

SUMMARIES OF PH.D. THESES

TEMPORAL COORDINATION IN SPEECH DEVELOPMENT A study on voicing contrast and assimilation of voice

author: Cecile T.L. Kuijpers

promotor: L.C.W. Pols

co-promotor: F.J. Koopmans-van Beinum

date of defence: 7 April 1993

Summary

In this thesis we study several temporal aspects in the speech development of Dutch children, concentrating on the voiced-voiceless distinction and on assimilation of voice. The most important research question is: how does the process of temporal coordination develop in the speech of children, and, above all, how does it develop towards the adult model? The term 'temporal coordination' refers to how segment durations are realized, and how the durational aspects of segments are influenced by each other. This theme returns each time in the subsequent chapters.

First, we pay attention to an existing language-independent model which presents the development of sound production in infants. This model shows that, during the first year of life, phonation and articulation interact. We consider the interaction of voicing and non-voicing as an early process of temporal coordination. The various experiments described in this thesis show that temporal coordination remains an important process in speech development. Actually, in each experiment we study the process from a different angle, and the experiments are characterized by several 'underlying aspects' (*Chapter I*).

Next, we argue for a natural setting in phonetic experimental research with children. This is illustrated by means of three pilot studies. The first two pilot studies relate to the difference between imitative and spontaneous speech; we concentrate on syllable durations, vowel durations, and initial voiced and voiceless stops in the speech of two Dutch children at the ages of 2;3 and 2;6 (in years; months). In the third pilot study we go further into the development of the voicing contrast of initial /p b/ vs. /t d/. We study a group of children the age of 1;5 and 3;9. The results show that the realization of /b/ and /d/ is often characterized by the insertion of a schwa-like sound. Actually, the stops are in intervocalic position now.

These observations give rise to the main research questions: how do the temporal parameters of intervocalic stops develop (in speech production *and* perception), and how do they develop in two-obstruent sequences which are characterized by assimilation of voice. Before passing on to a description of the production and perception experiments, we pay attention to the mechanism of voice production as well as to several measurement procedures that are generally used to chart voicing. Next to it, we concentrate on the acoustic parameters that characterize the intervocalic voicing contrast in Dutch, and to the measurement procedure that we actually chose (*Chapter II*).

In the production experiment we investigate the intervocalic voicing contrast (/b/ vs. /p/ and /d/ vs. /t/), and the vowel contrast (short vs. long). Therefore, we analyse closure duration, burst duration, and preceding vowel duration in spontaneous but controlled speech utterances of four-year-olds, six-year-olds, twelve-year-olds, and adults. In adult speech, closure duration of voiced stops are relatively short, whereas those of voiceless stops are relatively long. Furthermore, the adult speakers display a 'temporal compensation': vowels before voiced consonants are relatively long, and vowels before voiceless consonants are relatively short. In both contexts, the total duration of the vowel-consonant (VC) sequence remains the same.

Several developmental tendencies are deduced from this part of the study. With respect to both voicing and vowel contrast, we express the durational contrast first in terms of a 'relative contrast' (ratio voiced/voiceless, and ratio short vowel/long vowel). Our data show that two contrastive sounds become more and more distinct with age, and we call this an increasing 'distinctivity'. We express this developmental aspect by means of a measure of distinctivity which is based upon the overlap of two frequency distributions: the smaller the overlap, the higher the distinctivity. The temporal compensation takes form between the age of four and six, and it develops from a sequential coordination (no temporal compensation) towards a progressive coordination (incomplete compensation), to finally reach the complete coordination (total compensation).

Furthermore, the four-year-old children realize the voiceless stop /k/ with a relatively short closure duration as compared to /p/ and /t/. We attempt to explain this fact by the lack of the velar contrast /k/ vs. /g/ in Dutch, as well as by some physiological factors (*Chapter III*).

Another research question relates to the perception of the word-medial voicing contrast. How does it develop, is there a parallelism between production and perception, or do children perceive more or better than they can produce? Therefore, we study the same age groups and we use natural but manipulated speech. The silent interval is manipulated from 10 msec up to 130 msec which, in general, results in a categorical perception of voiced and voiceless stops. We are forced to use nonsense words, and we invent a game as identification experiment.

The statistical analyses are based upon curve fitting: the data are transformed to z-scores and a regression analysis is carried out. The most important differences are found between the two younger age groups (4 and 6) and the two older age groups (12 and adults). With respect to the phoneme boundary (the 50% crossover), which tends to decrease with age, the data show no significant differences between the age groups. The most important difference between the younger and the older age groups concerns the phoneme boundary width (the interval between the 25% and 75% point on the identification function). This can also be deduced from the increasing steepness of the function.

As children grow older, they respond more and more accurately to the durational difference that brings about the voiced-voiceless distinction. Apparently, the acoustic difference they need to perceive the voicing distinction diminishes. We show that this age-dependent, auditive acuity relates to the consistency of judgments in the different age groups. The adult listeners behave as a homogeneous group, whereas the young children display a variable perceptual behaviour. On the basis of our data, we show that the relation between production and perception of the voicing contrast is characterized by a parallelism in development. In both experiments distinctivity of the phonemic contrast increases with age (*Chapter IV*).

Furthermore, we examine the development of assimilation of voice in intervocalic two-obstruent clusters. Assimilation of voice occurs rather frequently in speech of Dutch adults (e.g. 'stropdas' (tie) pronounced as [strobdas]). In general, this phenomenon is represented by an ordered set of phonological rules which indicate

when voice assimilation occurs, when it is progressive (voiceless cluster), and when it is regressive (voiced cluster). Until now, phonetic research on assimilation of voice in Dutch has only been performed with adults, and we do not know whether voice assimilation occurs in the speech of children.

We analyse spontaneous but controlled speech utterances of six-year-olds, twelve-year-olds, and adults. The four-year-old children did not participate in the experiment for practical reasons. All clusters are sequences of two obstruents, viz. stop-stop, fricative-stop, stop-fricative, and fricative-fricative. In addition, we examine the clusters in compound words (e.g. 'voetbal' (football)), as well as across a word boundary (e.g. 'groot beast' (big animal)). We call this a difference in 'linguistic context'. With respect to the children's speech utterances, we make use of adjusted criteria in order to classify the different types of assimilation, and to compare the number of occurrences with those found in adult speech.

The results show that the clusters with a rightmost fricative are always assimilated progressively, by the children as well as the adults. This confirms the phonological rule. The other phonological rules, however, are less clearly present in the children's data. Concerning the fricative-stop and stop-stop clusters, the most important difference between the groups of children (6 and 12) and the adults, relates to the type of assimilation. Children do not assimilate less than adults, but they assimilate differently. Whereas regressive assimilation is predominant in the adult speech utterances, progressive assimilation predominates in the speech of both groups of children. The children tend to devoice the two-obstruent clusters irrespective of the linguistic context.

From a phonetic point of view, children do not seem to be able to anticipate on the voicedness of the rightmost obstruent, and they possibly cannot realize or maintain vocal cord vibration during a relatively long obstruction. Consequently, the cluster becomes voiceless. From a phonological point of view, it seems that children generalize the voiceless character to clusters in compound words as well as in two-word items, i.e. they insert the unmarked value [-voice]. We interpret the children's data within an autosegmental framework, which leads to a parsimonious description of children's voice assimilation and the possible learning process (*Chapter V*).

In the final chapter we integrate the main findings of the experiments in a general discussion. We indicate the limitations of the present research, but we also make suggestions for future research. This is done in conformity with the underlying aspects, as described in the first chapter. All comments are arranged according to language-related and person-related aspects. All in all, we hope that the many experimental choices and findings reported in this thesis will contribute to further firmly-based phonetic research (*Chapter VI*).

SPECTRO-TEMPORAL FEATURES OF VOWEL SEGMENTS

author: Rob van Son
promotor: L.C.W. Pols
date of defence: 3 September 1993

Summary

In this thesis we have investigated several aspects of the spectro-temporal structure of vowel segments, both concerning vowel production as well as vowel perception. Chapter 1 contains a summary of current models on vowel production and perception. Models of vowel pronunciation try to explain why vowel realizations vary so much in natural speech. It is known that vowel production is influenced in highly systematic ways by context, stress, and speaking style (among others). The classical explanation is that of the target-undershoot model. This model states that vowel articulation is limited by the speed of the articulators (e.g., jaw, tongue, lips). Each vowel has a unique target-position for each of the articulators which will produce the ideal, or canonical, realization of that vowel. When vowel realizations are very long, there is ample time for the different articulators to reach their respective target positions. However, when vowel duration is short and the context forces the articulators to cover relatively large distances, there is not enough time and the articulators are stopped short of their targets. The resulting vowel realizations show "undershoot" in their articulatory movements as well as in the resulting formant frequencies, hence the name of the model: target-undershoot.

The classical quantitative study of Lindblom (1963) on the relation between vowel duration and formant-undershoot is discussed in depth. It showed that formant-undershoot increased exponentially with a decrease in vowel duration. However, subsequent studies gave ambiguous results. Some studies did find clear evidence for articulatory- and formant-undershoot. Others showed that there were numerous cases where no relation between vowel duration and target-undershoot could be found. Especially, changes in stress and speaking style could bring about changes in duration that were not accompanied by changes in target-undershoot. In our opinion, these conflicting results can be explained by assuming that target-undershoot is planned by the speaker. In this view, the undershoot serves a purpose that depends on factors like context, prosody, and speaking style. From this it follows that, irrespective of vowel duration, the undershoot itself should not change if the purpose of the undershoot does not change and vice versa.

Considering the conflicting reports in the literature, it seems that any test of the target-undershoot model should introduce changes in vowel duration without changing stress, speaking style, or other prosodic factors that were known to cause ambiguous results. In this study, we settled for changes in speaking rate. A long, meaningful text, read at a normal and at a fast rate, would induce a speaker to use the same stress assignments and the same "style" of speaking, irrespective of reading speed. At the same time, a difference in speaking rate would change the duration of all the vowels. In this study (chapters 2-4), we used all realizations of seven different vowels and some realizations of the schwa (/ə/). If vowel duration could control formant-undershoot all by itself, then an increase in speaking rate should induce an increase in undershoot.

However, if formant-undershoot is planned, then a change in speaking rate should not necessarily result in a change in formant-undershoot.

In chapter 2, we measured formant frequencies in the *vowel kernel*. Vowel realizations uttered at the normal speaking rate were compared to the corresponding realizations uttered at the fast speaking rate. No spectral vowel reduction was found that could be attributed to a faster speaking rate. There was also no change in the amount of coarticulation or stress-induced reduction as a result of speaking rate. The only systematic effect was a higher F_1 value in fast-rate speech irrespective of vowel identity. This possibly suggests a generally more open articulation of vowels, speaking louder, or some other general change in speaking style by our speaker when he speaks fast.

In chapter 3 we looked at the effects of speaking rate on *vowel formant track shape*, using the same material as in chapter 2. The formant track shape was assessed on a point-by-point basis, using 16 samples at the same relative positions in the vowels. Differences in speaking rate only resulted in the same uniform change in F_1 frequency already found in chapter 2. Within each speaking rate, there was only evidence for a weak leveling off of the F_1 tracks of the open vowels /a a/ with shorter durations. When considering sentence stress or vowel realizations from a more uniform, alveolar-vowel-alveolar context, these same conclusions were reached.

In chapter 4 we again looked at the effects of speaking rate on formant track shape. This time we used a more elaborate method for assessing formant track shape. Legendre polynomial functions were used to model and quantify the shape of time normalized formant tracks. No differences in these normalized formant track shapes were found either that could be attributed to differences in speaking rate. A uniform higher F_1 frequency in fast-rate speech relative to normal-rate speech was found. Within each speaking rate, there was only evidence of a weak leveling off of the F_1 tracks of the open vowels /ε a a/ with shorter durations. Again, as in chapter 3, separately inspecting vowel realizations from a more uniform, alveolar-vowel-alveolar context, did not alter our conclusions.

The target-undershoot model of vowel production inspired a complementary model of vowel perception (Lindblom and Studdert-Kennedy, 1967). As vowel formant tracks will systematically undershoot the canonical target values in natural speech, it was suggested that listeners would compensate for this undershoot automatically by systematically overshooting the formant frequencies actually reached in perception, i.e. perceptual-overshoot. Early studies with synthetic speech did indeed find this kind of perceptual-overshoot. However, it showed to be rather difficult to prove the existence of an automatic mechanism for perceptual-overshoot in natural speech.

At the moment, there are two classes of models on vowel perception. The first class are models with dynamic-specification. In these models it is assumed that listeners use dynamical information from the Consonant-Vowel and/or Vowel-Consonant transitions to improve the recognition of the, stationary, vowel nucleus. Perceptual-overshoot is just one of such models. The second class of models is based on the assumption that a single, spectral, cross-section of the kernel of a vowel realization contains all information necessary to recognize it. In these models the vowel on- and offset transitions are of minor importance in vowel recognition.

The difference between these two types of models is the position of the Consonant-Vowel transition (in the vowel on- and offset). Is it used in vowel recognition, as is stated by models using dynamic-specification, or is it not, as stated by target models? There is evidence for perceptual-overshoot in synthetic speech. It is also known that presenting syllables without a vowel kernel, i.e. with only the vowel on- and offset transitions, hardly impairs vowel recognition. Still, there is no undisputable proof that the recognition of isolated, monphthongal, vowel segments is improved by adding

dynamical information to the formant tracks. Exactly such an improvement is expected when listeners use dynamic-specification of vowels.

In natural speech, the amount of variation in durations, vowel formant frequencies and track shapes is limited. These various types of variation are furthermore strongly correlated. It is therefore better to use synthetic speech, for which it is possible to control all features. With synthetic speech, it is also possible to detach formant track shape from formant frequency. This way, the effects of formant track shape can be studied independently of vowel identity and vowel duration. We therefore choose to use synthetic speech to study how vowel duration and formant track shape influence vowel identity. Especially we looked for any evidence for perceptual-overshoot. The result of this study is presented in chapter 5 (see below). In chapter 6 we took a closer look at the existing literature in order to try to find an explanation for the disagreement between our results and those presented in several earlier papers.

In chapter 5 we used synthetic vowels to investigate whether listeners use vowel duration and formant track shape to determine vowel identity. The synthetic vowels had level or parabolically-shaped formant tracks and variable durations. They were presented in isolation as well as in synthetic CVC syllables. There was no evidence of perceptual compensation for expected target-undershoot due to token duration or context. The only asserted effects of duration and context were in the number of long- and short-vowel responses. There was also no evidence that the listeners used the formant track shape or slopes independently to identify the synthetic vowel tokens. Tokens with curved formant tracks were generally identified near their formant offset frequencies.

The results of chapter 5 contradicted claims made in the literature about the way listeners use dynamical information to identify vowel realizations. The literature on vowel perception itself also contains contradictory claims regarding the use of information from CV-transitions in vowel recognition. Our own experiments showed that the information in formant track shape was not always used to compensate for formant-undershoot. In chapter 6 a re-evaluation of the literature is attempted. A closer study of the most relevant papers shows that evidence for compensatory processes, i.e. perceptual-overshoot and dynamic-specification, was only found when vowel realizations from different, and appropriate, context were contrasted. Some studies show that vowel recognition deteriorated when vowel segments were presented out of context. Together, these facts suggest that the presence of an appropriate context is essential for any perceptual compensation of coarticulatory changes. This speculation might be used as a starting hypothesis for further research on vowel perception.

Finally, in chapter 7 we summarize and discuss our findings. We recapitulate the methods used in chapters 2 to 4 to study the effects of speaking rate on formant-undershoot. We argue that, under the circumstances used, any excess undershoot due to an increase in speaking rate should have been detectable, but did not show up. We therefore conclude that, for our speaker, speaking rate did not influence the amount of vowel formant-undershoot or the formant track shape. Therefore, we can conclude that changes in vowel duration alone do not change the amount of target-undershoot and that the undershoot that does occur is probably planned.

The listening experiments presented in chapter 5 showed that our listeners did not use a perceptual-overshoot mechanism or dynamic-specification to help them identifying the synthetic vowel tokens. In general, they seemed to use the offset part of each vowel realization to identify it. We therefore conclude that listeners do not automatically and unconditionally compensate for the formant-undershoot that can be predicted from the formant track shape.

FINALLY A WORD: A sensori-motor approach of the mother-infant system in its development towards speech

author: Jeannette M. van der Stelt
promotores: L.C.W. Pols & F.X. Plooi¹
co-promotor: F.J. Koopmans-van Beinum
date of defence: 14 september 1993

Summary

In this thesis an approach is presented of mother and infant as a *sensori-motor system* which develops in a speech communication system. In the approach three fundamental characteristics of human communication systems are in focus: *intersubjectivity*, *intentionality*, and *turntaking*. These are present right after birth, although not yet in forms generally known in adult communication systems.

In normal mother-infant interaction both partners adapt their behaviours to create a context of social exchanges. These set the stage for the further development of speech communication. Any abnormalities in this development without obvious physical or mental causes (such as a hearing loss or Down Syndrome) are proposed to originate from early mother-infant interactions.

Two normal mother-infant pairs with different interaction patterns were chosen as test-cases for the approach. The development of these pairs appeared clearly to differ during the research period (from birth to the second birthday of the infants).

In the first chapter the reader is introduced to the human mother-infant interaction in its unique configuration. The fact that mothers and infants are successful in their development towards establishing conversations, that are also understandable for other humans, leads to the idea of *underlying processes*, generally present in mother-infant interaction. Psycholinguistic and psychobiologic literature is presented and related to publications on speech production and language development originating from linguistics, medicine, ethology, and primate evolution. It is concluded that mother and infant form a system that cannot be fully described by characteristics of these two individuals. The two persons mutually regulate each other's behaviour -to an extent not yet fully understood- which is called *coping*.

In previous research in collaboration with Koopmans-van Beinum (1979, 1986), I have described speech motor landmarks in infant sound production that are basic to adult sound production. In the framework of the Netherlands Prevention Fund project, I have observed mother-infant interactions by focusing on their movements during the first two years. These experiences have led to some working hypotheses on the development of speech in infants and on styles of interaction.

In the present approach, the literature data and the practical experiences have merged. Mother-infant interaction was evaluated as early as possible, in single pairs, and in naturalistic home situations. Only the movements were described because such an approach is independent of language and the interpretation of the observers. The

¹ Prof.dr. F.X. Plooi, Vakgroep Psychologie, Groningen

three common characteristics of human communication systems are treated in separate chapters. However, intersubjectivity, intentionality, and turntaking are related: intentionality presupposes an intersubjective orientation towards another person, while turntaking occurs upon transmitted intentions.

The second chapter introduces *two mother-girl pairs*, their medical histories over the two years, psycho-social characteristics (like infant temperament and scores on the Bayley Scales on infant development), and linguistic scores. The differences between the two pairs at the end of the observation period, i.e. when the children are two years old, are supposed to result from the different *interaction* patterns already present soon after birth.

Video-recording procedures, equipment, frequency, and durations are presented. These components originate from a Netherlands Prevention Fund project. Subsequently, the video-recordings of the two pairs as made during the two years have been transcribed in detail by means of a micro-analytic transcription system for movements. All movements of the mother and the infant, which occurred during five minutes per recording, were coded with regard to the body parts moving and the sounds produced. This results in a 16-channel behavioural score, similar to a musical composition for different instruments. This transcription was computer-assisted, and performed by one sole transcriber. Consistency of the transcriber was checked and appeared to be satisfactory (84% as a mean).

Not all movements made by one partner are actually seen by the other. For example, when the infant is looking at the camera, she surely will not see a smile movement on the face of the mother. The procedure to decide upon the classification 'transmitted or not transmitted movements' is described as a sensori-motor transmission model, in which memory for previous movements is neglected.

A computer program FP used for counting was adapted for duration measures. This enabled the calculation of the overall and median durations of specific codes per recording. The micro-analytic data were processed by the program PROGRAAF. This program can select specific channels or codes, indicated for mother or infant from the original transcriptions. In this manner the decomposed movement patterns in the 16 channels can be compiled selectively to obtain more complex behavioural patterns.

Intersubjective tuning is discussed in the third chapter. It is the first characteristic of mother-infant communication systems, and stands for the mutual notion that another human being is present. In the literature on mother-infant interaction it is described in positive terms like togetherness and bonding. In a way, intersubjectivity is already present before the birth, i.e. when the mother is thinking of the baby as a new person.

Our approach employs the transcriptions of movements of mother and infant, and thus intersubjectivity must be translated into movements in which mother and infant mutually orient towards each other.

Three forms of tuning by means of the visual and vocal-aural channels were selected for evaluation. A comparison was made per recording and per pair of the percentage of time and the frequency of the instances of (1) the mother and infant looking at each other's face, (2) their simultaneously producing sounds, and (3) sound production being simultaneous during face-to-face contact.

These three forms of intersubjective tuning appeared to be different for the two pairs in different periods of the development. In one pair (Claire and mother EVE) the presence of face-to-face contact appeared systematically to be less frequent than in the other pair (Fanny and mother SUSAN). Simultaneous sound production was more frequent for Claire and EVE in the first five recordings only. The frequency of vocalisation in unison during face-to-face contact appeared to be higher for Claire and

EVE in the first five months, and lower than for Fanny and SUSAN after the first five months.

The impact of these results for the development of speech communication is discussed. After the fifth month Claire and EVE used the two channels more selectively than Fanny and SUSAN who preferred to use the two channels simultaneously. In a book-reading-situation, Claire and EVE no longer looked at each other but visually focused on a picture; this can immediately be given an audible label, which is an efficient way of communication.

In the fourth chapter *transmission of intentions* is discussed. It is related to the frequencies that a person can see, hear, and interpret movements of a partner. In the literature, intentionality of young infants still is a matter of discussion, in which consciousness and goal-directedness play a major role. In mother-infant interaction an inequality seems to be present, but the mutual readiness to interpret and react to the partner's movements functions as if intentions are transmitted.

During face-to-face contact mother and infant can perceive each other's movements. In my approach, *visual intentions* of a person are assumed when mimical and head movements are seen by his partner during face-to-face contact. As *audible intentions* are assumed those sound productions that occur during face-to-face contact, and as *intense intentions* those combined visual and audible intentions.

The two mother-infant pairs were compared with regard to the three kinds of transmitted intentions. Intra-pair comparisons were made because the mother is expected to transmit more audible intentions to the infant than the infant to the mother, probably thereby instructing the infant about the mother tongue. Inter-pair comparisons of the mothers and the infants were also made of the percentage of time and of the frequencies because intersubjective tuning was different. Equally, the infants were compared, to check if they offered comparable amounts of intentions to their mothers' interpretation.

During face-to-face contact the transmission of *visual intentions* appeared not to be different for the mother and the infant of one pair. However, when comparing the children, Claire appeared to transmit more visual intentions to her mother than Fanny did. During face-to-face contact EVE transmitted more visual intentions to Claire than SUSAN to Fanny, but this difference was not yet significant in the first five months.

Mothers transmitted, as expected, significantly more *audible intentions* to their children than vice versa. Already in the first five months this difference was present, although more clearly for Claire and EVE than for Fanny and SUSAN. The children did not differ, while the mothers differed only with regard to the percentage of time and not for the frequency. This means that EVE's sentences had a longer duration than SUSAN's during face-to-face contact.

The transmission of *intense intentions* was not different for both pairs: the mothers and infants were roughly similar. Within the pairs EVE, however, differed from Claire, because she systematically used the intense intentions during mutual gaze.

The impact of these differences on the development of speech communication is interpreted in the realm of speech instruction, in which the visual information about sound production (the audible intentions) are expected to become redundant.

Turntaking in its simplest form is treated in this thesis in the fifth chapter. It is a well-known aspect in communication systems, and can be regarded as a kind of feedback mechanism. Turntaking implies intentionality and intersubjectivity. In the literature cyclic behaviour is described from an early age onwards, like in gazes at the face of the mother and away from it. After about the fourth month alternated sound production becomes more prominent in mother-infant interaction.

Turntaking by the mother is described upon landmark sound productions (laryngeals, simple articulations, babbling sounds, and words) of the infants. The landmark sounds represent, on the one hand, the ongoing speech motor development of the infants and new sound productions, and, on the other hand, these sounds increasingly resemble adult speech sounds. The mothers are supposed to take audible turns upon these sounds within a certain inter-speaker switch-pause. The mother's turntaking was analysed only with regard to the *onsets* of her utterances because the mothers differed in the amount of sound productions. Per group of landmark sounds the percentages of infant sounds *with* a mother-turn were compared for the two pairs.

Both infants produced sounds in the four groups of sound productions studied. Two of these groups (laryngeals and simple articulations) had their onset in the first two recordings of the infants. EVE took her turns abundantly upon the sounds of Claire. SUSAN took some turns upon Fanny's early landmark sounds, but did so more consistently when the babbling sounds occurred. Fanny was then 32 weeks old. Feedback on sound production started much later for Fanny than for Claire. Fanny produced many more babbling sounds than Claire, possibly because she finally got audible reactions of her mother. One of the conclusions is that *feedback* on later appearing sound productions cannot compensate for the lack of it during the first five months.

The impact for speech development is clear: parents should play the conversational game with their very young infant and should enjoy even the simple sound productions. They will recognise words in the sound stream, and probably sooner than they expected.

The final chapter integrates the previous chapters, and discusses the chosen approach in relation to the results. A surprising result is the crucial impact of interaction patterns, especially during the first five months, upon the outcome of the speech developmental processes at the age of two. The sensori-motor approach has enabled us to formulate suggestions about how the fundamental characteristics of speech communication systems are gradually mastered by the mother and the infant.

Further research is suggested in line with the possibilities of the sensori-motor approach. When speech developmental problems can be predicted already early in mother-infant interaction, such problems can probably be prevented to a large extent as well.

An outline is given for a method to evaluate mother-infant interaction in a laboratory setting. Depending upon the further elaboration of the present ethological approach, and practical and economical consequences, mother and infant pairs that are at-risk for communicative problems, may request for early guidance.