Institute of Phonetic Sciences
University of Amsterdam
Proceedings 23 (1999), 91-102.

AMSTIVOC: TESTING AND ELABORATING THE AMSTERDAM SYSTEM FOR TRANSCRIPTION OF INFANT VOCALIZATIONS

Florien J. Koopmans-van Beinum

Abstract

The need to transcribe infant sound productions from birth onwards by means of universally applicable coding tools has been basic to the development of our AMSTIVOC classification system. In this system we describe early infant vocalizations by means of a sensori-motor approach. And although the system has been used already in several of our own institutional research projects, as well as in a few external ones, the need existed to test the learnability and universality of the system. This paper reports on the experiences and discussions of a small group of international Ph.D.-students who were recently trained in applying the Amsterdam transcription system.

1. Introduction

Since nowadays early speech development and language acquisition are generally seen as a continuous process starting from birth or even beyond, with a close relationship between speech perception and production, researchers are more and more convinced of the necessity to describe the course in the development of infant sound productions from a very early stage onwards. However, since traditional adult phonological transcriptions are not applicable as long as no infant utterances are heard as belonging to a specific language yet, and therefore no target words can be recognized, it is more adequate to base the description tools on universal principles of movements of the human speech production instrument. By applying a source-filter based sensori-motor approach, using global characteristics of place and manner of sound production instead of a phonetic-linguistic description, we developed our **AM**sterdam **S**ystem for **T**ranscription of **I**nfant **VOC**alizations (AMSTIVOC) (Koopmans-van Beinum & Van der Stelt, 1986, 1998; Koopmans-van Beinum, 1990) and applied the system in several of our institutional research projects (e.g., *normalization research:* Van der Stelt & Koopmans-van Beinum, 1986; *diagnostic research:* Koopmans-van Beinum et al. 1990; Jansonius-Schultheiss, 1999; Clement et al., 1995; Van den Dikkenberg-Pot et al., 1998; Koopmans-van Beinum et al., to appear in 2000; *mother-infant interaction research:* Van der Stelt, 1993). In each of these projects slight specific adaptations were made, but basically the system turned out to be very useful. This lead some international English-speaking Ph.D.-students (from Seattle, Grenoble, and Sheffield) to the question whether it might be possible to be trained in using the transcription system for their own research and clinical work on early infant vocalizations as well. The diversity of the students' projects and their scientific aims, and the variety of their individual mother tongue (Hungarian, German, and British English) created a real challenge for testing and elaborating the system in order to make it internationally accessible. As an introduction to working together during the course, each of the students presented briefly her own scientific background and gave a short description of her Ph.D. project, explaining the need for this specific sensori-motor transcription system. As far as possible the structure of the course included theory in the mornings and practical transcription work in the afternoons. One of the sessions was presented by Jeannette van der Stelt, concerning pragmatic application of infant speech acquisition studies and of mother-child interaction studies in particular; all other sessions concerned the actual transcription training.

2. Training procedure

The training procedure was structured in such a way that first the theoretical background of the transcription system was presented, a list of literature concerning early infant vocalizations was read and discussed (a.o. Koopmans-van Beinum & Van der Stelt, 1986, 1998; Oller, 1980; Stark, 1980), and overhead sheets and demonstration materials were studied. Starting point in the theoretical part of the training procedure was the vision that development of speech communication is the result of a cyclic dynamic process in which perception, production, and interaction are closely working in concert. Subsequently the various aspects of classification were explained and perceptually trained. As an exercise the transcription system was applied to prerecorded infant vocalization material. To this purpose use was made of the audio-recorded data from the Dutch project on vocalizations of deaf infants (Clement et al., 1995; Clement, to appear; Van den Dikkenberg-Pot et al., 1998, 1999; Koopmans-van Beinum et al., to appear in 2000). In order to compare the results of perceptual analyses with those of acoustic analyses use was made of the speech analysis program PRAAT (Boersma & Weenink, 1996). The diversity of the students' background and the small group working together allowed for different levels of expertise and for a range of knowledge in acoustics and objective measurement.

In the remainder of this paper the main steps in the transcription procedure will be described together with specific problems or points of discussion that came up during these days of working together. Solutions were chosen after joint deliberation and consultation of the teacher, and if necessary, adaptations in the transcription system were made upon mutual agreement and approval by the teacher. Depending on specific topics references were given each day since the students were highly motivated to read throughout the course. First thing in the morning was a discussion session on the experiences so far. These daily discussion sessions turned out to be very useful and stimulating.

3. Transcription steps and points of discussion

Basic in the AMSTIVOC transcription system is the sensori-motor approach: describing gestures of the infant speech production instrument, instead of a phonetic-linguistic approach in which phonemes are described.

Gestures concern larynx movements (*phonation*) and supraglottal vocal tract movements (*articulation*). Unit of segmentation is bounded by the respiration cycle: an utterance is defined as 'the infant sound produced during one breath unit'. So respiration, phonation, and articulation are the basic elements that have to be classified separately. Although vegetative and discomfort sounds may be considered as essential parts in the infant's speech development as well (cf. Stark et al., 1975), we decided to concentrate on the non-vegetative comfort sounds (Fig. 1).

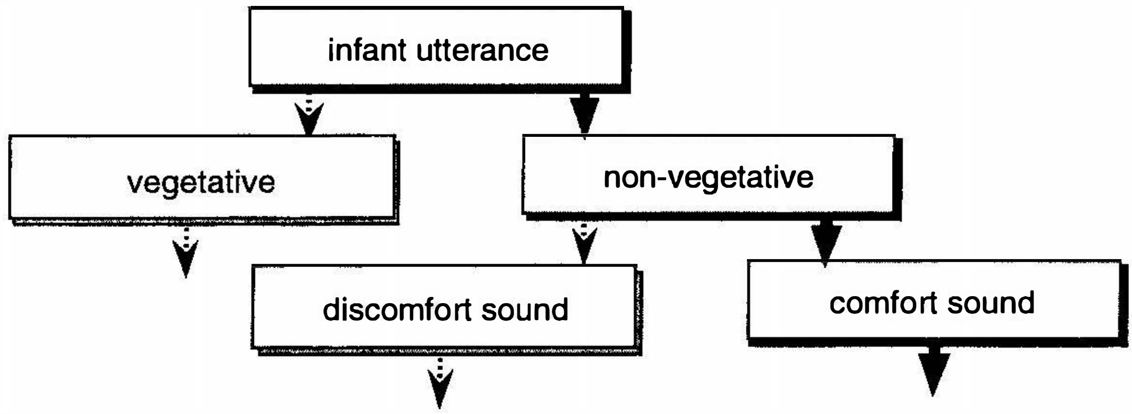

Fig. 1 Decisions in classifying early infant sound productions.

Most times one complete respiration cycle suffices to perceive whether the sound is a crying sound, but the decision whether a sound indicates discomfort is not always clear. Depending on the specific research question one may decide to include these 'grumbling' sounds. By paying attention to phonatory and articulatory aspects separately (Fig. 2) it is much easier to concentrate on specific type of movements made in the infant's speech production instrument.

Aspects of *phonation* to be discerned are:

- occurrence of phonation
- direction of phonation (egressive or ingressive)
- continuity of phonation (continuous or interrupted)
- duration of phonation
- voice onset (aspirated, normal, glottal stop)
- voice quality
- intonation

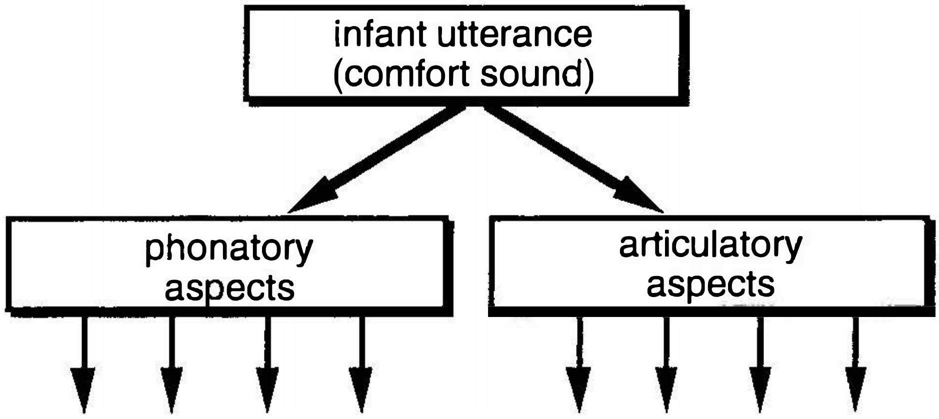

Fig. 2. Classification of phonatory and articulatory aspects separately.

Aspects of *articulation* to be discerned are:
- occurrence of articulation
- number of articulatory movements
- location of articulation within utterance (beginning, middle, or end)
- type of articulatory movements (same or different)
- place of articulation
- manner of articulation

In relationship to the research questions within a certain study it is possible to decide to go into more or less detail. In our own projects concerning speech development of infants in their first year of life (see 1. Introduction) we normally start with classifying phonation in five main classes, depending on continuous or interrupted phonation (the latter requiring a more elaborated coordination of the respiration muscles and the laryngeal muscles), and on variability in intensity, duration, and intonation. The five resulting classes in transcribing phonation then are: Phon 0 (no phonation), Phon 1 (simple, uninterrupted phonation), Phon 2 (simple interrupted phonation), Phon 3 (variation in intensity, duration, and intonation with uninterrupted phonation), and Phon 4 (variation in intensity, duration, and intonation with interrupted phonation).

With respect to articulation three main classes are used, depending on the number of articulations within one respiration cycle: Art 0 (no articulation), Art 1 (one articulatory movement), Art 2 (two or more articulatory movements in a two- or more-syllabic utterance). Here the notion syllable is introduced, on the whole perceptually in line with the acoustical requirements for a canonical syllable as defined by Oller (1986). However, syllabicity caused by simple interruption of phonation without any articulatory movements (*NoArt + IntPhon* and *NoArt + VarIntPhon*) are included in our system as well. To test the applicability and the perceptibility of the notion 'syllable' we did a small perception experiment some years ago. Listeners (n=10) when asked to simply indicate the number of syllables within infant sound productions (n=100), randomly selected from a database of one infant's vocalizations from birth to the age of eight months, agreed up to a level of about 85 % (Koopmans-van Beinum, 1993). Most difficult for indicating number of syllables are utterances classified as *NoArt + VarUnIntPhon* with strong variations in intensity and/or intonation. However, by means of the acoustical requirements of Oller (1986) with respect to intensity, F_0, and duration, these problems can be solved to satisfaction.

Listening to phonatory and articulatory aspects of early infant sound productions separately, turns out to be of great help in the transcription task. The classification results can be summarized subsequently in a matrix for phonation and articulation combined, giving a possibility for further data processing (Table 1). It is worth mentioning here that phonatory aspects like aspiration and glottal stops are classified as non-articulations, in contrast with the traditional phonological transcriptions that indicate them as consonantal.

As mentioned in several previous publications, application of this classification system nicely reveals the developmental course in speech development with a clear hierarchical structure: simple skills have to be mastered before they can be used as building blocks in the more complicated skills (Fig. 3). However, it always should be kept in mind that if a child shows up with a new, more complicated skill, this does not mean that sounds from the previous stages do not occur any more. It is important to make a distinction between a prospective description of stages reached at within the developmental course of the child, and a retrospective description or classification of prelexical utterances at that moment, seen in the light of the final aim of mature speech communication. Detailed descriptions and more specific definitions can be found in Koopmans-van Beinum and van der Stelt (1998).

Table. 1 Matrix for summarizing the results of separate classification of phonation and articulation movements. Impossible combinations are struck out.

Phonation Type	Articulation Type			
	Art 0 *(NoArt)*	Art 1 *(OneArt)*	Art 2 *(TwoArt)*	**Total**
Phon 0 *(NoPhon)*	~~ ~~			
Phon 1 *(UnIntPhon)*				
Phon 2 *(IntPhon)*				
Phon 3 *(VarUnIntPhon)*				
Phon 4 *(VarIntPhon)*				
Total				

Since most studies using the AMSTIVOC transcription system aim at describing the speech development course towards understandable speech, and since the transition to more traditional, linguistic transcriptions is needed, it turned out to be worth while to include a classification of place and manner of articulation for all utterances with one or more articulatory movements. Although the rather inexperienced infant production instrument sometimes creates great difficulties as for perceiving place or manner of articulation, it turned out that the following classification fits well in practice (Clement & Koopmans-van Beinum, 1995, 1999; Clement, to appear).

A global threefold differentiation for *place of articulation* suffices quite well:

- front (labial, labio-dental)
- central (dental, alveolar, palatal)
- back (velar, pharyngeal)

whereas for *manner of articulation* the following differentiation turns out to be quite useful for describing infant sound productions in the first year of life:

- fricatives and trills combined
- stops (including affricates)
- glides (/j/- or /w/-like)
- nasals
- laterals (/l/-like)

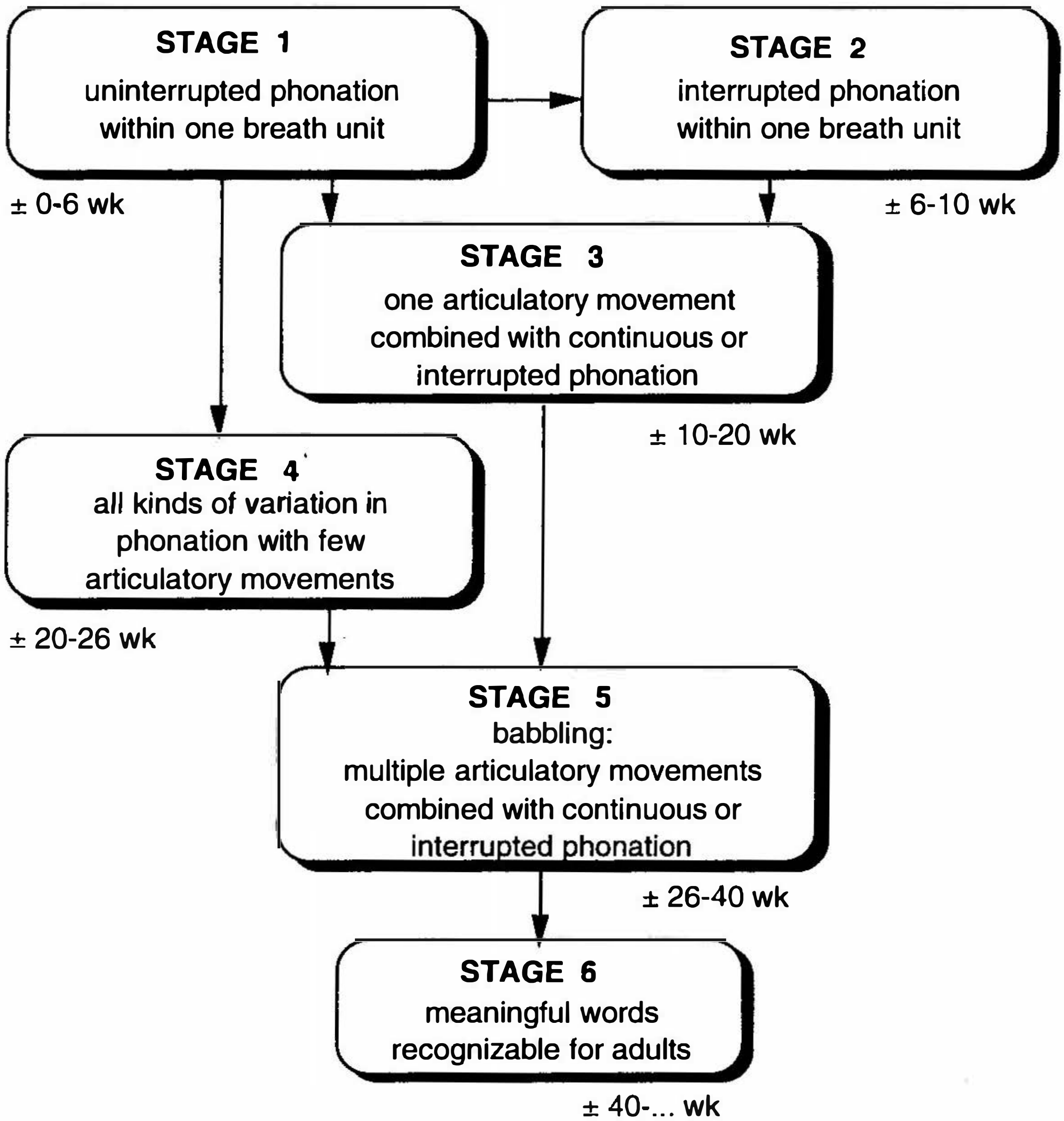

Fig. 3 Hierarchical scheme of speech development characterized on the base of the sensori-motor classification system for early infant sound productions. Arrows indicate the mastering and combination of simple skills in more complicated ones. Ages should be considered as very global indications only.

It will be clear that it is a very simple step to go from this differentiation to a more detailed one, in agreement with the traditional phonological features, for utterances of children in their second year of life when recognizable words emerge.

During the practical work of the course a solution for two specific problems was wanted: how to classify vocalic sounds in more detail and how to transcribe multisyllabic utterances with multiple articulatory movements. In the next section we will describe the solutions we decided to be most useful.

4. Transcription of vocalic sounds

The Amsterdam transcription system so far did not supply the possibility to indicate vocalic characteristics in a more detailed way. However, since it is our aim to connect the results of our transcription system optimally to existing databases (e.g., Davis & MacNeilage, 1995) it is worth trying to classify vocalic parts or 'vocoids' (i.e. NoArt + Phon 1, 2, or 3) by means of the same type of sensori-motor characteristics. In this respect two main aspects presented themselves for classification: position of tongue body and degree of jaw opening. With respect to position of tongue body we differentiated in front, central, and back, whereas for jaw opening we differentiated in four degrees: close, close-mid, open-mid, and open. Subsequently in the same way as was done with respect to the results of the separate classification in articulation and phonation, the results could be combined and summarized in a vocoid matrix (Table 2).

Another question showing up when applying the Amsterdam transcription system in practical exercises concerned the dynamicity of many of the vocalic parts. Especially in the second half of the first year of life children's utterances become more and more dynamic and complex, and the need to mark this in the transcription is obvious. Therefore in order to transcribe these movements within the vocoids it was decided to indicate both the initial and the final position of the tongue body together with the respective degree of jaw opening, in the combinations given in the matrix of Table 2. In Table 3 all possible movements are schematically presented, with this restriction of course that identical initial and final combinations do mean 'no movement'. Applying these movement indications led in general to a high degree of agreement between the students and seems to be a useful addition to the system so far.

Table. 2 Matrix for summarizing the results of classification of position of tongue body and degree of jaw opening for vocalic sounds.

degree of jaw opening	**position of tongue body**			
	front	central	back	**Total**
close				
close-mid				
open-mid				
open				
Total				

Table 3. Schematic representation of possible initial-to-final movements of degree of jaw opening and position of tongue body within vocalic parts of infant vocalizations. Identical initial and final combinations do mean 'no movement'.

possible movements in vocoids

initial position		final position
close / front		close / front
close / central		close / central
close / back		close / back
close-mid / front		close-mid / front
close-mid / central		close-mid / central
close-mid / back		close-mid / back
open-mid / front		open-mid / front
open-mid / central		open-mid / central
open-mid / back		open-mid / back
open / front back		open / front back
open / central		open / central
open / back		open / back

As a consequence of the preceding solutions the question came about whether it might be possible to quantify the degree or the weight of these movements. It is clear that for a child a rather simple movement, e.g., going from open/back to open/central, might be much easier to make than a more complex one, e.g., going from open/back to close/ front. In order to quantify the complexity of these movements we introduced a tentative weighing scheme (see Fig. 4).

Whether this tentative weighing scheme meets the requirements of quantification in a sufficient way could not be tested yet, since the recorded infant utterances used in the practical exercises were too diverse and too scarce to be apt for reliably testing this type of weighing. Nevertheless, since the scheme covers the articulatory vowel chart quite well, it can be expected that it fits the complexity of vocalic movements and will be well applicable, especially in a period that children start to learn phonological contrasts as well. In future work we will try to apply and test this weighing scheme on existing databases.

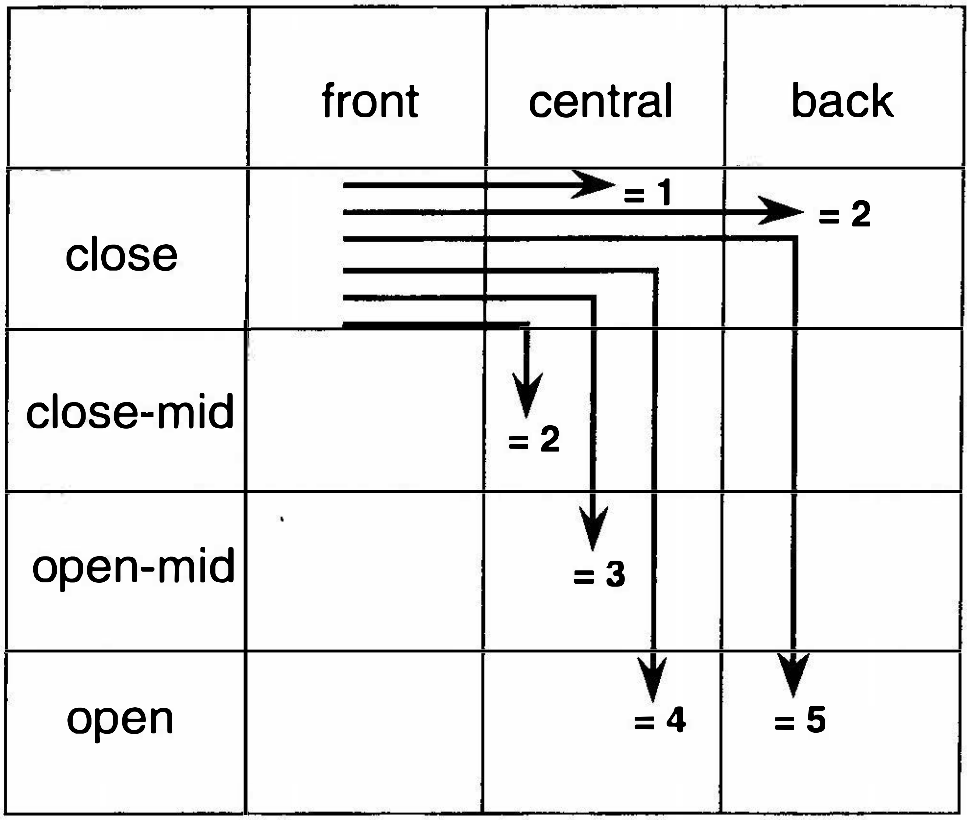

Fig. 4. Tentative weighing scheme for movements within vocalic parts of infant utterances.

5. Transcription of multisyllabic utterances

The question of how to transcribe multisyllabic utterances with multiple articulatory movements (Art 2) becomes more and more relevant in the second half of the first year when children start babbling. Especially in the case of variegated babbles transcription might be rather complicated.

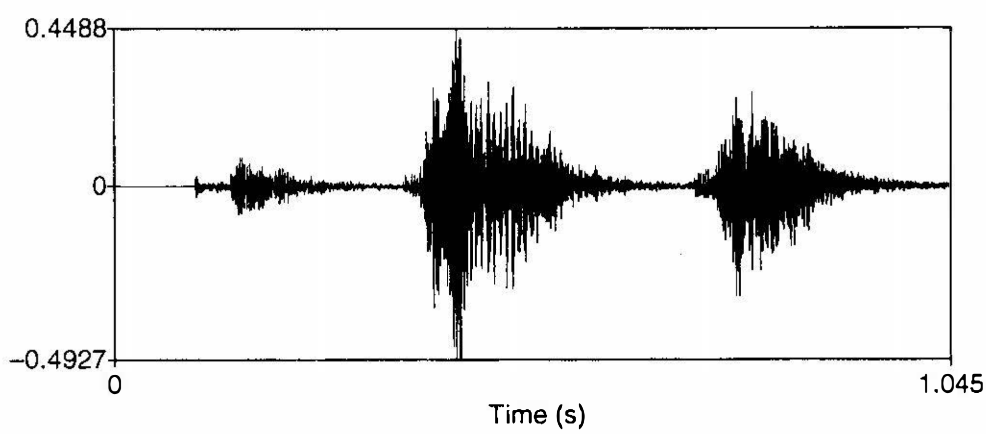

Fig. 5. Variegated babble utterance of an eight-months-old normally hearing boy.

Table 4. Tentative classification sheet for the transcription of multisyllabic utterances with multiple articulatory movements (Art 2). The description of a babbled utterance is given as an example. For further explanation see text.

utterance	rh8h32		
• number of syllables	3		
• phonation type	VarIntPhon (Phon 4)		
• articulation type	TwoArt (Art 2)		
• articulation position	1) middle	2) middle	
• place of articulation	1) central	2) front	
• manner of articulation	1) stop	2) fricative	
• vocoid initial	1) close-mid/central	2) open/central	3) close/central
• vocoid final	1) close-mid/central	2) open/central	3) open-mid/front
• comments			

To cover the various aspects in the transcription of such an utterance a classification sheet is developed as presented in Table 4. A variegated babble of a normally hearing boy at the age of eight months is used here as an example (Fig. 5).

First of all the code number of the utterance is registered *(rh8h32)*. Next the number of syllables is counted *(3)*, the phonation type *(VarIntPhon)* and the articulation type *(TwoArt)* are classified, both over the whole utterance. Since the utterance exists of more than one syllable the next items for each of the articulations and for each of the vocoids have to be transcribed separately. The utterance starts with a vocalic part *(non-changing close-mid/central)* followed by an articulatory movement *(central stop)*, a vocalic part *(non-changing open-mid/central)*, an articulatory movement *(front fricative)*, and ends in a vocalic part *(changing from close/central to open-mid/front)*. By indicating vocoid initial and vocoid final the possibility has been created to indicate dynamicity in the vocalic part for each of the vocoids occurring in the utterance. Finally a possibility is provided to give comments if needed.

As said before (p. 94) strong variations in intensity and/or intonation sometimes make a reliable syllable count quite difficult. Then visualization of the waveform and of the pitch contour by means of a good speech editing program like PRAAT (Boersma & Weenink, 1996), turns out to be of great help.

6. General conclusion

For both teacher and students the intensive transcription course has been a very positive experience. Training a small group of international Ph.D.-students during almost two weeks and discussing problems and possible solutions together was a fruitful way to test and to improve the usefulness of the AMSTIVOC. Our future work will follow one of the suggestions of the students: to make a webpage for users of the system on our institute's website. This will allow a discussion forum for system users, sharing of new ideas of how to work with the system, and maintaining consistency in the way the system is used. Making the whole course available on internet is another option.

Acknowledgements

The author wishes to express her gratitude to Kristina Zajdo, Stefanie Brosda, and Caroline Pickstone, the enthusiastic Ph.D.-students who attended the transcription course. They all contributed fresh ideas and useful suggestions to improve our AMSTIVOC.

References

Boersma, P. & Weenink, D. (1996): *Praat, a system for doing phonetics by computer.* Institute of Phonetic Sciences, University of Amsterdam, report 132. [see also http://www.fon.hum.uva.nl/praat/ for the present version].

Clement, C.J. (to appear): Ph.D.-thesis University of Amsterdam.

Clement, C.J. & Koopmans-van Beinum, F.J. (1995): "Influence of lack of auditory feedback: vocalizations of deaf and hearing infants compared", *Proceedings of the Institute of Phonetic Sciences Amsterdam* 19, 25-37.

Clement, C.J. & Koopmans-van Beinum, F.J. (1999): "Characteristics of vocalizations of deaf and hearing infants in the first year of life", *Proceedings ICPhS'99.* San Francisco, USA, Vol. 3, 1929-1932.

Clement, C.J., den Os, E.A. & Koopmans-van Beinum, F.J. (1995): "The development of early vocalizations of deaf and normally hearing infants in the first eight months of life", *Proceedings ICPhS'95.* Stockholm, Sweden, Vol. 1, 138-141.

Davis, B.L. & MacNeilage, P.F. (1995): "The articulatory basis of babbling", *Journal of Speech and Hearing Research* 38, 1199-1211.

Jansonius-Schultheiss, K. (1999): *Twee jaar spraak en taal bij schisis* (Two years of speech and language in Cleft Palate Children, with an English summary), Ph.D.-thesis, University of Amsterdam, IFOTT / LOT 17.

Koopmans-van Beinum, F.J. & Stelt, J.M. van der (1986): "Early stages in the development of speech movements", In: B. Lindblom & R. Zetterström (Eds.), *Precursors of Early Speech.* Wenner Gren. Int. Symp. Series 44, New York: Stockton Press, 37-50.

Koopmans-van Beinum, F.J. & van der Stelt, J.M. (1998): "Early speech development in children acquiring Dutch. Mastering general basic elements", In: S. Gillis & A. de Houwer (Eds), *The Acquisition of Dutch.* Amsterdam/Philadelphia: John Benjamins, 101-162.

Koopmans-van Beinum, F.J. (1990): "The source-filter model of speech production applied to early speech development", *Proceedings of the International Conference of Spoken Language Processing 1990,* Kobe, Japan, Vol. 2, 1321-1324.

Koopmans-van Beinum, F.J. (1993): "Cyclic effects of infant speech perception, early sound production, and maternal speech", *Proceedings of the Institute of Phonetic Sciences Amsterdam* 17: 65-78.

Koopmans-van Beinum, F.J., Clement, C.J. & van den Dikkenberg-Pot, I. (2000): "Babbling and the lack of auditory speech perception: a matter of coordination?", To appear in: *Developmental Science.*

Koopmans-van Beinum, F.J., Jansonius-Schultheiss, K. & Van der Stelt, J.M. (1990): "De invloed van een gehemelteplaatje op de spraakontwikkeling en -interactie in de eerste levensjaren van baby's met een gehemeltespleet", Final Report to the Dutch Prevention Fund, *Report 110 of the Institute of Phonetic Sciences Amsterdam.*

Oller, D.K. (1980): "The emergence of the sounds of speech in infancy", In: G.H. Yeni-Komshian, J. Kavanagh & C.A. Ferguson (Eds), *Child Phonology: Vol. I: Production.* New York: Academic Press, 93-112.

Oller, D.K. (1986): "Metaphonology and infant vocalizations", In: B. Lindblom & R. Zetterström (Eds.), *Precursors of Early Speech.* Wenner Gren. Int. Symp. Series 44, New York: Stockton Press, 21-35.

Stark, R.E. (1980): "Stages of speech development in the first year of life", In: G.H. Yeni-Komshian, J. Kavanagh & C.A. Ferguson (Eds), *Child Phonology: Vol. I: Production.* New York: Academic Press, 73-92.

Stark, R.E., Rose, S.N. & McLagen, M. (1975): "Features of infant sounds: the first eight weeks of life", *Journal of Child Language* 2, 205-221.
Van den Dikkenberg-Pot, I., Clement, C.J. & Koopmans-van Beinum, F.J. (1999): "Influence of lack of auditory speech perception on sound productions of deaf infants", In: H. Schutte et al. (Eds), *Proceedings of the 24th IALP Congress Amsterdam,* Vol. II, 924-927.
Van den Dikkenberg-Pot, I., Koopmans-van Beinum, F.J. & Clement, C.J. (1998): "Influence of lack of auditory speech perception on sound productions of deaf infants" (extended version), *Proceedings of the Institute of Phonetic Sciences Amsterdam* 22, 47-60.
Van der Stelt, J.M. (1993): *Finally a word. A sensori-motor approach of the mother-infant system in its development towards speech.* Ph.D.-thesis, University of Amsterdam, Amsterdam: IFOTT Studies in Language and Language Use 4.