Lund University, Centre for Languages & Literature, Dept. of Linguistics & Phonetics
Working Papers 52 (2006), 1–4

Acoustic Analysis of Phonetically Transcribed Initial Sounds in Babbling Sequences from Infants with and without Cleft Palate

Emilia Ahlberg, Julia Backman, Josefin Hansson, Maria Olsson, and Anette Lohmander
Institute of Neuroscience and Physiology/Speech Pathology,
Sahlgrenska Academy at Göteborg University
{gusahemi|gusbacju|gushanjos}@student.gu.se, kanelstang@hotmail.com,
anette.lohmander@logopedi.gu.se

Abstract
The aim of this study was to compare acoustic analysis of initial sound in babbling with corresponding phonetic transcription. Two speech pathologists had transcribed eight babbling sequences with total disagreement about whether the initial phoneme was a vowel or a plosive. After discussion, however, a consensus judgment was reached. To decide about the initial phoneme, an acoustic analysis was performed. Because of the deficient quality of some of the recordings, the results of the acoustic analysis were not completely reliable. However, the results were in relatively good agreement with the consensus judgments and indicate that the two methods should be used as complements to each other.

1 Background and introduction
Perceptual judgment is most commonly used in clinical practice of speech pathology. However, what a human listener perceives is dependent of many different components, among these a listener's expectations have great importance. Categorical perception signifies the tendency of humans to classify speech sounds into different categories (Lieberman and Blumstein 1988; Reisberg 2001). This may cause various listeners to sort a phone that is in between categories in different phonemic groups. In addition, perception is always subjective. Training with direct feedback is considered to be of great importance to create a common frame of reference for two different judges and to increase the perceptual awareness (Shriberg, 1972).

Children with cleft palate often have difficulties attaining a sufficient intra-oral pressure when producing plosives. In babbling, the phonetic features of the speech sounds are also less distinctive than in later speech due to maturation. In the acoustic analysis plosives may therefore appear with a less distinct mark in the spectrogram.

The purpose of this study was to examine the agreement between perceptual and acoustic analysis in judging the initial sound of babbling sequences of infants with and without cleft palate.

2 Method

2.1 Material

Babbling sequences at 18 months of age from 41 children with and without cleft palate had been phonetically transcribed by two listeners. The transcriptions were made independently, but in close connection with the transcriptions a consensus judgment was made. Consensus rules had been designed in advance to keep the judgments as equal as possible. Eight of these independent transcriptions by the two listeners had total disagreement regarding the initial phoneme. These as well as the consensus transcription were chosen for acoustic analysis.

2.2 Acoustic analysis

Acoustic analysis of the eight babbling sequences was made to establish the initial sounds (vowel or plosive). The judges were unaware of whether the children had a cleft palate or not. The analysis was made using the software Praat (Boersma & Weenink, 2005). The intension to be able to gather specific signs of a plosive had to be complemented by more subtle features. In the analysis the formant transitions and the intensity (of blackness) were observed in detail. Reduction of noise and in some cases filtering in frequency was also made.

2.3 Statistical method

To compare data from the judgments made by perceptual and acoustic analysis, the statistics software SPSS was used. The agreement between the different methods, as well as between the judges, was calculated using Cohen's Kappa.

3 Results and discussion

The acoustic analyses are presented in Table 1. In Table 2 the results from the acoustic and perceptual (phonetic transcriptions) analyses are presented.

Table 1. Description of the acoustic analysis of the initial sound from eight babbling sequences.

1.	**Plosive**	*Possibly approximant since there is no obvious burst.*
2.	**Vowel**	Vowel. Woman and child are speaking at the same time. The woman however is using "baby talk", meaning her F0 is higher and she is thereby sounding more like the child. Still, no natural formant-transitions for plosives are seen in the utterance.
3.	**Plosive**	Rather distinct burst, where formant-transitions are seen.
4.	**Plosive**	Distinct burst and formant-transitions are seen.
5.	**Plosive**	Obvious weakening of the formants initially, which implies a more closed mouth and could indicate a plosive.
6.	**Plosive**	At a detailed level a trace of a plosive is seen. Could also be an approximant? The result is very uncertain since an adult and the child is speaking at the same time.
7.	**Plosive**	A small formant-transition appears after a filtering of noise. Weak formants that could signify a nasal vowel or a more closed mouth are observed. Click/Bang sound is heard at the exact same time as the possible plosive. Uncertain result.
8.	**Vowel**	Nothing in the acoustic analysis indicates a plosive. 89

Table 2. The results of the three perceptual judgments (judge 1, judge 2 and consensus judgment) and acoustic analysis. P = plosive, V = vowel.

Child	*CP/Control*	*Judge 1*	*Judge 2*	*Consensus*	*Acoustic analysis*
1	CP	C	V	C	C
2	CP	V	C	V	V
3	Control	C	V	C	C
4	Control	C	V	C	C
5	Control	C	V	C	C
6	CP	V	C	C	C
7	Control	C	V	V	C
8	Control	V	C	V	V

A calculation of agreement between the consensus judgment and the acoustic analysis resulted in a Kappa value of **.71** (>.75 is considered good concordance) (Table 3).

Table 3. The values for agreement between acoustic analysis and the different listener conditions. Cons = consensus judgment, J1 = judge 1 and J2 = judge 2, Negative value = disagreement.

Judgments	*Cohen's Kappa*
Acous – Cons	.71
J1 – J2	-.88
Acous – J1	.71
Acous – J2	-.56
Cons – J1	.47
Cons – J2	-.41

4 Conclusions

According to the statistical analysis the agreement between the acoustic analysis and the consensus judgment is relatively good. Even though the Kappa value is .71, seven out of the eight judgments were equivalent. The limited amount of samples makes it difficult to draw conclusions. However, the results show that the consensus judgment was reliable. It also implies that judge 1 had better agreement with the acoustic analysis than judge 2. In fact judge 1 had as good agreement with the acoustic analysis as the consensus judgement but based on different samples. According to the acoustic analysis some initial sounds appeared more like approximants than plosives, which can explain the uncertainty in the perceptual judgments. Since there were only two possible options (plosive or vowel) in the acoustic analysis no consideration was taken to the approximantic signs in the spectrogram. This also makes the perceptual judgments less reliable as well as the results of the concordance between the perceptual judgments and the acoustic analysis.

Three out of the eight babbling sequences were produced by children with cleft palate. Children with cleft palate have difficulties building up a sufficient intra-oral pressure, which can result in a less distinct burst in the spectrogram. In the acoustic analysis two specific sounds were interpreted as plosives (possibly approximants). These sounds were produced by children with CP, which explain the approximantic appearances.

Since the result of the acoustic analysis is interpreted by a human, it is a subjective judgment. The anatomy of the speaker, the design of the room and the recording method are examples of things that are of importance for the analysis. In theory, the acoustic signs are well described, but are difficult to interpret when used clinically. Interfering sounds and noise

in the spectrogram are difficult to separate from the babbling. This is also important for the results of the perceptual judgments and could explain disagreement between perceptual judges.

The fact that the judges in this study are disagreeing is not unique. In a study by Shriberg (1972), conclusion are drawn that training with a key is of importance for the concordance between judges. In this study the judges had experience of transcribing together but without direct feedback or a key.

In conclusion, the results from this study show that neither the perceptual nor the acoustic judgment, gave reliable answers. However, one could assume that these both methods complement each other. In order to increase the reliability, both perceptually and acoustically, it is important that the recordings are of sufficient quality. This can be achieved by using high quality equipment, carefully consider the placing of the microphone and by using a designed recording room where the risk for disturbing sounds is minimal (for example by using soft toys when recording children). To get a more valid judgment it is also suggested to exclude utterances where competing sounds cannot be avoided.

References

Boersma, P. & D. Weenink, 2005. *Praat: doing phonetics by computer* (Version 4.3.33) [Computer program] Retrieved October 7, 2005, from http://www.praat.org/.

Lieberman, P. & S.E. Blumstein, 1988. *Speech Physiology. Speech Perception and Acoustic Phonetics*. Cambridge: Cambridge University Press.

Lindblad, P., 1998. *Talets Akustik och Perception*. Kompendium, Göteborgs Universitet.

Reisberg, D., 2001. *Cognition – Exploring the Science of the Mind*. New York: W.W. Norton & Company, Inc.

Shriberg, L.D., 1972. Articulation Judgments: Some Perceptual Considerations. *Journal of Speech and Hearing Research 15*, 876-882.

Lund University, Centre for Languages & Literature, Dept. of Linguistics & Phonetics
Working Papers 52 (2006), 5–8

Perception of South Swedish Word Accents

Gilbert Ambrazaitis and Gösta Bruce
Dept. of Linguistics and Phonetics, Centre for Languages and Literature, Lund University
{Gilbert.Ambrazaitis|Gosta.Bruce}@ling.lu.se

Abstract

A perceptual experiment concerning South Swedish word accents (accent I, accent II) is described. By means of editing and resynthesis techniques the F0 pattern of a test word in a phrase context has been systematically manipulated: initial rise (glide vs. jump) and final concatenation (6 timing degrees of the accentual fall). The results indicate that both a gliding rise and a late fall seem necessary for the perception of accent II, while there appear to be no such specific, necessary cues for the perception of accent I.

1 Introduction

In the original Swedish intonation model (Bruce & Gårding, 1978) the two tonal word accents (accent I and accent II) are assigned bitonal representations in terms of High plus Low (HL), representing the accentual F0 fall. These Highs and Lows are timed differently, however, in relation to the stressed syllable depending on dialect type. For all dialect types, the HL of accent I precedes the HL of accent II. In South Swedish, the HL of accent I is aligned with the stressed syllable, while the HL of accent II is instead aligned with the post-stress syllable.

A problem with the latter representation is that the stressed syllable in accent II words has no direct tonal representation. Thus this modelling does not reflect what should be the most perceptually salient part of the pitch pattern of accent II. Figure 1 shows prototypical F0 contours of the two word accents (minimal pair) in a prominent position of an utterance as produced by a male speaker of South Swedish (the second author).

This particular problem of intonational modelling has been the starting-point of a phonetic experiment aimed at examining what is perceptually relevant in the F0 contours of accent I and accent II in the South Swedish dialect type. More specifically, our plan has been to run a perceptual experiment, where the intention was to find out what are the necessary and sufficient cues for the identification of both word accents.

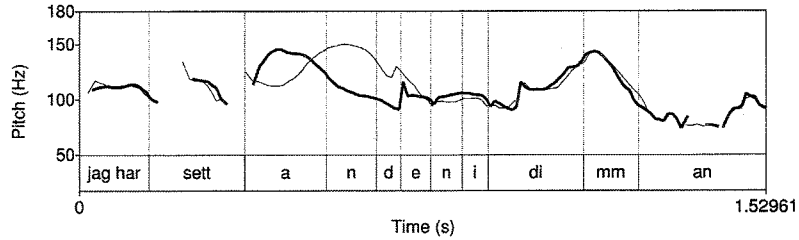

Figure 1. Prototypical F0 contours of the two word accents in a prominent position of an utterance as produced by a male speaker of South Swedish: *Jag har sett anden i dimman. ('I have seen the duck/spirit in the fog.')* Thick line: acc. I *('duck')*; thin line: acc. II *('spirit')*.