aperiodic portion in the signal. It is, therefore, possible that the productions, perceived as fricatives by the transcriber, could have be productions of aspiration that failed to meet the perceptual criteria of the adult listener. Alternatively, the aspiration produced in the adult model may have been mistaken for a fricative when perceived by the child. These explanations should be considered speculative; further and more detailed is needed to tease these two hypotheses teased apart.

After the initial period where consonant clusters are reduced to a single consonant, complex syllable onsets start to be formed. A strong dominance of the Fricative+Plosive sequence was observed once complex syllable onsets had been established (Table 8.5). Alternative production patterns were, however, manifested. These alternatives involved substitution of the manner of articulation of one of the consonants or a substitution of the manner of articulation of one consonant together with a reversal of the sequence (Plosive+Fricative). The most frequently occurring substitution was a Fricative+Nasal sequence. These substitutions, however, occurred during a limited frame in Development and mainly occurred before the age of 30 months (Figure 8.5).

Two-element clusters manifested after the age of 30 months are produced in an adult-like manner in terms of MANNER OF ARTICULATION, indicating a stabilised proficiency in the production of the Plosive+Fricative sequence. This in turn indicates that a perceptual effect behind the presence of productions transcribed as single-element clusters in Figures 8.1–8.4 after the age of 30 months is plausible. It is argued that children may confuse the target form to be produced, resulting in a structurally simple syllable onset. However, when the form is correctly perceived, the entire sequence
in MANNER OF ARTICULATION is manifested in an adult-like manner. This claim is given further support by the manifestations of overly complex productions of target words with an initial fricative (Figure 8.6) or an initial aspirated plosive (Figure 8.7), as these productions are either Fricative+Plosive or Plosive+Fricative after the age of 29 months in Figure 8.6 and after 40 months in Figure 8.7. If these productions were caused by a maturing articulatory system, it is argued that the productions of overly complex syllable onsets would be minimal, especially after the age when this production pattern has been established in an adult-like manner in productions of structurally complex targets.

Over time, the productions increase in level of likeness with the adult target model. As was shown in Figures 8.9–8.12, stable productions in terms of manner were first achieved in plosive targets, followed by fricatives in singleton onsets and later clusters. As seen in the calculated age of achievement milestone production rates (Table 8.7), cluster production rates reached levels of over 80% after the age of 4 years (48 months). For two out of the four cluster target words, levels of over 80% are not reached in the time frame investigated in this Thesis. In addition, it should be noted that the production rates on which the calculated milestones are based are optimistic estimates due to the upper truncation of the age range. Thus, the real acquisition rate, when collected from a more extended period of development, may show slightly lower estimations of the production rate.

When comparing the achievements in MANNER OF ARTICULATION with the overall achievement of structural complexity, tentative signs of a parallel development can be observed. As shown in Table 8.8, the achievement of a 50% level of correct production in terms of manner is delayed by between 16 months and 28 months compared to structural complexity. However, this developmental gap decreased with each milestone reached, and not present at all in the target words that reached the 90% milestone (Table 8.8); This effect may, however, be caused by the truncation imposed. Further research, conducted on a larger group of participants and one that has an older upper age is needed in order to determine whether this is a sign of a true merger, establishing an adult-like production pattern, or an artifact of the method used.
Acquisition of place of articulation

A consonant’s place of articulation may move relative to the adult model in initial productions made by children. For example, plosives have been noted to be produced in the front of the mouth, resulting in a fronting process and in target words with two plosives, a spreading or assimilation of velar place of articulation to non-velar plosives has been observed (Ingram 1976, Grunwell 1987). This chapter investigates the manifestation of place in simple and complex syllable onsets.

In Section 9.1 the output forms produced as singleton consonants are compared to the other targets within the quadruple. Then in Section 9.2 the Place of Articulation of the first and second consonant in complex productions is investigated. Productions with adult-like complexity are contrasted across onset types and comparisons are made with overly complex productions. In Section 9.3 the focus shifts to the stabilisation of adult-like places of articulation before in Section 9.4 the results regarding the acquisition of a contrast in Place of Articulation are discussed.

9.1 Place of articulation in simple syllable onsets

It was noted in Chapter 7 that productions with structurally simple syllable onsets may occur both for target words with a simple syllable onset and target words with a complex syllable onset. In this section, the Place of Articulation for these structurally simple target onsets is investigated by contrastive analysis of the manifestation of the target words. (Figures 9.1–9.4).

For the skal quadruple (/skal/, /kal/, /sakl/ and /gkal/), the distribution of productions across Place of Articulation is displayed in Table 9.1. A majority of the productions of this quadruple are perceived to have been made at the velar Place of Articulation. However, in accordance with previous investigations (Ingram 1976, Grunwell 1987), a strong presence of dental/alveolar productions are also present in the UC3 Corpus for this velar plosive quadruple. For the target words within the skal quadruple, dental and alveolar places of articulation are most often used in the target word /skal/, and productions of /gkal/ are produced slightly more often as a dental than /kal/. Thus, in contrast to the manifestation of manner (Table 8.1),
the manifestations of voiced plosives are closer to the reduced version of the cluster than the aspirated plosives.

As the quadruple /spək/, /spɒk/, /spæk/ and /brək/ have velar codas, a frequent substitution of the syllable onset labial plosives for velar plosives due to the application of the velar harmony process might be expected. However, Table 9.2 shows that does not occur as frequently as might be expected in the target words where both labial and velar plosives are present. Unexpectedly, velar productions (all of them plosives) in the onset are observed relatively more frequently for the /sək/ target word. These velar productions, however, cease to occur in the recorded speech samples approximately after the age of 36 months for all target words (See Figure 9.2). The last occurrence is observed at 36 months for /sək/, 31 months for /spək/ and 24 months for /brək/ (a single occurrence). Thus, after the age of three years, the children never utilise the velar spreading process. This observation agrees well with the one made previously for English speaking children that this is an early process in the development of speech in children.
Table 9.1: Distribution of productions across Place of Articulation for each target word within the quadruple /sa1/, /ka1/, /ga1/ and /ska1/. The total number of productions were 750. Values are in percent.

<table>
<thead>
<tr>
<th></th>
<th>Alveolar</th>
<th>Dental</th>
<th>Glottal</th>
<th>Labial</th>
<th>Laryngal</th>
<th>Palatal</th>
<th>Retroflex</th>
<th>Velar</th>
</tr>
</thead>
<tbody>
<tr>
<td>gal</td>
<td>0.5</td>
<td>12.3</td>
<td>1.8</td>
<td>2.3</td>
<td>0.0</td>
<td>0.9</td>
<td>82.2</td>
<td></td>
</tr>
<tr>
<td>kal</td>
<td>0.5</td>
<td>9.8</td>
<td>0.0</td>
<td>1.5</td>
<td>2.9</td>
<td>0.5</td>
<td>84.8</td>
<td></td>
</tr>
<tr>
<td>sal</td>
<td>34.9</td>
<td>30.7</td>
<td>0.0</td>
<td>3.6</td>
<td>10.8</td>
<td>16.3</td>
<td>3.6</td>
<td></td>
</tr>
<tr>
<td>skal</td>
<td>0.6</td>
<td>20.5</td>
<td>1.2</td>
<td>5.0</td>
<td>0.0</td>
<td>0.0</td>
<td>72.7</td>
<td></td>
</tr>
</tbody>
</table>

Table 9.2: Distribution of productions across Place of Articulation for each target word within the quadruple /sa1/, /pa1/, /ba1/ and /spa1/. The total number of productions were 854. Values are in percent.

<table>
<thead>
<tr>
<th></th>
<th>Alveolar</th>
<th>Dental</th>
<th>Glottal</th>
<th>Labial</th>
<th>Laryngal</th>
<th>Palatal</th>
<th>Retroflex</th>
<th>Velar</th>
</tr>
</thead>
<tbody>
<tr>
<td>bak</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>98.7</td>
<td>0.9</td>
<td>0.0</td>
<td>0.0</td>
<td>0.4</td>
</tr>
<tr>
<td>pak</td>
<td>0.0</td>
<td>1.1</td>
<td>0.0</td>
<td>95.3</td>
<td>3.2</td>
<td>0.0</td>
<td>0.4</td>
<td>0.0</td>
</tr>
<tr>
<td>sak</td>
<td>38.5</td>
<td>20.7</td>
<td>1.1</td>
<td>4.6</td>
<td>12.6</td>
<td>1.1</td>
<td>10.9</td>
<td>10.3</td>
</tr>
<tr>
<td>spak</td>
<td>1.2</td>
<td>1.2</td>
<td>1.2</td>
<td>91.8</td>
<td>2.3</td>
<td>0.0</td>
<td>0.6</td>
<td>1.8</td>
</tr>
</tbody>
</table>

Except for the velar productions in the target word /sa1/, productions of the target words within the spak quadruple with a labial plosive are manifested in terms of manner early in development. Confusions, when present, do not show a pattern of being systematic in their distribution across age and target word. Thus, no signs of a maturation effect is observed in these productions.

The frequency of velar spreading in the spak quadruple of target words was found to be an infrequently occurring process. However, comparison of the summary presented in Table 9.2 with the summary presented in Table 9.3, shows that back places of articulation, such as Velar, Laryngeal and Glottal, occur more frequently when the syllable coda consists of a velar. In the /spor/ quadruple, which has labial plosives in the syllable onsets, but no velar plosive in the codas, back productions occur less frequently. In all other respects, the productions are similar in distribution. Thus even though the velar spreading process does not show a strong presence in the spak quadruple, the presence is stronger in this quadruple than in the /spor/ quadruple. Velar spreading, though infrequent, has been shown to increase in frequency when placed in an appropriate context.

For target words within the /stot/ quadruple, a strong preference for fronted productions is observed (See Table 9.4). The dental Place of Articulation receives the highest relative frequency of productions for target words with a dental plosive in the syllable onset. Structurally reduced forms of the cluster target word stabilise in terms of place earlier than the other target words. Non-adult productions, where dental and alveolar productions are pooled, do not occur after the age of 34 months (See
Figure 9.2: Relative frequency of occurrence for each Place of Articulation in simple productions of the /sak/, /pak/, /bak/ and /spak/ target words.

Table 9.3: Distribution of productions across Place of Articulation for each target word within the quadruple /sor/, /por/, /bor/ and /spor/. The total number of productions were 1036. Values are in percent.

<table>
<thead>
<tr>
<th>Word</th>
<th>Alveolar</th>
<th>Dental</th>
<th>Glottal</th>
<th>Labial</th>
<th>Laryngal</th>
<th>Palatal</th>
<th>Retroflex</th>
<th>Uvular</th>
<th>Velar</th>
</tr>
</thead>
<tbody>
<tr>
<td>bor</td>
<td>0.0</td>
<td>2.0</td>
<td>0.8</td>
<td>96.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.4</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>por</td>
<td>0.0</td>
<td>0.9</td>
<td>0.3</td>
<td>94.5</td>
<td>4.3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>sor</td>
<td>37.2</td>
<td>23.2</td>
<td>0.0</td>
<td>10.1</td>
<td>9.7</td>
<td>2.9</td>
<td>14.0</td>
<td>0.0</td>
<td>2.9</td>
</tr>
<tr>
<td>spor</td>
<td>1.2</td>
<td>6.4</td>
<td>0.8</td>
<td>90.0</td>
<td>0.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.8</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Figure 9.3: Relative frequency of occurrence for each Place of Articulation in simple productions of the /sor/, /por/, /bor/ and /spor/ target words.
Figure 9.4). In the single plosive target words, however, substitutions of place occur until the age of 47 months. Thus, structurally reduced productions of [st] stabilise earlier as dental or alveolar than the syllable target onsets consisting of aspirated or voiced plosives.

Figure 9.4: Relative frequency of occurrence for each Place of Articulation in simple productions of the /so/, /to/, /do/ and /sto/ target words.

9.2 Place of articulation in complex syllable onsets

It was observed in Chapter 7 that productions with a structurally complex syllable onsets occurred for target words with either simple or complex syllable onsets in the UC³ Corpus. In this section, the Place of Articulation manifested in these two-consonant sequences is considered.

The two-consonant sequences produced for target words with a complex syllable onset are presented in Figure 9.5 for [sk] onsets, in Figure 9.6 for [sp] onsets and Figure 9.7 for [st] onsets. The most frequently occurring production pattern for [sk] is the adult-like sequence of an alveolar consonant followed by a velar consonant (See Table 9.5). The same is true for [sp] targets (Table 9.6) and [st] targets (Table 9.7).
Table 9.4: Distribution of productions across Place of Articulation for each target word within the quadruple /so/, /to/, /do/ and /sto/. The total number of productions were 961. Values are in percent.

<table>
<thead>
<tr>
<th></th>
<th>Alveolar</th>
<th>Dental</th>
<th>Glottal</th>
<th>Labial</th>
<th>Laryngeal</th>
<th>Retroflex</th>
<th>Velar</th>
</tr>
</thead>
<tbody>
<tr>
<td>do</td>
<td>0.0</td>
<td>91.6</td>
<td>0</td>
<td>3.5</td>
<td>0.0</td>
<td>0.4</td>
<td>4.4</td>
</tr>
<tr>
<td>so</td>
<td>37.5</td>
<td>34.0</td>
<td>0</td>
<td>1.5</td>
<td>6.9</td>
<td>17.8</td>
<td>2.3</td>
</tr>
<tr>
<td>sto</td>
<td>1.0</td>
<td>87.9</td>
<td>1</td>
<td>2.4</td>
<td>1.0</td>
<td>1.0</td>
<td>5.8</td>
</tr>
<tr>
<td>to</td>
<td>0.7</td>
<td>90.0</td>
<td>0</td>
<td>4.8</td>
<td>0.7</td>
<td>0.4</td>
<td>3.3</td>
</tr>
</tbody>
</table>

This indicates that the investigated children manage to acquire an adult-like production pattern in terms of place of the articulation for both consonants within the investigated timeframe. Adult-like production of place for cluster targets is achieved to a lower degree than for Manner of Articulation (c.f. Table 8.5).

Table 9.5 shows that some of the fricatives analysed as adult-like in terms of manner of articulation in Chapter 8 are produced as a retroflex. As shown in Tables 9.5–9.7 this production pattern is the second most frequently occurring pattern in productions of complex syllable onsets manifested as two consonants, and exhibits only a small variation in percent occurrence across target onset type. The retroflexion of the initial consonant is, thus, not dependent on the following consonant. No clear age preference in the distribution of this process is observed.

The variation in manifested place of articulation is larger than for Manner of Articulation in the productions in the UC3 Corpus. However, Tables 9.5–9.7 show that the relative frequency of alternative production patterns only occur in 6% or less of the two-consonant productions of target words with a complex syllable onset. Further, an age trend for these productions is not observed in Figures 9.5–9.7. Thus, the presence of these less frequent alternative production patterns may be the result of the random sampling from a larger set of possible values in the phonetic description of place compared to Manner of Articulation.

### 9.2.1 Place of articulation in overly complex productions

Target words with a structurally simple syllable onset may be produced as a structurally complex syllable onset. The Manner of Articulation of these sequences was discussed in Section 8.2. As a complement to this description, the Place of Articulation of overly complex productions was investigated in terms of relative occurrence frequency for each Place of Articulation sequence produced for target words with structurally simple syllable onsets (Table 9.8).

The Place of Articulation sequences presented in Table 9.8 show only partial evidence of systematic use. Confusion occurs most frequently in target words with an initial velar plosive. The voiced variant of this syllable onset was manifested as an adult-like Alveolar+Velar sequence in 33% of the cases. The voiceless aspirated velar plosive was produced as an adult-like place-sequence in 25% of the cases and as the reversed sequence in 25% of the productions. Thus, there is evidence of a production...
Figure 9.5: Place of Articulation for two-consonant onsets produced for [sk]-target words. The total number of productions are 79.

Table 9.5: Summary table of the relative frequency of the combination of Place of Articulation for the first and second consonant in clusters produced for [sk]-target onsets. Values are in percent of the total number of two-consonant productions for a complex target.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alveolar + Alveolar</td>
<td>1</td>
</tr>
<tr>
<td>Alveolar + Dental</td>
<td>6</td>
</tr>
<tr>
<td>Alveolar + Velar</td>
<td>47</td>
</tr>
<tr>
<td>Dental + Velar</td>
<td>1</td>
</tr>
<tr>
<td>Glottal + Velar</td>
<td>1</td>
</tr>
<tr>
<td>Laryngal + Dental</td>
<td>5</td>
</tr>
<tr>
<td>Laryngal + Velar</td>
<td>1</td>
</tr>
<tr>
<td>Retroflex + Dental</td>
<td>5</td>
</tr>
<tr>
<td>Retroflex + Velar</td>
<td>28</td>
</tr>
<tr>
<td>Velar + Velar</td>
<td>4</td>
</tr>
</tbody>
</table>
Figure 9.6: **Place of Articulation** for two-consonant onsets produced for complex [sp]-target words. The total number of productions are 213.

Table 9.6: Summary table of the relative frequency of the combination of Place of Articulation for the first and second consonant in clusters produced for [sp] target onsets. Values are in percent of the total number of two-consonant productions for a complex target.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alveolar + Dental</td>
<td>&lt;1</td>
</tr>
<tr>
<td>Alveolar + Labial</td>
<td>62</td>
</tr>
<tr>
<td>Glottal + Palatal</td>
<td>2</td>
</tr>
<tr>
<td>Labial + Alveolar</td>
<td>1</td>
</tr>
<tr>
<td>Labial + Labial</td>
<td>3</td>
</tr>
<tr>
<td>Laryngal + Dental</td>
<td>1</td>
</tr>
<tr>
<td>Laryngal + Glottal</td>
<td>2</td>
</tr>
<tr>
<td>Laryngal + Labial</td>
<td>5</td>
</tr>
<tr>
<td>Retroflex + Labial</td>
<td>23</td>
</tr>
</tbody>
</table>
Figure 9.7: **Place of Articulation** for two-consonant onsets produced for complex [st]-target words. The total number of productions are 92.

Table 9.7: Summary table of the relative frequency of the combination of **Place of Articulation** for the first and second consonant in clusters produced for [st] target onsets. Values are in percent of the total number of two-consonant productions for a complex target.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alveolar + Alveolar</td>
<td>1</td>
</tr>
<tr>
<td>Alveolar + Dental</td>
<td>59</td>
</tr>
<tr>
<td>Dental + Alveolar</td>
<td>1</td>
</tr>
<tr>
<td>Dental + Dental</td>
<td>1</td>
</tr>
<tr>
<td>Laryngal + Dental</td>
<td>9</td>
</tr>
<tr>
<td>Retroflex + Dental</td>
<td>28</td>
</tr>
<tr>
<td>Velar + Velar</td>
<td>1</td>
</tr>
</tbody>
</table>
of adult-like, or possibly reversed, place sequences which may be interpreted as signs of confusion with the place of the [sk] cluster.

For labial and dental plosives, the manifestations of adult-like or near adult-like productions of the cluster target onset within the same quadruple occur less frequent than for velar plosives. For labial plosives, the most frequent place sequence is Velar+Labial. The adult-like sequence Alveolar+Labial never occurs, but near adult-like sequence Dental+Labial occurs in 20% of the voice labial plosive productions and 7% voiceless labial plosives productions.

For dental onsets, the correspondence between the overly complex production and the cluster target within the same quadruple is less pronounced. The Alveolar+Dental sequence is produced only in productions of [t] syllable onsets with the remaining productions made as Dental+Palatal, which bears no resemblance with the cluster onset within the quadruple. The onset [t] is never produced as an Alveolar+Dental sequence, but is produced as a slightly retracted Retroflex+Dental or Dental+Retroflex sequence in 50% of the productions.

The overly complex productions also display poor agreement with the productions made for complex target words. Comparing the overly complex productions with adult-like complexity within the skal quadruple shows that 50% of the sequences manifested in overly complex productions are not manifested in attempted productions of /skal/ (Figure 9.5). In addition, 70% of the sequences manifested in the /skal/ productions are not found in the overly complex productions of target words in the same quadruple. Similar observations can be made for labial and dental targets.

Thus, in contrast to the manifestation of MANNER OF ARTICULATION, the overly complex productions made by the children show fewer clear signs of confusion of target words. The place sequences in overly complex productions do not agree well with the adult production of the complex target words or with the children’s manifestations of the complex target words in the UC^3 Corpus.

9.3 Achievement of stable production of place

A visual representation of the progression towards adult-like PLACE OF ARTICULATION is provided in Figures 9.8–9.11, one figure for each quadruple. Each bar indicates the percentage of adult-like sequences produced relative to the total number of productions of the target word made after the indicated point in development. Thus, the bar displayed at 40 months in the top-left panel of Figure 9.8 shows the percentage of occurrence of a single Alveolar consonant produced towards the target word /skl/ after the age of 39 months. The corresponding bar shown at 40 months in the /skl/ panel indicates the percentage of occurrence of a Alveolar+Velar sequence after the age of 39 months. The age at which certain milestones are reached in terms of progressive relative occurrence is shown in Table 9.9.

Similarly to the progression of stability in manner (see Section 8.3), PLACE OF ARTICULATION is first achieved in the target words with a single plosive onset. The target words /tɔ/, /dɔ/, /pɔk/, /kɔl/ and /gɔl/ achieved an 80% production rate and and /bɔk/, /bɔr/ and /pɔr/ a 90% production rate at the onset of the study (18 months of age). A 100% production rate was achieved at 48 months for most single plosive targets, though some targets reach this milestone earlier; the target /bɔk/ reached 100% at 43 months, /pɔr/ at 44 months and /bɔr/ at 36 months. Thus, by the age of
Table 9.8: Summary table of the relative frequency of the combination of Place of Articulation for the first and second consonant in clusters produced for structurally simple target onsets.

<table>
<thead>
<tr>
<th></th>
<th>s</th>
<th>b</th>
<th>p</th>
<th>d</th>
<th>t</th>
<th>g</th>
<th>k</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alveolar + Dental</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>33</td>
<td>0</td>
<td>33</td>
<td>0</td>
</tr>
<tr>
<td>Alveolar + Labial</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Alveolar + Laryngal</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Alveolar + Velar</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>33</td>
<td>25</td>
</tr>
<tr>
<td>Dental + Alveolar</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>33</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Dental + Labial</td>
<td>0</td>
<td>20</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Dental + Palatal</td>
<td>0</td>
<td>0</td>
<td>67</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Dental + Retroflex</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>25</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Glottal + Alveolar</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Labial + Alveolar</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Labial + Dental</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Labial + Velar</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Laryngal + Alveolar</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Laryngal + Dental</td>
<td>6</td>
<td>0</td>
<td>20</td>
<td>0</td>
<td>25</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>Laryngal + Labial</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Laryngal + Velar</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>Retroflex + Alveolar</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Retroflex + Dental</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>25</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Retroflex + Glottal</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Retroflex + Labial</td>
<td>4</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Velar + Alveolar</td>
<td>9</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>Velar + Dental</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Velar + Labial</td>
<td>0</td>
<td>60</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 9.9: Age of stable production of adult-like place of articulation of target words. Each column presents the age (in months) at which the rate of following productions increased above the level indicated in the column headers.

<table>
<thead>
<tr>
<th>Target word</th>
<th>50%</th>
<th>60%</th>
<th>70%</th>
<th>80%</th>
<th>90%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>sto</td>
<td>43</td>
<td>44</td>
<td>48</td>
<td>48</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>so</td>
<td>34</td>
<td>48</td>
<td>50</td>
<td>50</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>do</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>24</td>
<td>48</td>
</tr>
<tr>
<td>to</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>38</td>
<td>48</td>
</tr>
<tr>
<td>spak</td>
<td>43</td>
<td>48</td>
<td>48</td>
<td>48</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>sak</td>
<td>33</td>
<td>36</td>
<td>42</td>
<td>48</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>bak</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>43</td>
</tr>
<tr>
<td>pak</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>25</td>
<td>48</td>
</tr>
<tr>
<td>skal</td>
<td>48</td>
<td>49</td>
<td>49</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>sal</td>
<td>34</td>
<td>37</td>
<td>42</td>
<td>43</td>
<td>49</td>
<td>49</td>
</tr>
<tr>
<td>gal</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>33</td>
<td>48</td>
</tr>
<tr>
<td>kal</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>spor</td>
<td>41</td>
<td>44</td>
<td>48</td>
<td>48</td>
<td>49</td>
<td>49</td>
</tr>
<tr>
<td>sor</td>
<td>36</td>
<td>41</td>
<td>48</td>
<td>48</td>
<td>48</td>
<td>48</td>
</tr>
<tr>
<td>bor</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>36</td>
</tr>
<tr>
<td>por</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>44</td>
</tr>
</tbody>
</table>

48 months, production stability in Place of Articulation has been reached in all single plosive targets.

In agreement with the acquisition of manner (Section 8.3), stable production of Place of Articulation in the single fricative is delayed in development relative to the single plosive target onsets. Table 9.9 indicates a progressive relative frequency of 60–90% for fricative target words at the point of development where all a plosive target words have reached the 100% milestone. In addition, the target word /so/ fails to reach a stable Place of Articulation due to the strong presence of retroflex fricatives observed in Figures 9.1–9.4.

Not surprisingly, Place of Articulation for complex syllable onsets is acquired late in each quadruple. Two of the complex target words, /skal/ and /sto/ failed to reach the 100% production milestone. For /spak/ and /spor/, a rapid increase from 50% or 60% to 100% around the age of 48 months. However, it should be noted that this rapid increase may in part be an artifact of the natural truncation imposed due to the lower number of recordings conducted beyond this age. The true point of achieving a 100% progressive production rate may therefore be shown to be delayed relative to what is presented here in a more extensive investigation covering a larger age range.

9.4 Discussion

Transformations in terms of Place of Articulation are expected to occur in early productions of consonants. Processes frequently observed in manifestations of plosives
Figure 9.8: Progressive relative frequency of adult-like Place of Articulation for all elements in the onset for target words /sal/, /kal/, /gal/ and /skal/.

Figure 9.9: Progressive relative frequency of adult-like Place of Articulation for all elements in the onset for target words /sak/, /pak/, /bak/ and /spak/.
Figure 9.10: Progressive relative frequency of adult-like Place of Articulation for all elements in the onset for target words /sor/, /por/, /bor/ and /spor/.

Figure 9.11: Progressive relative frequency of adult-like Place of Articulation for all elements in the onset for target words /so/, /to/, /do/ and /sto/. 
Table 9.10: Delay in stable production of adult-like Place of Articulation compared to production of Manner of Articulation in complex syllable onsets. Each column presents the delay in months for Place of Articulation relative to Manner of Articulation in the achievement of a progressive relative frequency indicated in the row labels.

<table>
<thead>
<tr>
<th>% CC</th>
<th>50%</th>
<th>60%</th>
<th>70%</th>
<th>80%</th>
<th>90%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>sto</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>so</td>
<td>5</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>do</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>18</td>
</tr>
<tr>
<td>to</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>spak</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>sak</td>
<td>11</td>
<td>7</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>bak</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>pak</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>skal</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sal</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>gal</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>kal</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td>spor</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>sor</td>
<td>2</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>bor</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>−12</td>
</tr>
<tr>
<td>por</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>−8</td>
<td>0</td>
</tr>
</tbody>
</table>

and fricatives involve context insensitive fronting (Grunwell 1987, Ingram 1976), den- talisation, and the context sensitive process of velar harmony, which is triggered by a velar consonant present in the word (Grunwell 1987). Thus, development in terms of Place of Articulation is regarded as movement away from usage of these processes in production, towards more adult-like production patterns.

The phonological process of fronting has been shown to occur in early productions and reduce in frequency before the age of 3 years (Grunwell 1987). The processes of fronting and velar spreading both occur in UC3 Corpus. For target words in the skal quadruple, the fronting process occurs in 11.3–25.5% of the productions of a target word with a velar plosive (Table 9.1).

Dental plosives, in comparison, show an more resistant fronting; movement from a dental to a labial Place of Articulation occurs in only 2.4–4.8% of the productions of target words /st0/, /sto/ and /d0/. The distribution of these fronted productions covers most of the age range, but cease to occur between 44–46 months of age for the skal quadruple (Figure 9.1) and at 47 months for /st0/ (Figure 9.4). Thus, by the end of the fourth year, fronting of plosives are no longer perceived by the adult observer, which is a full year later than was noted by Ingram (1976) and Grunwell (1987) for children acquiring English.

The process of velar harmony has been argued to co-occur in time with the fronting process and to decrease in frequency to minimum by the end for the third year. It has been argued that this process is triggered by a velar plosive being present in the
produced word. Hence this process was expected to occur more frequently in spak quadruple. This expected result, however, only occurred in productions of fricatives. As shown in Table 9.2, 10.3% of the productions of fricatives are made at the velar Place of Articulation, which is substantially higher than for /sɔr/ (2.9%, as seen in Table 9.3). Productions of /sɔk/ at the laryngeal and velar Places of Articulation account for 22.9% of the productions, which is higher than the relative frequency of dental productions. It is also higher than the combined relative frequency of laryngeal and velar productions for /sɔr/, which was found to be 12.6% (Table 9.3). Thus, evidence is provided of an increase in productions of fricatives in the back of the oral cavity when the target word contains a velar consonant in the syllable coda, yet no evidence was found that a similar process acted on plosives in UC³ Corpus recordings.

For plosives, only productions of /spɔk/ show movement towards a velar Place of Articulation. The relative frequency of occurrence of this process is between 0.4–1.8% of the productions, compared to 0% for targets in the spår quadruple (Tables 9.2 and 9.3). The combined production rate at the laryngeal and velar Places of Articulation is comparable across the spak and spår quadruples (0–4.3% vs. 1.3–5.1%).

Thus, it is argued that the combined effect of backing of obstruents and the infrequent substitution of plosives for other manners of articulation (Chapter 8) pushes some of the plosives back to the only available place where plosives are produced, i.e. the velar Place of Articulation. For fricatives, more options are available, which creates a more general backing of the consonant produced. Thus, the evidence indicates that the process described as velar harmony in the literature may possible also be described as a more general backing process triggered by a velar consonant.

Not all movements in terms of Place of Articulation observed in the UC³ Corpus are uni-directional. Some of the movements are directed towards the dental Place of Articulation, regardless of whether this is fronting or not. For the /pɔr/ and /bɔr/ targets, 0.9–2% of the productions are made as dental plosives (Table 9.3), which are more retracted than the labial targets. The corresponding frequency for the /pɔk/ and /bɔk/ target words is 1.1–1.2% (Table 9.2). This dentalization process occurs in the data set until the age of 31–36 months in the /spɔr/ quadruple and 22-26 months in the spak quadruple. Hence, the dentalization of plosives regardless of target stops occurring earlier than the unidirectional fronting process.

Productions of target words containing a single fricative in the syllable onset are most often produced at the alveolar Place of Articulation (37-38.5% of the productions). The most frequent substitution is the dental Place of Articulation (21.7–43%), followed by a retroflex production (10.9–17.8%). Target words with an [s] onset are thus produced in the front of the mouth in a majority of productions, indicating little variation from the adult target Place of Articulation. This is in contrast to what was found for Manner of Articulation, where the manner of the fricative target was frequently changed to a plosive (Tables 8.1–8.4). Thus, in contrast to plosives, the Place of Articulation of fricatives is relatively more robustly manifested than Manner of Articulation.

With time, the productions of cluster target words start to be produced as a consonant cluster onset. In the early stages of these productions, the syllable onsets manifested may, however, differ in the sequence of Place of Articulation of the elements from the adult target. It is shown in Tables 9.5–9.7, that the adult-like Manner of
**Articulation** is the dominant manifestation across the entire age range. The second most frequently occurring sequence is the adult-like sequence with the initial element produced as a retroflex rather than an alveolar. Substitution for a dental consonant occurs only in productions of [st] onsets (Table 9.7).

The reversed adult-like production (resulting from a metathesis) does not occur frequently. Fully reduced sequences are observed in 1% of the productions of [sp] and [st] target words (Tables 9.6 and 9.7). Productions where the adult sequence is reversed and the second segment (first in the adult sequence) substituted by a **Place of Articulation** one step away from that of the adult target also occur in the UC³ Corpus. The set of single-feature metathesis includes 3% of the productions of [sp] and 1% of the productions of [st]. Thus, 2-4% of the productions of [sp] and [st] and 0% of the productions of [sk] may be viewed as potentially resulting from a metathesis process affecting **Place of Articulation**. Productions of metathesis or single-feature metathesis in complex target words therefore to occur relatively infrequently in the UC³ Corpus.

As observed in Chapters 7–8, structurally simple syllable onsets are produced as consonant clusters in the UC³ Corpus. The sequences of **Place of Articulation** observed in these overly complex productions are presented in Table 9.8. It is observed that the most frequently produced **Place of Articulation** sequence for velar plosives is the adult-like sequence of the cluster word /skul/, i.e. Alveolar+Velar. Productions with **Place of Articulation** sequences matching those of the cluster target words also occur in productions of [d], but not for [t] or any of the labial plosives. Thus, not all overly complex productions may be argued to be the result of perceptual confusion with the complex target word in the same quadruple.

Instead, a number of alternative forms occur in the data set. Overly complex productions of target words with an initial [p] and [k] also occur as a metathesis applied to the cluster target word in the same quadruple. 25% of the productions of [k] are produced as Velar+Alveolar sequences and 7% of the [p] productions are produced as Labial+Alveolar sequences. Thus, full reversals of **Place of Articulation** sequences occur relatively more frequently in overly complex productions of [k] and [p] compared to productions of [sk] and [sp] target clusters.

As seen in Figures 9.8–9.11, the **Place of Articulation** in structurally simple onsets that stabilises first as adult-like is the plosive. The summary of ages of achievement of specific milestones in progressive relative frequency (Table 9.9), show that the delay in achievement in **Place of Articulation** may be up to 32 months, depending on milestone investigated; fricative **Place of Articulation** is acquired much later than plosive **Place of Articulation**.

The late acquisition of fricative **Place of Articulation** indicates that the stopping rule is not the only possible cause of the late onset of adult-like productions for the fricatives. Two further explanations are possible: One, it may be argued that the easier articulatory parameters of plosives compared to fricatives are not the only factors in this substitution. Instead, the frequent substitution of the alveolar fricative for a retroflex may be taken as an indication of a lack of detail in the underlying specification of the perceived fricative. Thus, the child may not be noticing the linguistic significance of the acoustic differences between the output form produced by the child and the model form of the adult speaker. Two, an interaction of two processes of articulation produce the observed pattern of productions in fricatives, 1) the fricative is
produced as a plosive due to the simpler production pattern of complete closure rather than near closure of articulators, and 2) the tongue is moved to a position that is less extended from a neutral mid-cavity position, producing a sound perceived as a retroflex by the adult speaker.

In the absence of articulatory investigations of these productions, it is not possible to know which of these alternative explanations may best help explain the observed data. However, retroflex productions occur in a variety of contexts in the productions in the UC³ Corpus, both before (Tables 9.6–9.8) and after (Table 9.8) a more fronted consonant. If the child was aware of the distinction between retroflex and alveolar fricatives at the time of production, it would be difficult to propose that this distinction was dropped due to articulatory ease, while the more fronted Place of Articulation of the preceding or following consonant is realised. Thus, it is proposed that the child does not yet have a good grasp of the retroflex–alveolar distinction. The productions at a more retracted Place of Articulation may occur by chance when the alveolar Place of Articulation is not reached, but only approximated.

With time, the Place of Articulation sequence of the entire consonant cluster will reach an adult-like manifestation. The age at which each milestone in progressive relative frequency is reached is presented in Table 9.9. Two out of the four clusters do not reach a level above 80% in progressive relative frequency within the time frame investigated in this Thesis. The remaining two target words reached the 90% and 100% milestones after the age of 48 months. However, the possibility of a truncation effect introduced at the edge of the investigated frame of development should also be taken into account. It is therefore possible that the achievement of a 100% progressive relative frequency of adult-like Place of Articulation for the entire consonant cluster may fall well into the fifth year of life.

At least two factors suggest that stable production of Place of Articulation should be delayed relative to Manner of Articulation. One, Manner of Articulation was shown to be more perceptually robust by Singh (1971) indicating that the acoustic cues to Manner of Articulation differences may be more salient to the child than Place of Articulation. Two, there are more phonetic levels of Place of Articulation available to the child and the adult observer than Manner of Articulation. Hence, the chance of making an adult-like production for which positive feedback is given to the child is greater for Manner of Articulation than Place of Articulation.

In the UC³ Corpus as expected, stable and adult-like productions are achieved at the same time or later for Place of Articulation than for Manner of Articulation. As shown in Table 9.10, the delay between achievement of Place of Articulation relative to Manner of Articulation at a specific progressive relative frequency can be substantial. For two target words, /bɔr/ and /pɔr/, the achievement of milestones in Place of Articulation precede that of Manner of Articulation. However, for both of these target words, the progressive relative frequency lay at a level above 90% adult-like Place of Articulation at the onset of the study. In contrast, manner of articulation for /pɔr/ was at 89.8% at 18 months and remained just below 90% until 26 months. Thus, the delay in manner relative to Place of Articulation for /pɔr/ may be an artifact of the milestones defined or the truncation of the developmental sequence performed by choosing a specific age range to investigate. The earlier achievement of 100% adult-like production of Place of Articulation
relative to MANNER OF ARTICULATION for /bɔɾ/ is considered to be the only example deviating from the manner–place order of acquisition.
Chapter 10

Acoustic investigation of refinement in place of articulation

The perceptually based presentation of development of Place of Articulation, presented in Chapter 9, was made in terms of manifestations of structurally reduced syllable onsets, overly complex onsets and the increase in stability in adult-like production of place. In this chapter, the place quality of the production is investigated in terms of developmental trends in acoustic correlates of place of articulation for plosives and fricatives. The place correlates are contrasted both across target forms and the perceived places of articulation of the produced consonants.

The investigation into children’s manifestations of the acoustic features of Place of Articulation is divided into three sections. In Section 10.1 acoustic correlates of Place of Articulation for plosives are investigated. The correlates considered include the measurement of spectral moments of the release burst, and the spectral tilt at release compared to at the point of voicing. Section 10.2 then investigates the gradual refinement of fricative place, using acoustic measurements of spectral moments of the fricative spectrum, $F_2$ onset frequency and the degree of co-articulation with $F_2$ in the middle of the following vowel. The results are summarised and discussed in Section 10.3.

10.1 Acoustic correlates of place of articulation in plosives

A number of proposals have been made concerning the acoustic correlates of Place of Articulation in plosives. Stevens & Blumstein (1978) proposed that the release burst contained enough information for a classification. Velar plosives have a pointed release spectrum with a centrally located peak. Viewing the spectrum of the release as a statistical distribution of sample points, development towards adult-like production is proposed to be quantifiable by measurements of spectrum variance and kurtosis (Section 10.1.1).
Productions of adult-like velar plosives should display mid-range median values, minimal skewness, small variance and a strong kurtosis. In contrast, alveolar and labial plosives were argued by Stevens & Blumstein (1978) to have a more wide-spread, diffuse, spectrum compared to velar plosives. Alveolar plosives were described by Stevens & Blumstein (1978) as having a rising spectrum with a high-frequency peak and labial plosives were described as having a peak at a low frequency and with a fall in amplitude with increasing spectral frequency. Thus, a plausible quantification of the spectral distinction between labial and dental plosives should be measurements of release burst skewness values (Section 10.1.2). Adult-like productions of alveolar is expected to have a low skewness value. Labial plosives are expected to show a high skewness value.

Alternatively, the child may produce a distinction between labial and alveolar productions using more dynamic acoustic cues, such as the change in spectral tilt from release to voicing onset suggested by Kewley-Port (1983) and Kewley-Port et al. (1983) and quantified by Lahiri et al. (1984). Children are expected to refine their productions towards more frequent adult-like classifications in terms of Lahiri et al.’s (1984) quantification of the spectral tilt change, within the plosive (Equation A.3 on page 183). This acoustic correlate will be put in a developmental perspective in Section 10.1.3.

10.1.1 Spectral diffuseness

Stevens & Blumstein (1978) proposed that the plosives produced in the front of the mouth, alveolar and labial, could be characterised as having a more diffuse spectrum than velar plosives. Viewing the spectrum as a probability distribution, the spectrum of front plosives could be described as having a large variance from the spectral mean. Velar plosives however, not being characterised as having a diffuse spectrum, would be expected to show a relatively small spectral variance in the release of the plosive.

From a developmental perspective, Grunwell (1987) showed that velar plosives may be produced at a more fronted place of articulation in the early stages of development. This finding was confirmed for the UC Corpus data in Chapter 9. Thus, development towards an increasingly adult-like production of velar targets could be described by the decrease in spectral variance in the burst spectrum with age.

In order to investigate the change of variance as a developing cue to PLACE OF ARTICULATION, the burst spectrum variance is presented as a box-and-whisker plot for each AGE in Figure 10.1. Each panel shows the acoustic measurements obtained divided by PLACE OF ARTICULATION of the target plosive. A linear regression line, for which the slope and intercept is given in the plot, is fitted to the data set within the panel. In addition, a dashed LOESS line (Cleveland 1979) is fitted to the dataset with the span of 1. The median of each AGE is indicated by a black dot in the centre of the box. The top and bottom of the box (the “hinges”) indicate the first (bottom line) and third (upper line) quantile. The whiskers extend 1.5 times the length of the corresponding hinges. Data points outside the whiskers are viewed as outliers and are indicated by unfilled circles.

The variance measures presented in Figure 10.1 show no clear sign of development towards a more adult-like production. Slope values are, contrary to expectation, smallest in for productions towards velar targets. A similar observation may be made in Figure 10.2, showing the data set divided by the perceived place quality of the plosive. The slope of the regression line shows that velar productions are not progressing to-
Figure 10.1: Variance of the release spectrum against Age of the participant. The data set is divided by Place of Articulation of the target plosive.

Figure 10.2: Variance of the release spectrum against Age of the participant. The data set is divided by Place of Articulation transcribed by the author.
Towards a less diffuse spectrum and that dental plosives have the lowest intercept and slope values. Thus, the measurements of spectrum *diffuseness*, quantified as spectrum variance, do not show a developmental trend. This could reflect a lack of a group developmental trend for spectral *diffuseness* in the children of the UC\textsuperscript{3} Corpus or be because spectral variance does not capture the change in distribution wideness that occurs.

Therefore, an alternative quantification of spectral diffuseness, distribution kurtosis, was investigated for developmental trends. Figure 10.3 shows the kurtosis of productions in the form of a box-and-whisker plot for each age, divided into panels by the **Place of Articulation** of the target plosive. Figure 10.4 shows the same data divided into panels by **Place of Articulation** perceived by the adult transcriber.

The regression lines in the panels of Figures 10.3–10.4 do not show a development towards a more adult-like separation between the plosives in terms of release burst spectrum diffuseness. In the case of a development towards an increasingly adult-like separation of the place contrasts, a negative slope would be expected in velar plosives, decreasing the diffuseness of the spectrum. No such trend is observed. The left-most panels of Figures 10.3 and 10.4 both show a non-negative slope and no distinction from the competing plosive places of articulation in terms of intercept. The dataset, therefore, does not indicate an overall decrease in release burst spectrum with **Age** for attempted or realised velar plosives.

However, from the LOESS line in the panels, it can be observed that a change in progression direction may be occurring at approximately three years of age. At that time, for the velar plosives, the LOESS line stops following the linear regression line of the entire data set displayed in the panel and shows a decreasing trend. This pattern of development is observed both in the division of the data set by place of articulation of the target plosive (Figure 10.3) and by perceived place of articulation (Figure 10.3).

Thus, productions of dental and labial plosive are becoming more pointed with **Age** and velar productions show a rapid change of production pattern towards an increasingly flat spectrum after the age of 38 months. This pattern of development is the opposite of what would be expected if the productions were developing towards the adult model.

### 10.1.2 Spectral skewness

The diffuseness of the release burst spectrum and kurtosis were shown in the previous section not to provide evidence for a refinement towards a separation between velar and front plosives with age. However, Stevens & Blumstein (1978) proposed differences in placement of distribution centre; labial plosives should develop towards a positive skewness, velar towards no skew, and dental plosives towards a negative skewness.

The progression in release burst skewness was therefore investigated using a comparative analysis of productions made towards a specific **Place of Articulation**. Figure 10.5 shows box-and-whisker plots of release burst skewness for each age investigated, divided into panels by **Place of Articulation** of the target plosive. A linear regression line, for which the slope and intercept is given in the graph, is superposed on the box-and-whisker display of the data. In addition, a LOESS line shows the local linear trend in the cell.
Figure 10.3: Kurtosis of the release spectrum against Age of the participant. The data set is divided by Place of Articulation of the target plosive.

Figure 10.4: Kurtosis of the release spectrum against Age of the participant. The data set is divided by Place of Articulation transcribed by the author.
The results indicate a developmental trend towards more adult-like productions in terms of release burst skewness. Productions towards dental plosives have the smallest intercept value for the regression line and reach the lowest skewness centre of the alternative places of articulations. Thus, productions towards dental plosives progress towards the *rising* spectral shape described by Stevens & Blumstein (1978).

As expected labial productions in terms of release burst shape did not change as rapidly as velar plosives in the progression towards the adult model. Thus, in the last ages investigated, attempted productions of labial plosives have the highest skewness values, indicating a *falling* spectral release burst shape, attempted velar productions have skewness distribution centred around zero or with a slightly positive skewness, indicating a more centralised release burst centre, and attempted dental productions have the lowest skewness values, indicating a *rising* distribution with spectral frequency. Thus, the acoustical separation in terms of burst spectrum skewness increases with age.

A mixed effect model of skewness by *Age* of the participant and *Place of Articulation* of the target plosive (with participant as a fixed effect) showed a significant main effect of *Age* ($F_{(1,3672)}=11.5, p<0.01$) and *Place of Articulation* ($F_{(2,3672)}=51.24, p<0.001$). Group-wise comparisons of *Place of Articulation* showed a significant difference between all target places of articulations ($[F_{(1,2846)}=7.3, p<0.01]$ for the velar–labial contrast, $[F_{(1,2846)}=96.0, p<0.001]$ for the dental–labial contrast and $[F_{(1,2846)}=23.1, p<0.001]$ for the dental–velar contrast).

The results can be interpreted in two ways: One, the gradual separation in skewness is indicative of a refinement in the acoustic characteristics of the production, leading to the increase in productions perceived as adult-like (as seen in Figures 9.8–9.11) and Two, that the plosives produced are acoustically specified in an adult-like manner when produced, but that the trend observed in Figure 10.5 illustrates that the number of instances when the child realises which plosive is to be produced increases with age. In order to tease these two possibilities apart, an analysis of the productions grouped by the perceived place quality was conducted.

The results obtained from the division of the productions by perceived *Place of Articulation* (Figure 10.6) are similar to the data set divided by *Place of Articulation* of the target (Figure 10.3). The separation obtained with development accords with what would be expected from a progression towards an increasingly adult production. Dental plosives develop towards a small skewness value, labial plosives show the highest intercept value and a skewness at the highest relative range, and velar plosives are produced with a skewness that lies between dental and labial production skewness throughout the investigated *Age* range.

There is therefore evidence for a developmental trend in skewness of the release burst spectrum, both when investigating trends by the target of the production and the perceived place quality of the plosive produced. Release burst spectrum skewness can be viewed as an acoustic cue to *Place of Articulation* which matures over the investigated *Age* interval.

### 10.1.3 Change in spectral tilt

The cue of spectral tilt change has been shown in both visual classification (Kewley-Port 1983), perceptual (Kewley-Port et al. 1983) and production experiments (Lahiri et al. 1984) to be a significant acoustic cue to plosive *Place of Articulation*. Lahiri
Acoustic cues of place

Figure 10.5: Skewness of the release spectrum against Age of the participant. The data set is divided by Place of Articulation of the target plosive.

Figure 10.6: Skewness of the release spectrum against Age of the participant. The data set is divided by Place of Articulation transcribed by the author.
et al.

Acoustic cues of place

et al. proposed that the quotient between the spectral slope at the release burst and the corresponding slope at the onset of phonation (See Equation A.3) could serve as a basis for classification of plosives characterised by Stevens & Blumstein (1978) as diffuse. According to Lahiri et al.’s (1984) classification, dental and alveolar plosives are characterised by a $q < 0.5$ or a $(c - a) < 0$, and labial plosives by $q > 0.5$.

If there is a developmental progression in the spectral tilt change during the plosive in the productions made by children, this progression could be quantified in terms of the $q$ quotient; for a developmental effect in terms of $q$ to be present, an increase in relative frequency of adult-like classification by the $q$ quotient is expected to occur.

In order to investigate the developmental trends in slope dynamics within the plosives produced, the relative frequency of classifications into alveolar/dental or labial was investigated. Figure 10.7 shows the frequency of production that was classified as alveolar/dental or labial by the Lahiri et al.’s (1984) quantification, divided by Age and place of articulation of the attempted target plosive. Figure 10.8 shows the same data set divided by the Place of Articulation perceived by the adult transcriber. In Figure 10.8, only plosives perceived as either dental, alveolar or labial are included.

A clear preference is found in Figure 10.7 towards classification into the alveolar/dental category for both target plosive places of articulation. If a change in the within-plosive spectrum dynamics is a part of the developmental strategy of children, a progression towards an increased classification the labial category of productions made towards a labial plosive target would be expected. Figure 10.7, however, does not show such a developmental trend.

On the other hand, a developmental trend is observed in Figure 10.8. The relative frequency of occurrence of classification into the labial category increases with age, reaching a production rate of >50% after the age of 44 months. Thus, a gradual maturation of the produced plosives is observed within the grouping by perception that is not observed in the grouping by target plosive.

10.2 Acoustic correlates of Place of Articulation in fricatives

The reviewed of the literature on acoustic cues for Place of Articulation in fricatives presented in Section 2.3.4 showed a strong influence of spectral moments of the fricative on the perception of place of articulation (Forrest et al. 1988, Jongman et al. 2000). $F_2$ onset frequency and degree of co-articulation with the following vowel were also shown to provide Place of Articulation information (Jongman et al. 2000, Soli 1981). Nittrouer & Studdert-Kennedy (1987) and Nittrouer & Miller (1997a, 1997b) observed a shift in focus from co-articulation cues towards spectral cues was in the children they investigated. Measurements of spectral mean, spectral variance, skewness, kurtosis and $F_2$ onset frequency of the fricatives found in the UC3 Corpus may therefore reveal developmental trends.

10.2.1 Spectral mean

The research conducted on [s] productions in terms of spectral mean values has indicated a progression from the 8 kHz range in children’s speech to spectral means between
Figure 10.7: Categorisation of plosives by the spectral tilt change in from the release to the onset of voicing. Each point in the graph represent the relative frequency of categorisations by the \( q \) quantity, divided by the Place of Articulation of the target plosive.

Figure 10.8: Categorisation of plosives by the spectral tilt change in from the release to the onset of voicing. Each point in the graph represent the relative frequency of categorisations by the \( q \) quantity, divided by perceived Place of Articulation of the dental or plosive. Only dental and labial productions are shown since the \( q \) was designed for this distinction.
Acoustic cues of place

8–6.5 kHz for 7-year-olds and adults. The mean spectral frequency of the fricatives produced in the UC3 Corpus are displayed in Figure 10.9. In contrast to Nissen & Fox (2005), the frequency values were not converted to ERB, since the observed frequencies frequently exceeded 6500 Hz and “... the expression should only be considered valid for values of \( f \) between 0.1 and 6.5 kHz” (Moore & Glasberg 1983, p 752).

Figure 10.9 shows a slowly decreasing trend in terms of spectral mean values that does not agree with the progression found by Nittrouer (1995) and Nissen & Fox (2005) when investigating children’s production of [\( \mathrm{f} \)]. A refinement towards an increasingly adult-like production pattern could, however, be observed in the division of the data by perceived place of articulation (Figure 10.10), productions perceived as alveolar display an increasing trend in mean spectral frequency, in agreement with the results of investigations of children acquiring English (Nittrouer 1995, Nissen & Fox 2005). Thus, a gradual refinement within the achieved perceptual place category is observed.

The statistical significance of the developmental trends presented in Figures 10.9 and 10.10 were tested using a mixed effects model of spectral mean frequency by \textit{Age} and \textit{Place of Articulation} of the produced plosive. \textit{Participant} was used as a fixed effect. The results showed a significant main effect of \textit{Place of Articulation} \((F(2,3042)=69.2,p<0.001)\) but no significant effect of \textit{Age} \((F(1,3042)=0.06,p=0.81,\text{n.s.})\) across all places produces. A test of the \textit{Age} effect within the cell “alveolar” cell only also produced a non-significant result \((F(1,684)=0.9,p=0.35,\text{n.s.})\). Thus, the measurement of spectral mean failed to provide a statistically significant trend of development in the direction of the adult target plosive.

10.2.2 Variance

Nissen & Fox (2005) showed that spectral variance was able to separate the four places of articulation for the fricatives, and that [\( \mathrm{f} \)] productions had the lowest variance values both in children’s and adult target production. The data also suggest a decreasing trend in spectral variance with \textit{Age} of the participant.

Figure 10.11 shows the spectral variance of the fricatives produced in the syllable onset in the form of a box-and-whisker plot for each month investigated. A linear regression line, for which the slope and intercept are indicated in the graph, is superposed to the plot. In addition, a LOESS line (calculated with span=1) is shown as a dashed line. The almost identical path of the LOESS line and the regression line however indicates a strong linearity in the measurements. The positive slope indicated for the regression line is viewed as good description of the progression in the full data set.

A positive slope is observed across all fricatives produced in the syllable onset. Thus, Figure 10.11 gives no indication of fricatives produced in the syllable onset progressing towards [\( \mathrm{f} \)]. The spectral variance in fricatives produced in the UC3 Corpus therefore does not indicate a progression towards the an adult fricative target.

A decreasing trend was not observed in the measurements of spectral variance in fricatives divided by perceived \textit{Place of Articulation}. Figure 10.12 shows the variance measurements divided into separate panels by perceived place for the three categories most frequently occurring in the UC3 Corpus. A positive slope was found for all perceived places of articulation. However, the deviating slope of the LOESS line in the labial panel indicates that the linear regression line may not be an accurate description of the true trend in the data set.
Figure 10.9: Spectral mean of the fricative spectrum against Age of the participant. The data set is divided by Place of Articulation of the target plosive.

Figure 10.10: Spectral mean of the fricative spectrum against Age of the participant. The data set is divided by Place of Articulation of the transcribed plosive.
Figure 10.11: Variance of the fricative spectrum against Age of the participant.

Figure 10.12: Spectral variance of the fricative spectrum against Age of the participant. The data set is divided by Place of Articulation of the transcribed plosive.
The observed trends were tested for significance using a mixed-effects model with \textsc{Age} and \textsc{Place of Articulation} as independent variables, spectral variance as a dependent variable and \textsc{Participant} as a fixed effect. The results showed a significant main effect of both \textsc{Age} ($F_{(1,3042)}=20.1, p<0.001$) and \textsc{Place of Articulation} ($F_{(1,3042)}=4.5, p<0.05$). Thus, spectral variance reveals a significant developmental trend that is the opposite to that observed in the data presented by Nissen & Fox (2005) in terms of direction.

### 10.2.3 Skewness

Previous research conducted on fricative spectrum skewness has indicated a progression towards an increasing skewness value with \textsc{Age} for productions of $[s]$ (Nittrouer 1995, Nissen & Fox 2005). There is, however, considerable variation in the reported skewness values for $[s]$ productions: Nittrouer (1995) reported mean skewness values of 0.46 for children at the \textsc{Age} of three years, 0.22 at the age of five and 0.65 for adults. Nissen & Fox (2005) reported mean skewness values of approximately -2.3 for three-year children and -1.7 for adults. It seems that the direction of the development is the determining factor.

The spectral skewness of the fricatives produced in the syllable onset in the UC3 Corpus are shown in Figure 10.13 in the form of a box-and-whiskers plot for each month. A linear regression line indicates the linear trend of the entire data set and a (dashed) LOESS line indicates the local trend of the data set. The slope and intercept of the linear regression line are indicated at the top of the plot.

The data presented in Figure 10.13 show a similar progression to that observed between \textsc{Age} three and five by Nittrouer (1995), although the slope of the linear trend is steeper than in Nittrouer (1995). This steeper slope results in skewness values by the \textsc{Age} of 5:0 that are in the same approximate range as the mean skewness values for adult speech reported by Jongman et al. (2000). The progression, however, does not resemble the data of Nissen & Fox (2005), who reported an increasing trend.

A progression similar to the one reported by Nissen & Fox (2005) is, however, observed for the skewness measurements divided by perceived \textsc{Place of Articulation} (Figure 10.14). The spectral skewness of alveolar productions was found to be closer to Nissen & Fox’s (2005) findings than the Nittrouer’s (1995) findings and show an after 36 months. Thus, the developmental trend in the data set approaches the mean values of the adult productions (Jongman et al. 2000).

The \textsc{Age} effects observable in Figures 10.13 and 10.14 were shown to be not significant by a mixed effect model testing the effect on skewness measurements obtained ($F_{(1,769)}=0.86, p=0.80, n.s.$). The place differences were, however, shown to be significant ($F_{(2,769)}=0.86, p<0.001$). However, as a whole the developmental data of fricative spectrum skewness failed to provide clear evidence of a developmental trend towards an adult model.

### 10.2.4 Kurtosis

There are two scales (kurtosis proper and kurtosis excess) on which the kurtosis of a distribution may be indicated (See Abramowitz & Stegun, 1972) These differ only by the point at which the distribution has a kurtosis similar to that of the normal
Acoustic cues of place

Figure 10.13: Skewness of the fricative spectrum against Age of the participant. The data set is divided by Place of Articulation of the target plosive.

Figure 10.14: Spectral skewness of the fricative spectrum against Age of the participant. The data set is divided by Place of Articulation of the transcribed plosive.
(Gaussian) distribution. On the scale of kurtosis proper, the normal distribution has a kurtosis of 3; on the scale of kurtosis excess, the normal distribution is at zero.

For adult productions, Jongman et al. (2000) reported a mean kurtosis for alveolar fricatives of 2.36 and Nissen & Fox (2005) reported mean adult skewness values of approximately 3. Neither group of researchers indicated which kurtosis scale was used in the calculations, nor on which scale their values should be interpreted.

However, regardless of the scale used, Nissen & Fox (2005) provided evidence of a decreasing developmental trend in spectral kurtosis in alveolar fricatives from three years. They reported a mean kurtosis value of approximately 4.3 at three years that decreased linearly to approximately 3.6 in the productions of their five-year-old speakers.

For the UC3 Corpus, the kurtosis values calculated from the spectrum of fricatives produced in the syllable onsets are presented in Figure 10.15 in the form of box-and-whiskers plots of kurtosis proper for each age investigated. Similarly to the developmental data presented by Nissen & Fox (2005), the data shows a decreasing trend with age. However, the kurtosis values found and presented in Figure 10.15 are somewhat lower ($\bar{x} = 3.22$ at the age of 36 months) than those of Nissen & Fox (2005). A similar progression is observed in the UC3 Corpus data set when divided by perceived Place of Articulation (Figure 10.16). Productions perceived as alveolar progress in a (near) linear manner from 3.8 at the age of 18 months to 3.1 at the age of 50 months.

As for the other spectral moments, fricative spectral kurtosis was investigated for significant effect of Age of the participant and perceived Place of Articulation using a mixed effects model. Participant was used as a fixed factor in order to account for the fact that the data consisted of repeated measurements on fricatives produced by the same speaker. The results showed a significant main effect of perceived Place of Articulation ($F_{(2,769)}=4.7, p<0.05$), but not for Age ($F_{(1,769)}=1.13, p=0.28, \text{n.s.}$). These results are similar to the results from Nittrouer (1995) and Nissen & Fox (2005) in regard to the significant effect of Place of Articulation and Nissen & Fox (2005) in regard the non-significant effect of Age.

10.2.5 Onset frequency of F<sub>2</sub>

The spectral properties of the fricatives have been shown to decrease in perceptual importance as the child grows older. Nittrouer et al. (1989) reported that “The extent to which speakers differentiated between /F/ and /S/ increased with age, while the extent to which they co-articulated each fricative with its following vowel decreased.” (p 120). Two questions immediately arise: 1) do children learning Swedish also show this shift of attention from the F<sub>2</sub> towards spectral cues with age, and 2) do the children’s usage of the F<sub>2</sub> onset frequency cue give additional information regarding the specificity of Place of Articulation in the phonological system of the children.

Three hypotheses are proposed in connection to level of detail acquired in the phonological system. One, if the fricatives produced by children are intended productions of alveolar fricatives that miss the acoustic mark and are therefore miss-perceived as fricatives produced at other places of articulation, a change in usage of F<sub>2</sub> onset frequency should be observed. With this change, the children should be trying to find a production pattern that more accurately fits the adult model of production and results in more positive feedback.
Figure 10.15: Kurtosis of the fricative spectrum against Age of the participant.

Figure 10.16: Spectral kurtosis of the fricative spectrum against Age of the participant. The data set is divided by Place of Articulation of the transcribed plosive.
Two, if a contrast between output forms produced and the adult target form is not present, a developmental change in the F$_2$ onset frequency used is instead expected to occur within a produced place category. No change would indicate no linguistic use of the cue.

Three, if a shift from usage of co-articulation towards spectral properties occurs as a part of speech development in Swedish children, the F$_2$ onset frequency used in the fricative should show less influence of the F$_2$ frequency observed in the vowel. Thus, the absolute difference between the frequency of F$_2$ in the fricative and at the middle of the following vowel should increase with age until a stable pattern of production is achieved. A decrease in absolute difference in F$_2$ within the fricative compared to the vowel would indicate increased co-articulation with age, which would disagree with the results of Nittrouer et al. (1989).

In order to test whether a change in F$_2$ onset frequency usage across all produced fricatives is present in the UC$^3$ Corpus data, measurements of F$_2$ onset frequencies from all pre-vocalic fricatives were extracted. These are presented in Figure 10.17. It can be observed that the linear regression line fitted to the data set is close to horizontal. The slope value indicated at the top of Figure 10.17 shows a minimal deviation from zero, indicating no developmental effect. The same is observed for the dashed LOESS line, which follows the linear regression line closely and, therefore, indicates that no change in production behaviour is present across the entire data set of pre-vocalic fricatives.

The visual impression obtained from Figure 10.17 was confirmed using an ANOVA performed on a mixed-effects model of age effects on F$_2$ onset frequency in the set of pre-vocalic fricatives. The results showed that age did not significantly affect the use of F$_2$ onset frequency ($F_{(1,2913)}=0.22, p=0.64, \text{n.s.}$). Thus, no overall developmental effect on the use of F$_2$ onset frequency was found in the data set.

In order to investigate whether a change in usage of F$_2$ is present within a category upheld by combined effort of acoustic cues, the data set of F$_2$ measurements was divided by place of articulation perceived by the adult observer. Figure 10.18 shows a decreasing trend in F$_2$ onset frequency in fricatives perceived as alveolar. In addition, the dashed LOESS line indicates an increased negative slope after the age of 38 months. The fricative productions perceived as alveolar therefore supports the hypothesis that a change in production behaviour is manifested in F$_2$ onset frequency. The age trend observed, however, failed to reach a statistical significant level ($F_{(1,2913)}=0.22, p=0.64, \text{n.s.}$).

In order to investigate the development of the use of co-articulation, the difference between F$_2$ onset frequency and the F$_2$ frequency in the middle of the vowel was calculated (See Figure 10.19). In order for a decrease in the use of co-articulation to be considered to have occurred, a decrease in the difference between F$_2$ at onset and in mid-vowel is expected. The results presented in Figure 10.19 do not follow this trend. The linear regression line indicates no sign of increase in F$_2$ distance. In addition, the dashed LOESS line does not deviate from the linear trend at any point. The effect of age on F$_2$ distance was shown by statistical testing using a mixed-effects model of F$_2$ distance by age of the participant, and with participant as a fixed effect to non-significant ($F_{(1,656)}=0.22, p=0.64, \text{n.s.}$). Thus, no change in co-articulation usage was observed for pre-vocalic fricatives produced in the UC$^3$ Corpus.
<table>
<thead>
<tr>
<th>Age (months)</th>
<th>Fricative F₂ onset frequency (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>1800</td>
</tr>
<tr>
<td>22</td>
<td>2000</td>
</tr>
<tr>
<td>26</td>
<td>2200</td>
</tr>
<tr>
<td>30</td>
<td>2400</td>
</tr>
<tr>
<td>34</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>

Slope = 0.184008
Intercept = 2134.490183

Figure 10.17: F₂ onset frequency in pre-vocalic fricatives against AGE of the participant.

<table>
<thead>
<tr>
<th>Age (months)</th>
<th>Fricative F₂ onset frequency (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>1800</td>
</tr>
<tr>
<td>22</td>
<td>2000</td>
</tr>
<tr>
<td>26</td>
<td>2200</td>
</tr>
<tr>
<td>30</td>
<td>2400</td>
</tr>
<tr>
<td>34</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>

Slope = 1.699564
Intercept = 2193.550824

Figure 10.18: F₂ onset frequency in pre-vocalic fricatives perceived as alveolar against AGE of the participant.
10.3 Discussion

According to the presentation of Place of Articulation development in Chapter 9, the adult observer may perceive the attempted plosives and fricatives as highly variable in terms of where the consonants are produced. However, as it is possible that the observed variance in place is an artifact of the categorizations that are based on perceptual judgements made by a single observer acoustic correlates of Place of Articulation have been investigated as possible indicators of refinement in Place of Articulation.

In order to both provide a description of the productions manifested and to acquire insight into nature of the phonological awareness present in the child, a general methodology of categorisation of the results by both target Place of Articulation and perceived Place of Articulation was employed. Thus, in the case a progression may be observed in the categorisation by target place towards an adult target value, children are viewed to have perceived the target correctly but are unable to produce it. In this case, then, the variance observed is caused by children’s failure to meet the demands of adult perception. Thus, the imperfect productions are viewed as caused by articulatory factor and the variance observed in Chapter 9 is viewed as an artifact of the adult perceptual system.

In contrast, if a development is observed in the acoustic measurements divided by perceived Place of Articulation, children are argued to show signs of having attempted productions at more than one place of articulation. In turn, this suggests that the child may have an imperfect phonological representation of the target, possibly caused by perceptual processes. Thus, the consonants produced may be closer to an adult output form of the produced consonant than the adult output form of the
Acoustic cues of place

A consonant present in the target word elicited. The variance observed in Chapter 9 is in this case viewed as representative of the child’s intended production.

Of the four acoustic cues of Place of Articulation for plosives proposed in the literature, only spectral skewness and the spectral tilt change cue were shown to display a developmental effect towards the adult target. Skewness was shown to develop towards an increased separation in distribution skewness that agreed with what was predicted from data gathered from adult speakers of English. At the end of the investigated Age interval, dental plosives most strongly agreed with a diffuse-rising spectral description, labial plosives agreed best with the diffuse-falling description and velar plosive skewness showed the most centred spectral distribution.

As this developmental trend was observed both in the data set when divided by target plosive place (Figure 10.5) and by perceived place (Figure 10.6), it is argued that this effect is caused by the overlap between two categorisations of the data points. As seen in Figures 9.8–9.11, adult-like production of Place of Articulation is achieved at a high level early in the productions of children. This observation is interpreted as an indication that enough separation in terms of release spectrum skewness is achieved earlier than the earliest ages investigated in this Thesis. Thus, place differences may in many cases is perceived using only the minimal separation displayed by the small differences displayed in Figures 10.5–10.6 at the age of 18 months. The developmental trends observed in Figures 10.5–10.6 are thus viewed as within-category refinement in order to take care of ambiguous cases.

The cue of spectral tilt change from the release burst to the vowel proposed by Lahiri et al. (1984) also provided evidence of a developmental trend. This trend was observed only in the division of the data set by Place of Articulation perceived by the adult observer. This developmental trend was not observed in the division of data by target Place of Articulation. Thus, it is proposed that the cue of spectral tilt indicates within-category adjustment of the cue rather than an attempt to achieve an increased adult target likeness.

However, it should be noted that Lahiri et al.’s (1984) spectral tilt quotient was partially successful in separating labial plosives from dental plosives. Productions of labial plosives were classified as labial more often than productions of dental targets. The quotient was found to be more biased towards a dental/alveolar classification of attempted labial plosive productions than the adult observer. Hence, this cue may not be suitable for investigations of speech development.

For fricatives, three out of the four investigated acoustic cues of Place of Articulation were shown to change in usage with development. The spectral mean cue showed a developmental trend towards increasing adult values. This progression was, however, only observed in the division of the data set by perceived Place of Articulation, indicating a within-category refinement of the produced Place of Articulation. The lack of development towards the adult model found in the division by place of the target indicates that children do not use the cue of spectral skewness in order to push the output form closer to that of the target; thus, children may not be aware of the difference between how the target form and the output form produced are perceived by the adult observer.

Spectral skewness also changed in usage with development. For this feature, two developmental directions were observed. When dividing the data set by Place of Articulation of the target plosive, the results showed a downward trend in skewness
Acoustic cues of place

155

with age (Figure 10.13). This trend parallels that observed by Nittrouer (1995) for children acquiring English.

However, the opposite trend, a positive slope, is observed in fricatives perceived as alveolars (Figure 10.14). This slope contrasts with the trend observed by Nittrouer (1995), but agrees with the trend observed by Nissen & Fox (2005). Due to this diversity in developmental trends, the data skewness measurements obtained from productions within the UC3 Corpus do not afford an interpretation in terms of the driving force behind development and no general conclusion can be drawn.

In addition to the cues of spectral mean and skewness, F2 onset frequency was shown to be used in the development of Place of Articulation for pre-vocalic fricatives. It was argued by Nittrouer et al. (1989) that F2 values produced in the latter part of the fricative decreased with age. This effect is not observed in the UC3 Corpus when dividing the data by Place of Articulation of the target plosive (Figure 10.17), but is observed in the linear regression line fitted to the set of F2 onset frequencies divided by perceived Place of Articulation (Figure 10.18). Thus, the data set of F2 onset frequency measurements gathered from the UC3 Corpus agrees with the data from previous investigations, when divided by perceived Place of Articulation. The data, therefore, indicate a usage of F2 within produced categories rather than as a way to increase adult model.

Nittrouer et al. (1989) proposed a shift in attention from focus on cues of co-articulation (F2) to spectral based cues do occur as a part of development. In the UC3 Corpus, a change in direction of the LOESS line in Figure 10.18 at 38 months may be taken as support of a change in usage of the F2 cue in favour of spectral cues. However, Nittrouer et al. (1989) suggested that this change in behaviour was due to a change in focus from cues of co-articulation to spectral cues more bound to a segment. In this case, the F2 in the fricative should be closer to that of the F2 frequency in the middle of the vowel in early productions and be further from it in the later productions.

The results from the investigation of F2 frequency differences between that of the fricative and in the middle of the vowel, however, did not lend support to the hypothesis of decreased co-articulation with age. Hence, in the absence of an increase in F2 frequency separation, it is concluded that the effect of F2 frequency may not be seen as an effect of formant frequency change, but rather as a static spectral cue to the Place of Articulation, even for children.
Acoustic cues of place
Part V

Synthesis of the acquisition
Chapter 11

Synthesis of the findings

This Thesis has provided a description of the advances made by children in their manifestations of place, manner and voicing/aspiration as they progress towards increasingly adult-like s+plosive clusters. The results are discussed in terms of a developmental sequence for place, manner, voicing and structural complexity. Acoustic cues showing progression towards a new contrast of an enhancement of already acquired contrasts and periods in development, where confusion is observed, are summarized. Finally, suggestions are made for future research.

11.1 Towards a developmental sequence

The data presented in this Thesis can be interpreted in terms of a developmental of stable productions of Place of Articulation, Manner of Articulation and Onset Complexity.

The perceptual development order found in the UC3 Corpus for Place of Articulation and Manner of Articulation paralleled that of Singh (1971) and Karlsson (2005) perceptual stability a stable production of manner of articulation precedes a stable production of Place of Articulation in single-consonant onsets. This earlier acquisition of Manner of Articulation, compared to Place of Articulation, in singleton onsets holds for both plosives and fricatives. The Manner of Articulation for fricatives is acquired later in than for plosives, causing a smaller delay in Place of Articulation in fricatives compared to Place of Articulation for plosives. This difference decreases with Age (c.f. Tables 8.7 and 9.9).

Structural complexity is acquired in complex target words after manner and Place of Articulation in target words with a single plosive. However, the point at which acquisition of structural complexity in comparison to Place of Articulation and Manner of Articulation is judged to occur is dependent on the milestone selected. For example, complexity is acquired before Manner of Articulation for fricatives until the 70–80% milestone (See: Tables 7.1–8.7), after which manner achieves the more demanding milestones before structural complexity. The acquisition of Place of Articulation and Manner of Articulation merges in productions of complex
targets after complexity has been achieved. As seen in Table 9.10 there is no time
difference in the time of Place of Articulation and Manner of Articulation
for complex targets. Thus, once complexity has been achieved, the delay in Place
of Articulation compared to Manner of Articulation acquisition disappear.
This constrasts with the progression found in singleton onsets, where this merger is
not found.

The relative production sequence described in this Thesis shows a convergence in
Place of Articulation, Manner of Articulation and Onset Complexity at
the time of these features reaching an 80% adult-like production rate in consonant
cluster targets. However, there are two factors that should be kept in mind when inter-
preting the data. One, as the number of analysed productions is not as large at the tails
of the age interval as in the distribution centre, the estimate of the relative frequencies
of adult-like productions are weaker at the distribution tails. Two, the selection of
a specific age interval to investigate imposes restrictions on the interpretation of the
data set. Since recordings stopped at an age at which not all investigated features had
been acquired at the 100% stable production level, the data set should be viewed as
a truncated distribution. The effect of this is that as the age at which the progres-
sive relative production frequency is calculated progresses towards the positive tail of
the distribution, the truncation effect causes an increase in the confidence interval of
estimation of the statistic.

Due to the combined effects of the decrease in production frequency and the trun-
cation effect, the estimation of production rates at the positive tail of the production
distribution is associated with an increasingly wide confidence interval. Since a pro-
gression is made towards an increase in adult-like productions, the estimate is increased
in likelihood of being an overestimate of the true progressive relative frequency. Thus,
the progressive relative frequencies of adult-like productions may be overly positive
estimates of the achievement of a distinction between output forms in production.

Karlsson, Czigler & Sullivan (2005), based on a subset of UC³ Corpus reported that
the acquisition of a voicing and aspiration contrast shows signs of being detached from
the acquisition of Place of Articulation and Manner of Articulation. This
finding was confirmed for the full UC³ Corpus data set. Movement from a unimodal
towards a bimodal distribution of VOT for voiced unaspirated and voiceless aspirated
plosives was observed as early as 24 months (Figure 4.4), and a pattern of frequent
productions of prevoicing was observed in voiced and voiceless unaspirated plosives
from the age of 24 months (Figure 4.4). These observations provide evidence of the
early manifestation of all output forms.

However, an increased production of false positives was observed for VOT; long
lag productions of unaspirated plosives still occurred at the 48–54 months interval
(Figure 4.5). This presence of false positives suggests that the aspiration distinction is
not achieved at the time of onset of long lag productions, and that a 100% production
distinction had not been achieved even in the last age interval investigated. Since
VOT was shown to be the only acoustic cue to voicing that provided evidence of
developmental use, a stable production of a contrast in voicing is viewed as being
acquired at the offset of the age period investigated in this Thesis.
11.2 Acoustic correlates to development

In this Thesis, several acoustic cues of phonetic distinctions have been investigated as potential indicators of development. For the voicing aspiration contrast, the acoustic cues of VOT, RELATIVE ASPIRATION AMPLITUDE and \( f_0 \) onset frequency were investigated for signs of being manipulated to achieve an increased agreement with the model production presented by adults.

The results showed a gradual refinement with age in the VOT cue, but no developmental use of the \( f_0 \) cue. RELATIVE ASPIRATION AMPLITUDE was shown to provide an auxiliary cue to the voicing distinction in ambiguous cases, i.e. when the produced VOT was close to the perceptual boundary between aspirated and unaspirated plosives. Thus, this Thesis has provided evidence of VOT being a primary cue that is used in the development of the contrast between unaspirated and aspirated plosives, and that VOT may be enhanced by the cue of aspiration amplitude.

For the distinction in PLACE OF ARTICULATION, only the spectral skewness of plosive burst spectral showed a developmental trend towards the adult model. This cue is, therefore, seen as being used to increase the production contrast between acoustically similar output forms. It was also concluded that the cues of spectral tilt change for plosives and spectral mean and \( F_2 \) onset frequency provide an effect of within-category refinement for already established contrasts.

11.3 Underlying representation confusion

An under-specified underlying representation of the target form can result in production confusion. Several instances of production patterns that agree with a confusion of output forms were found in the UC^3 Corpus. For example, long lag VOT productions were found in attempted productions of unaspirated target plosives and target words with a structurally simple syllable onset erroneously produced with a complex syllable onset were found in the UC^3 Corpus.

Kewley-Port & Preston (1974) argued that both prevoicing and long lag productions are motorically more complex production patterns than short lag productions. This helps to explain the initial preference of short lag productions. Hence, the presence of long lag VOT in productions where it makes the production less adult-like is interpreted as a sign of a confusion of output form.

Cluster reducing processes such as cluster reduction and coalescence are widely reported in the Literature. Together with the prevalence of structurally simple syllable onset productions in the initial productions found in the UC^3 Corpus and reported in Chapter 7, it is possible to view structurally simple productions as the preferred production pattern. Thus, there is no reason for structurally simple targets being produced as structurally complex due to production reasons. Hence, overly complex productions can be interpreted as instances of confusion between the output forms of rival targets.

All productions differing from an adult target may, however, not be interpreted as an indication of a perceptual confusion between rival output forms. In Chapter 5 it was demonstrated that productions in the initial stages had stronger RELATIVE ASPIRATION AMPLITUDE than those in the later stages. Hence, the significant effect of AGE
on Relative Aspiration Amplitude should be seen as an instance of hyperspeech rather than as an effect of perceptual confusion with other output forms.

Further, some of the acoustic cues of Place of Articulation investigated in Chapter 10, spectral tilt change, $F_2$, spectral mean and spectral mean, displayed refinement within a place category perceived by the adult transcriber but not productionally towards the adult target. Since the developmental trends for these cues do not cause them to increase in acoustic likeness with the intended target production, it is proposed that these productions can be viewed as refinement of an established production pattern rather than an exploration of available contrasts; this progression is viewed as a development in production rather than a development in the child’s underlying specification of the target form.

11.4 Suggestions for future research

The research and the results presented in Thesis have generated a number of possible future research programmes including the following three.

One, as the calculated measurements of stability in production may have been influenced by the end of the investigated age range and the number of analysed productions decreased with age near the positive boundary the study should be replicated with an expanded age range and, if possible, a non-truncated data set.

Two, the UC³ Corpus does not provide an opportunity for sort of analysis conducted by Sussman et al. (1992) using locus equations. A programme of research that permitted the description of acquisition of Place of Articulation using locus equations using a target word corpus with a more varied vocalic context would complement the research presented in this Thesis. A suggestion would be to use target words with consonants followed by both front and back vowels as this would allow the relationship between $F_2$ in the consonant and $F_2$ in the vowel to be quantified by the locus equation and the development in Swedish-acquiring children to be tracked.

Three, the research presented in this Thesis did not include individual case-studies of development. Karlsson, Sullivan, Czigler & van Doorn (2002), Karlsson et al. (2003) and Karlsson & Sullivan (2004) are examples of case studies of participants in the UC³ Corpus. These case studies, however, each focused on a different acoustic cue. Full case studies that examine the full set of aspects examined in this Thesis are needed to explore individual children’s paths through the multidimensional space of a phonetic category may be explored and provide a complement to the group-based findings of this Thesis.
References


References


Fernald, A. (1989), ‘Intonation and communicative intent in mothers’ speech to infants: Is the melody the message?’, *Child Development* 60(6), 1497–1510.


References


References


References


References
Author Index

Abramowitz, Milton 147
Abramson, Arthur S. 1, 13, 14, 56, 69, 74, 181
Ambler, Stephen 69, 74
Aslin, Rickard N. 2

Baran, Jane A. 6, 15
Barton, David 6, 16, 17, 43, 48, 56
Bates, Douglas M. 55, 64, 89
Beskow, Jonas 28, 34
Bird, Steven 34
Black, J. W. 19, 111
Blumstein, Sheila E. 19–21, 135, 136, 138, 140, 142, 154, 182, 183
Bond, Z. S. 2, 6, 17
Braine, M. D. S. 1
Buder, Eugene H. 28, 30

Callow, Mo 69, 74
Caramazza, Alfonso 1
Carbone, Ettore 1
Catts, Hugh W. 6, 16, 17, 43, 48, 56
Cho, Taehong 1, 56
Ciocca, Valter 31
Cleveland, William S. 136
Cooper, F. S 13
Cooper, William E. 6
Czigler, Peter E. 59, 160, 162

Daniloff, Ray 6, 15
Davis, Barbara L. 4
de Boysson-Bardies, Bénédicte 5
DebRoy, Saikat 55, 64, 89

Dechovitz, D. 14
Delattre, P. C. 13
Docherty, Gerard 15
Donegan, P. 1
Dougall, Ronald N. 21, 142
Dyson, A. T. 11

Ebert, Robert R. 6
Eguchi, S. 6, 16, 50, 56
Eimas, Peter E. 18
Erickson, D. M. 14

Faulkes, Paul 15
Fernald, Anne 2, 15
Forner, L. L. 6
Forrest, Karen 21, 142
Fox, Robert Allen 6, 142, 144, 147, 149, 155, 182
Francis, Alexander L. 31
Fujimura, Osamu 69, 74

Garell, P. Charles 56
Gewirth, Letitila 21, 136, 140, 142, 154, 182, 183
Gierut, J. A. 3
Gilbert, John H. V. 6, 16
Glasberg, Brian R. 142
Grunwell, Pamela 11, 19, 97, 110, 115, 127, 130, 136

Haggard, Mark 6, 13, 14, 69, 74
Hawkins, Sarah 6, 88
Herriman, Eve 14, 67
Hirsh, I. J 6, 16, 50, 56
Hoemeke, Kathrine A. 6, 162
Holt, Lori L. 14, 69, 70
Howard, Matthew A. 56

Ingram, David L. 1, 19, 110, 115, 127, 130
International Phonetics Association 30

Jenison, Rick L. 56
Jennische, Margareta 11
Jongman, Allard 6, 21, 142, 147, 149
Jusczyk, Peter W. 2

Kamhi, Alan G. 6, 16, 17, 43, 48, 56
Karlsson, Fredrik 50, 55, 59, 88, 89, 159, 160, 162
Kent, Raymond D. 4, 6, 16, 50
Kewley-Port, Diane 16, 17, 20, 21, 42, 43, 51, 56, 57, 74, 136, 140, 161
Khami, Alan G. 6
Klatt, Dennis H. 13, 14
Klatt, Mary 17
Kluender, Keith R. 14, 56, 69, 70
Koenig, Laura L. 6, 47, 50, 55
Koenigsknecht, Roy A. 6, 16, 17, 56, 74
Kuhl, Patricia K. 2, 6

Lacerda, Francisco 15
Ladehoffed, Peter I. 56
Lahiri, Aditi 21, 136, 140, 142, 154, 182, 183
Landgren, Tomas 50
LaRiviere, Conrad 14, 67
Laufer, Marsha Zlatin 6, 15
Lee, Sunbok 6, 28, 30
Liberman, Mark 34
Lieberman, A. M. 13, 14
Lieberman, Philip 6, 17
Lindblom, Björn 2, 64, 65
Linell, Pär 11
Lisker, Leigh 1, 13, 14, 56, 69, 74, 181
Liu, Huei-Mei 2
Lotto, Andrew J. 14, 56, 69, 70
Luce, Paul A. 2
Lundberg, Katharina 24

Mack, Molly 6, 17
Macken, Marlys A. 1, 6, 16, 17, 43, 48, 56
MacNeilage, Peter F. 4
Malsheen, Bathsheba J. 15, 56
Mandler, R. 14
Mattys, Sven L. 2
Maxwell, E. M. 1
McCaffrey, Helen A. 6, 162
McCrea, Christopher R. 69, 74
McGowan, Rickard S. 22, 149, 151, 155
McLeod, Sharyne 3, 6, 11, 12
Meltzoff, Andrew N. 6
Menyuk, Paula 17
Milenkovic, Paul 21, 142
Miller, Joanne L. 18
Miller, Marnie E. 22, 142
Mody, Maria 14, 69, 74
Moore, Brian C. J. 142
Morgan, James L. 2
Morris, Richard J. 69, 74
Murray, Ann D. 4

Narayanan, Shriyanth 6, 28, 30
Newport, Elissa L. 2
Nissen, Shawn 6, 142, 144, 147, 149, 155, 182
Nittouer, Susan 22, 142, 144, 147, 149, 151, 154, 155, 182
Noh, Daniel 56

Ohde, Ralph N. 69, 70, 73–75
Oller, Kimbrough D. 4–6

Paden, E. P. 11
Pinheiro, José C. 55, 64, 89
Pisoni, David B. 20, 21, 42, 136, 140
Potamianos, Alexandros 6, 28, 30
Preston, Malcolm S. 16, 17, 42, 43, 51, 56, 57, 74, 161
Priestly, T.M.S. 3
Purves, B.A. 6

Reed, Vicki A. 11, 12
Repp, Bruno R. 14, 67, 68, 73, 74
Roberts, Martin 14, 69, 74

Saffran, Jenny R. 2
Sarkar, Deepayan 55, 64, 89
Sawusch, James R. 42
Shankweiler, Donald 18
Shockey, Linda 2
Singh, Sadanand 19, 111, 133, 159
Sjölander, Kåre 28, 34
Smith, Bruce L. 6
Smith, Niels V. 1
Soli, Sigfrid D. 6, 21, 142
Stampe, D. 1
Stegum, Irene A. 147
Steinschneider, Mitchell 56
Stevens, Kenneth N. 13, 14, 19–21, 69, 74, 135, 136, 138, 140
Studdert-Kennedy, Michael 18, 20–22, 136, 140, 142, 149, 151, 155
Sullivan, Kirk P. H. 50, 55, 59, 160, 162
Summerfield, Quentin A. 13, 14, 69, 74
Sundberg, Ulla 15
Sussman, Harvey M. 6, 162
Thelen, Esther 4
Tsao, Feng-Ming 2

van Doorn, Jan 3, 6, 11, 12, 59, 162
Vetenskapsrådet 24
Vihman, Marilyn May 5
Volkov, Igor O. 56
Wat, Dominic 15
Wayland, Ratree 6, 21, 142, 147, 149
Weismer, Gary 21, 142
Whalen, Douglas H. 14, 69, 74
Wilson, H. F. 6, 17
Winitz, Harris 14, 67
Wong, Serena 6, 21, 142, 147, 149
Yeni-Komshian, G. H. 1
Yu, Jojo Man Ching 31
Zetterholm, Elisabeth 50, 55
Zlatin, Marsha A. 6, 13, 16, 17, 56, 74
Zurif, E. B. 1
Author Index
Subject Index

acoustic correlates
  $f_0$, 6, 7, 13–15, 22, 28, 30, 67, 69–75, 161, 181
  $F_2$, 13, 21, 22, 28, 30, 31, 135, 142, 149, 151–153, 155, 161, 162, 182
relative aspiration amplitude, 22, 60–69, 73, 74, 161, 162, 182
syllable onset duration, 88, 92
utterance amplitude, 68, 69
Voice Onset Time, 13–17, 22, 39, 40, 42–44, 46–60, 63, 64, 66, 70, 160, 161, 181

acoustic cues
  kurtosis, 21, 22, 54, 135, 136, 138, 142, 147, 149, 150, 182
skewness, 21, 22, 50, 51, 54, 136, 138, 140, 142, 147, 148, 154, 155, 161, 182
spectral mean, 21, 22, 142, 161, 162
spectral tilt change, 21, 22, 136, 140, 142, 143, 154, 161, 182
spectral variance, 21, 22, 136, 138, 142, 144, 146, 147

analysis factors
age category, 42–44, 47, 51–54, 60–63, 70–73
manner of articulation, 11, 17, 19, 22, 79, 97–113, 121, 125, 130, 131, 133, 134, 159, 160
participant, 26, 34, 41, 68, 69, 73, 80, 83, 144, 147, 149, 151
session, 34, 69
syllable onset complexity, 84, 89, 159, 160
Voice Onset Time category, 39, 42–47, 56, 68, 72, 73
long lag, 16, 17, 39, 42–44, 47–51, 56, 57, 68, 70, 160, 161
prevoicing, 16, 17, 43, 44, 48, 49, 52, 57, 160, 161, 181
short lag, 16, 17, 42–44, 47, 50, 51, 56, 57, 68, 161
voicing category, 40, 41, 43–47, 53–56, 59–64, 68–71, 73, 74
cluster reduction, 11, 12, 23, 24, 161
cluster simplification, 12, 23, 24
cluster simplifying processes
  cluster reduction, 11, 12, 23, 24, 161
  cluster simplification, 12, 23, 24
  coalescence, 11, 12, 23, 24, 161
coalescence, 11, 12, 23, 24, 161
dentalisation, 130
$F_2$, 13, 21, 22, 28, 30, 31, 135, 142, 149, 151–153, 155, 161, 162, 182
<table>
<thead>
<tr>
<th>Subject Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>fronting, 23, 115, 130, 131</td>
</tr>
<tr>
<td>hyperspeech, 2, 5, 162</td>
</tr>
<tr>
<td>kurtosis, 21, 22, 54, 135, 136, 138, 142, 147, 149, 150, 182</td>
</tr>
<tr>
<td>long lag Voice Onset Time, 16, 17, 39, 42–44, 47–51, 56, 57, 68, 70, 160, 161</td>
</tr>
<tr>
<td>phoneme simplifying processes</td>
</tr>
<tr>
<td>dentalisation, 130</td>
</tr>
<tr>
<td>fronting, 23, 115, 130, 131</td>
</tr>
<tr>
<td>stopping, 19, 23, 107, 112, 132</td>
</tr>
<tr>
<td>velar harmony, 23, 116, 130, 131</td>
</tr>
<tr>
<td>phonological processes</td>
</tr>
<tr>
<td>cluster reduction, 11, 12, 23, 24, 161</td>
</tr>
<tr>
<td>cluster simplification, 12, 23, 24</td>
</tr>
<tr>
<td>coalescence, 11, 12, 23, 24, 161</td>
</tr>
<tr>
<td>dentalisation, 130</td>
</tr>
<tr>
<td>fronting, 23, 115, 130, 131</td>
</tr>
<tr>
<td>stopping, 19, 23, 107, 112, 132</td>
</tr>
<tr>
<td>velar harmony, 23, 116, 130, 131</td>
</tr>
<tr>
<td>phonological re-organisation, 57</td>
</tr>
<tr>
<td>place of articulation cues</td>
</tr>
<tr>
<td>kurtosis, 21, 22, 54, 135, 136, 138, 142, 147, 149, 150, 182</td>
</tr>
<tr>
<td>skewness, 21, 22, 50, 15, 145, 54, 136, 138, 138, 142, 147, 148, 154, 155, 156, 182</td>
</tr>
<tr>
<td>spectral mean, 21, 22, 142, 161, 162</td>
</tr>
<tr>
<td>spectral tilt change, 21, 22, 136, 140, 142, 143, 154, 161, 162, 183</td>
</tr>
<tr>
<td>spectral variance, 21, 22, 136, 142, 144, 146, 147</td>
</tr>
<tr>
<td>prevoicing, 16, 17, 43, 44, 48, 49, 52, 57, 160, 161, 181</td>
</tr>
<tr>
<td>relative aspiration amplitude, 22, 60–69, 73, 74, 161, 162, 182</td>
</tr>
<tr>
<td>short lag Voice Onset Time, 16, 17, 42–44, 47, 50, 51, 56, 57, 68, 161</td>
</tr>
<tr>
<td>skewness, 21, 22, 50, 51, 54, 136, 138, 140, 142, 147, 148, 154, 155, 161, 182</td>
</tr>
<tr>
<td>spectral mean, 21, 22, 142, 161, 162</td>
</tr>
<tr>
<td>spectral tilt change, 21, 22, 136, 140, 142, 143, 154, 161, 162, 182</td>
</tr>
<tr>
<td>spectral variance, 21, 22, 136, 142, 144, 146, 147</td>
</tr>
<tr>
<td>stopping, 19, 23, 107, 112, 132</td>
</tr>
<tr>
<td>syllable onset duration, 88, 92</td>
</tr>
<tr>
<td>target word</td>
</tr>
<tr>
<td>/bak/, 39, 88, 100, 107, 110, 116, 125, 131</td>
</tr>
<tr>
<td>/bor/, 39, 88, 101, 107, 110, 125, 131, 133, 134</td>
</tr>
<tr>
<td>/do/, 39, 88, 107, 110, 125, 130</td>
</tr>
<tr>
<td>/gal/, 24, 39, 83, 88, 92, 94, 98, 107, 115, 125</td>
</tr>
<tr>
<td>/kal/, 24, 39, 83, 88, 92, 98, 106, 115, 125</td>
</tr>
<tr>
<td>/pak/, 39, 88, 92, 100, 106, 110, 116, 125, 131</td>
</tr>
<tr>
<td>/por/, 39, 88, 106, 110, 125, 131, 133</td>
</tr>
<tr>
<td>/sak/, 88, 100, 106, 110, 116, 117, 125, 131</td>
</tr>
<tr>
<td>/sal/, 24, 80, 83, 88, 98, 106, 107, 115</td>
</tr>
<tr>
<td>/skal/, 24, 39, 80, 83, 88, 90, 92, 94, 100, 101, 103, 107, 111, 115, 127, 132</td>
</tr>
<tr>
<td>/so/, 88, 90, 106, 127, 130</td>
</tr>
<tr>
<td>/sol/, 88, 106, 131</td>
</tr>
<tr>
<td>/spek/, 39, 80, 100, 101, 103, 110, 116, 127, 131</td>
</tr>
<tr>
<td>/spik/, 24</td>
</tr>
<tr>
<td>/spor/, 39, 80, 90, 100, 101, 103, 110, 117, 127, 131</td>
</tr>
<tr>
<td>/stc/, 39, 80, 85, 100, 101, 103, 111, 117, 127, 130</td>
</tr>
<tr>
<td>/stcr/, 11</td>
</tr>
<tr>
<td>/to/, 39, 88, 92, 106, 110, 125</td>
</tr>
<tr>
<td>target word quadruple</td>
</tr>
<tr>
<td>skal, 83, 100, 115, 125, 130</td>
</tr>
<tr>
<td>spar, 131</td>
</tr>
</tbody>
</table>
Subject Index

spak, 100, 117, 131
velar harmony, 23, 116, 130, 131
Voice Onset Time, 13–17, 22, 39, 40,
42–44, 46–60, 63, 64, 66, 70,
160, 161, 181
Voice Onset Time category
long lag, 16, 17, 39, 42–44, 47–51,
56, 57, 68, 70, 160, 161
prevoicing, 16, 17, 43, 44, 48, 49,
52, 57, 160, 161, 181
short lag, 16, 17, 42–44, 47, 50, 51,
56, 57, 68, 161
voicing cues
$ f_0 $, 6, 7, 13–15, 22, 28, 30, 67, 69–75, 161, 181
relative aspiration amplitude, 22,
60–69, 73, 74, 161, 162, 182
Voice Onset Time, 13–17, 22, 39,
40, 42–44, 46–60, 63, 64, 66,
70, 160, 161, 181
Appendix A

Computation of acoustic correlates of phonetic features

This appendix provides an overview of the calculations done based on the acoustic features marked by the procedure presented in Chapter 3. The resulting quantities were used in this thesis as potential acoustic quantifications of the phonetic properties of voicing/aspiration, place of articulation and manner of articulation.

A.1 Derived temporal quantities

A.1.1 Phone and syllable duration

The duration of aspirated sections, phones and syllables were calculated as the difference in time between the offset mark and the onset mark of the element. The placement of onset and offset marks is described in Section 3.3.3.

Syllable boundaries were not marked separately. For syllable duration calculations, the start time of the first phone of the syllable onset was used as the start time of syllable and the end time of the last phone of the syllable coda was used as the end time of the syllable.

A.1.2 Voice Onset Time

The Voice Onset Time (VOT) of produced plosives was computed as the difference in time between the onset of voicing and the release of the plosive (Lisker & Abramson 1964). Thus, a negative VOT indicates an onset of voicing before the release (prevoicing) and a positive VOT indicates an onset of f_0 after the time of plosive release. See Section 3.3.3 for markup criteria used for plosive release and voicing onset.
A.2 Derived spectral quantities

A.2.1 Degree of $F_2$ co-articulation

The degree of $F_2$ co-articulation was only calculated for pre-vocalic fricatives. The degree of $F_2$ co-articulation was calculated as the difference between the $F_2$ onset frequency in the fricative and the $F_2$ frequency at the time of reaching a steady state in the following vowel.

A.2.2 Relative aspiration amplitude

The quantification of aspiration force provided by relative aspiration amplitude used the marked points of aspiration onset, aspiration offset and the onset of the following vowel to define the periods of amplitude extraction. The relative aspiration amplitude was calculated as the difference between the mean aspiration amplitude and the amplitude of a vocalic portion of the same length extracted from the following vowel. The mean aspiration amplitude (in dB) was calculated as the mean amplitude between the times of onset and offset of aspiration, and the amplitude of vocalic portion of the same length extracted from the following vowel the vowel (also in dB) was calculated from the onset of the vowel.

A.2.3 Spectral skewness

Spectral skewness, the third spectral moment, was calculated from the extracted spectrum. The formula used in the calculation of the skewness value is presented in Equation A.1, where $x$ is the spectral amplitude value, $N$ is the number of sample points and $s$ is the standard deviation of $x$.

$$\mu_3 = \frac{\left(\sum_{i=1}^{N}(x_i - \bar{x})^3\right)}{Ns^3}$$  (A.1)

A.2.4 Spectral kurtosis

Spectral kurtosis, the fourth spectral moment, was calculated from the extracted spectrum. In order to provide kurtosis values that are comparable to those reported in the literature (Nittrouer 1995, Nissen & Fox 2005), the formula for the kurtosis proper, presented in Equation A.2 was used. In Equation A.2, $x$ is the spectral amplitude value, $N$ is the number of sample points and $s$ is the standard deviation of $x$.

$$\mu_4 = \frac{\left(\sum_{i=1}^{N}(x_i - \bar{x})^4\right)}{Ns^4}$$  (A.2)

A.2.5 Spectral tilt change

The acoustic quantification of spectral tilt change from the release burst to the onset of the vowel proposed by Lahiri et al. (1984) was used in this thesis. The calculations are based on the amplitude found at 1500 Hz and 3500 Hz in the spectrum of the release ($a$ and $b$) and at 1500 Hz and 3500 Hz at the onset of voicing ($c$ and $d$). Based on these
amplitude measurements, Lahiri et al. (1984) proposed a metric of spectral tilt change shown in Equation A.3.

$$q = \frac{(d - b)}{(c - a)} \quad (A.3)$$

Lahiri et al. (1984) argued that labial plosives were characterised by $q > 0.5$ of $(c - a) < 0$. Dental and alveolar plosives were described as having a $q < 0.5$. 
Appendix B

Publications and presentations while a PhD student


