Nature's distributional-learning experiment: Infants' input, infants' perception, and computational modeling

Benders, A.T.

Publication date
2013

2

ALL MOMMY DOES IS SMILE! DUTCH MOTHERS' REALIZATION OF SPEECH SOUNDS IN INFANT-DIRECTED SPEECH EXPRESSES AFFECT, NOT DIDACTIC INTENT

An adapted version of this chapter is:
Benders, T. (under review).

Abstract

Exaggeration of the vowel space in infant-directed speech (IDS) is well documented for English, but not consistently replicated in other languages. A second attested pattern of change in IDS, which has received little attention in the literature, is an overall rise of the formant frequencies. The present study investigates longitudinally how Dutch mothers change their corner vowels /i/, /u/, /aː/, and /ɑ/, the fricative /s/, and their pitch when speaking to their infants at 11 and 15 months of age. Dutch mothers were found to raise the second formant (F2) of their vowels in IDS in comparison to adult-directed speech (ADS), especially of the back vowels. As a result, the vowel space became *smaller* in IDS than in ADS. Together with the raised spectral frequency of /s/ in IDS and the observation that F2 is raised more strongly for infants at 11 than at 15 months, these results show that smiling and enhanced positive affect are the main factors influencing Dutch mothers' realization of speech sounds in IDS. This study provides evidence that mothers' expression of emotion in IDS can influence the realization of speech sounds at the cost of speech clarity.

2.1 Introduction

Caregivers from most cultures use a different speech register for babies than for other adults (see for reviews Ferguson, 1977; Cruttenden, 1994; Soderstrom, 2007). This special way of speaking to an infant expresses positive emotions and maintains the infant's attention, but it also conveys the structure of the language (Ferguson, 1977; Fernald et al., 1989; Uther et al., 2007). Caregivers' positive affect is mostly carried by the pitch characteristics of infant-directed speech (IDS, Uther et al., 2007; Trainor et al., 2000). One linguistic aspect that caregivers from many languages seem to clarify in IDS as compared to adult-directed speech (ADS) is the auditory contrast between the corner vowels[1] (Bernstein Ratner, 1984; Kuhl et al., 1997; Burnham et al., 2002; Uther et al., 2007; Andruski et al., 1999; Liu et al., 2003). Uther et al. (2007) have claimed that the different realizations of speech sounds in IDS occurs independently of caregivers' affect. In the present paper I challenge the proposed dichotomy between didactic changes to the speech sounds and affective changes to the pitch, and test the hypothesis that the expression of affect is the main determinant of caregivers' realization of speech sounds in IDS.

2.1.1 *Didactic vowel space enhancement in IDS*

Enhanced auditory contrast between the corner vowels provides infants with clear examples of their native language's phoneme categories and is related to overall intelligibility and possibly to more precise articulations (Bradlow et al., 1996). It has been hypothesized that mothers enhance speech sound contrasts out of didactic consideration of their language-learning infant, because they similarly enhance their corner vowels in speech to adults learning a second language (Uther et al., 2007), but not in speech to pets (Burnham et al., 2002; but see Kim et al., 2006). Since mothers' enhancement of the vowel space in IDS is related to their infants' faster development of language-specific phoneme perception, these clear pronunciations in IDS may indeed promote infant language acquisition (Liu et al., 2003).

The occurrence of such vowel enhancement in English, Swedish, Russian, Japanese, and Mandarin IDS has led to the claim that it is a universal characteristic of IDS (Kuhl et al., 1997; Uther et al., 2007). However, the expansion of the vowel space in IDS is not found consistently across studies of American English (Green et al., 2010) and not found in all languages (Dodane and Al-Tamimi, 2007; Englund and Behne, 2006; Van de Weijer, 2001). In Norwegian, the vowel space is

1 The corner vowels are the vowels produced with the most extreme articulations physically possible. For most languages they are /i/ as in English *sheep*, /u/ as in English *shoe*, and one or two low vowel such as /ɑ/ as in English *shark*, or /æ/ as in English *sand*.

crucially *smaller* in IDS than in ADS (Englund and Behne, 2006). Other evidence against the universality of clear speech in IDS is the dependence of the infant-directed vowel space on infant characteristics: Infants that cannot hear their mother as a result of actual or simulated deafness receive less clear input than normally hearing infants (Lam and Kitamura, 2010, 2012). Consequently, the expansion of the vowel space in IDS cannot be considered a universal characteristic of IDS.

2.1.2 *Affective vowel formant increase in IDS*

A second attested pattern of change in infant-directed vowels is an overall increase of formant frequencies (Dodane and Al-Tamimi, 2007; Englund and Behne, 2005; Green et al., 2010). Formant frequencies depend on the shape and size of the vocal tract. An increase in formant frequencies results from a shortening of the vocal tract, which occurs when the lips are retracted for a smile (Tartter, 1980; Tartter and Braun, 1994; Waaramaa et al., 2008; Zacher and Niemitz, 2003; cf. Fagel, 2010, showing that the acoustic effect of smiling is vowel-dependent, and Aubergé and Cathiard, 2003, suggesting that formants are lower in speech with amused smiles). A very joyful smile is the predominant facial expression in interactions with infants (Stern, 1974; Chong et al., 2003). High formant frequencies of infant-directed vowels could well be a side effect of smiling (Englund and Behne, 2005), and as such a result of caregivers' enhanced positive affect when they speak to their infant.

There is a widespread consensus that the main acoustic vehicle of caregivers' positive affect in IDS is their pitch (Uther et al., 2007; Trainor et al., 2000). Cross-linguistically, caregivers use a higher average F0 (fundamental frequency, the main acoustic correlate of pitch) and a larger F0 range when speaking to their baby (Fernald et al., 1989). These infant-directed pitch modifications resemble those in emotional ADS (Trainor et al., 2000). Similar affective pitch changes are found in speech to pets (Burnham et al., 2002), but not in speech to foreigners Biersack et al. (2005); Uther et al. (2007). The affective pitch modifications are especially large in American-English IDS (Grieser and Kuhl, 1988; Fernald et al., 1989; Papoušek et al., 1991), while speakers of other languages mainly employ other means to express their positive affect in IDS. For example, Japanese mothers have a relatively restricted F0 range in IDS, which may be related to cultural restrictions on the vocal expression of emotions (Fernald et al., 1989). But Japanese mothers do establish emotional communication in IDS with attentional nonsense words and onomatopoeia (Toda et al., 1990; Fernald and Morikawa, 1993; Bornstein et al., 1992). Kitamura et al. (2002) argue that Thai mothers have a smaller pitch increase in IDS than English-speaking mothers, but express more positive affect in the content of their IDS. Possibly, raised formant frequencies as a re-

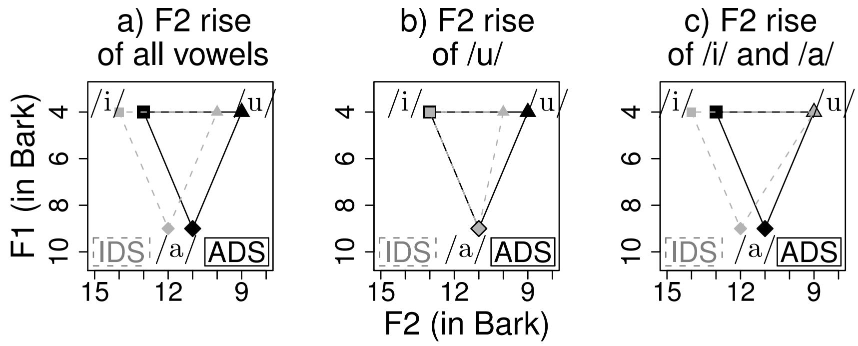

Figure 2: **Three possible interrelations between the height of F2 and the size of the vowel space in IDS as compared to ADS.** The vowel spaces are defined by F1 and F2 in Bark, with hypothetical vowel triangles encompassing /i/, /a/, and /u/ in ADS (black, solid line) and IDS (gray, dotted line). See the text for details.

sult of positive affect and smiling are yet another carrier of positive affect in the IDS of some languages. This possibility is important to explore, because the exact acoustic vehicles of positive affect in the voice quality are still unknown (Scherer, 2003).

2.1.3 *Testing didactic and affective changes in Dutch IDS*

The 'didactic' enhancement of the vowel space in IDS and 'affective' rise of the vowel formant frequencies in IDS are interdependent, as an increase of the formant frequencies in IDS can influence the size of the vowel space in various ways. When the first and second formant (F1 and F2) are raised equally along the auditory scale across the three corner vowels in IDS, the vowel space shifts without a change in the auditory contrast between the vowels. This is illustrated for a change in F2 in Figure 2a (cf. Dodane and Al-Tamimi, 2007; Green et al., 2010). A smaller vowel space in IDS can occur if F2 of /u/ is raised more in IDS than F2 of /i/ and /a/ (Figure 2, cf. Englund and Behne, 2006, 2005). A larger vowel space in IDS can be the consequence of a smaller F2-raise of /u/ than of /i/ and /a/ (Figure 2c, cf. the figures in Burnham et al., 2002; Kuhl et al., 1997). No study to date has investigated the size of the vowel space as well as the rise of the formant frequencies in IDS, so that the dependence of the size of the vowel space on the raising of the formant frequencies has remained largely unnoticed.[2] the absence of such a relation is crucial to Uther et al.'s (2007) claim that the size of the vowel space in IDS is the result of caregivers' linguistic-didactic efforts and not of their affect.

2 Englund and Behne investigate the rise of the formant frequencies in Englund and Behne (2005) and the size of the vowel space in Englund and Behne (2006), but do not directly relate the two.

For several reasons, Dutch is an interesting language to investigate in this respect. In the first place, if enhancement of the vowel space is a (near)-universal property of IDS, it should occur in the vast majority of the languages. However, Van de Weijer (2001) did not find consistent enhancement of the vowel space in IDS in a corpus consisting of one Dutch infant's input from three speakers: the mother, the father and the babysitter. Interestingly, the (native German but Dutch-speaking) mother's vowel space was smaller in IDS than in ADS and inspection of the vowel spaces reported in Van de Weijer (2001) suggests that the mother and the babysitter increased F2 of their vowels in IDS. Therefore, the present study investigates the size of the vowel space as well as the heights of the formant frequencies in Dutch IDS.

To investigate whether the pronunciation of vowels in Dutch IDS is primarily 'didactic' or 'affective', it is useful to consider that didactic and affective changes in IDS follow different age-related trends. Mothers enhance speech sound contrasts somewhat more in the period after the child's first birthday than before, although it is not clear how long they maintain this extra enhancement (Bernstein Ratner, 1984; Malsheen, 1980; Cristiá, 2010; see also Liu et al., 2003, 2009). The infant-directed pitch changes, on the other hand, become less pronounced over the course of the first year and thereafter (Stern et al., 1983; Amano et al., 2006; Warren-Leubecker and Bohannon, 1984; Stern et al., 1983; Amano et al., 2006; Garnica, 1977; Remick, 1976; but see Jacobson et al., 1983). Also the content of IDS becomes less affective when the infant grows older, as caregivers start speaking more about events in the outside world (Snow, 1977; Sherrod et al., 1978; Penman et al., 1983; Bornstein et al., 1992).

These developmental changes suggest that mothers trade affective speech for linguistic-didactic speech when their child enters the second year of life (Kitamura et al., 2002). If Dutch mothers didactically enhance their vowel space in IDS without raising the formant frequencies, they are expected to enhance their vowel space more to infants that are over one year of age than to infants who are just under one year of age. If Dutch mothers' infant-directed vowel space is primarily characterized by an affective increase of the formant frequencies, they are expected to change their vowels more to infants under one year of age than to older infants. The present study investigates longitudinal changes in Dutch IDS at two time points, when infants are 11 months of age and when they are 15 months of age.

An alternative to the smiling hypothesis of raised formant frequencies in IDS (Englund and Behne, 2005) is that the raised formant frequencies result from caregivers' attempts to imitate their infant (Dodane and Al-Tamimi, 2007). Infants have a smaller vocal tract than adults and therefore they produce their vowels with overall higher formant frequencies (Peterson and Barney, 1952). An investigation of the realization of /s/ in IDS can help to determine whether smiling or the imitation of children's speech leads to raised formant frequencies in IDS. In emotional speech, adult speakers realize /s/ with spectral energy on higher frequencies than in emotionally neutral speech (Kienast and Sendlmeier, 2000). Children realize /s/ with most spectral energy on lower frequencies than adults (Nissen and Fox, 2005). If mothers raise their formant frequencies in IDS, the smiling hypothesis and the imitation hypothesis for raised formant frequencies in IDS provide competing predictions with respect to mothers' realization of /s/. A single-parameter measure of the concentration of the energy distribution in a fricative is the center of gravity (COG, also spectral mean or first spectral moment, Forrest et al., 1988). Although there are many other acoustic parameters to fricatives (Jongman et al., 2000), it COG that differs between adult and child speech (Nissen and Fox, 2005) and is related to smiling (Kienast and Sendlmeier, 2000). Therefore, the present study investigates the COG of /s/ in Dutch IDS, in addition to the vowels.

2.1.4 *Summary of study objectives*

To summarize, in addition to being the first investigation of Dutch IDS with multiple mother-child dyads (cf. Van de Weijer, 2001, 1997), the present study is the first to investigate vocalic, consonantal, and prosodic modifications in IDS in the same group of mothers at two time points. The primary question is whether Dutch mothers change the size of their vowel space in IDS or shift their vowel space to higher formant frequencies in that register. If both patterns of change are found, the relation between the changed formant frequencies and the size of the vowel space will be determined. If the vowel space is shifted to higher formant frequencies, changes in infant-directed /s/ can help to interpret whether these formant changes are the result of smiling or of the imitation of children's speech. The second question is whether the speech sound and pitch characteristics of Dutch IDS change between the infants' age of 11 and 15 months. If age-related changes are found, the vowel changes may be primarily didactic or primarily affective. If raised formant frequencies are found, . By answering these questions, this study tests the claim that enhancement of the vowel space is a universal characteristic of IDS that results from mothers' attempts to clarify the structure of the language for their infant (Kuhl et al., 1997; Uther et al., 2007) and contrasts it with the hypothesis that, in some languages, the infant-directed vowels are characterized by raised formant frequencies, which result from affect.

2.2 Method

2.2.1 *Participants*

Eighteen mother-child dyads (6 boys, 12 girls) participated in this longitudinal study. Recordings were made when the child was 11 months of age (ranging from 311 to 352 days) and 15 months of age (ranging from 448 to 472 days). All children were born at a gestational age of at least 36 weeks and were from monolingual Dutch families. The mothers were native speakers of Dutch. Another 11 dyads had to be excluded from the analysis because the father instead of the mother came to the visits (n=2), an appointment for the second recording could not be scheduled (n=6), an older sibling interfered (n=1), or because of equipment failure (n=1) or experimenter error (n=1). Participants were recruited from a database that is maintained at the University of Amsterdam. All mothers gave written consent prior to participating in the study and afterwards received a small monetary compensation for their travel expenses and participation (€10).

2.2.2 *Procedure and Equipment*

Recordings took place in a sound-proofed studio. Recordings were made with an omni-directional head-mounted Samson QV microphone fitted to the mother and connected to the amplifier by a long cord to allow freedom of movement.[3] The stream was sampled at 44100 Hz and recorded together with a video recording of the scene using the program Enosoft DV Processor.

Prior to the recordings, mothers were told that the natural play interactions between mothers and children were the focus of the investigation. The mother and child were seated on the floor, on a blanket in a corner of the room. When the recording started, the experimenter first had a short conversation with the mother about the child's development in the past months. This introductory conversation was intended to make the participants feel at ease and the speech from this phase was not analyzed. Next the mother was given three bags with toys and instructed to unpack the bags with the child, name the toys for the child, and play with the toys. Mother and child were then left alone for approximately 10 minutes. After this period, the experimenter engaged the mother in a conversation about the play session to elicit the target words in an adult-directed register.[4]

---

3 To ensure that sessions were not lost due to contact between the mothers' face and the head-mounted microphone, parallel recordings were made with a free-standing Sennheiser HF condenser microphone MKH-105. These recordings proved not to be necessary.

4 The introductory adult-to-adult conversation was skipped if the child was impatient and for the first participants in the study. In that case, the experimenter and parent

Each bag contained items to elicit the vowels /i/, /u/, /aː/, and /ɑ/, and the fricative /s/ (Table 2). All items were selected so that they were either monosyllabic, or had the main stress on the first syllable.

Vowel	Items					
/i/	/fits/ fiets *bike*	/spixəl/ spiegel *mirror*	/xitər/ gieter *watering can*	/vlixtœyx/ vliegtuig *plane*		
/u/	/bukjə/ boekje *book*	/ku/ koe *cow*	/pus/ poes *cat*	/ɦut/ hoed *hat*		
/aː/	/sxaːp/ schaap *sheep*	/aːp/ aap *monkey*	/taːfəl/ tafel *table*	/kaːs/ kaas *cheese*		
/ɑ/	/tɑs/ tas *bag*	/bɑk/ bak *container*	/slɑk/ slak *snail*	/ɑpəl/ appel *apple*	/bɑt/ bad *bath*	/kɑst/ kast *cupboard*
/s/	/sinəzɑpəl/ sinaasappel *orange*					

Table 2: **The words for the stimuli used to elicit the four target vowels /i/, /u/, /aː/, and /ɑ/, and /s/.** For each word, the IPA transcription (row 1, between / /), Dutch spelling (row 2), and English translation (row 3, in italics) are given. Apart from the item that was only used to elicit /s/, the underlined items were also used to elicit /s/.

2.2.3 *Coding*

The sound recordings were isolated from the video and stored as WAV-files. Subsequent coding was done in Praat (Boersma and Weenink, 2011).

The coders were two undergraduate students, both native speakers of Dutch, with basic education in phonetics and specific training

talked about the child's development when the play session and the adult-to-adult conversation about the play session were completed.

in speech segmentation. They transcribed the phrases orthographically and marked the boundaries of the target words and the target vowels in the signal. The criteria for vowel segmentation were based on Machač and Skarnitzl (2009). In addition, the coders indicated to whom the mother was speaking (her child, the experimenter, or uncertain), noted external sounds that overlapped with the target vowels (such as the child or noise from the toys), and indicated atypical voice qualities. The fricative /s/ was segmented from the signal by a third coder. If part of the fricative overlapped with another sound, the coder segmented the non-overlapping part of the fricative.

2.2.4 *Acoustic measurements*

All acoustic analyses were conducted in Praat (Boersma and Weenink, 2011). Detailed information about the acoustic measurements is given in section 2.5

2.2.4.1 *Vowels*

The median F1 and F2 were measured in the central 40% of the vowels /i/, /u/, /aː/, and /ɑ/. Prior to the analyses, the formant values in hertz were converted to the psychoacoustic Bark scale (Zwicker, 1986) following the formula in Equation 2:

$$Bark(x) = 7\log\left(\frac{Hz(x)}{650} + \sqrt{1 + \frac{Hz(x)}{650}^2}\right) \tag{1}$$

2.2.4.2 *The fricative* /s/

The complete spectrum of each /s/-sound was high-pass filtered at 700 Hz in order to remove residual voiced parts from the signal prior to the analysis, and then COG was measured in the complete fricative.

2.2.4.3 *Pitch*

The median F0 of each phrase was measured in hertz (further: F0-median). The minimum and maximum F0 of each phrase were measured as well and the distance between these extremes in semitones (12 semitones = 1 octave) was divided by the utterance duration, yielding a measure of F0 excursions in semitones per second (further: F0-excursions, Fernald and Simon, 1984).

2.2.5 *Exclusion and Analyses*

Phrases and speech sounds were not included in the analyses for a number of reasons: if they overlapped with another sound; if the mother was singing, whispering, glottalizing, or had been using yet another voice quality that might have affected the acoustic measurements; if the coder considered the voice quality otherwise atypical; or if the coder was uncertain whether the infant or adult had been addressed. Phrases were also excluded if the coder indicated doubt about the transcription or if the analysis of F0 (see details in section 2.5) did not return a value. Table 3 gives the number of vowels, /s/s, and phrases in the full corpus and in the analyses.

The median was considered the appropriate measure of central tendency to summarize each mother's data because it is robust to outliers, which may occur due to incidental errors in the acoustical analyses. Medians were taken per mother for each measure (the formant values of the four vowels, the COG of /s/ and the pitch of the phrases), separating the speech addressed to her infant at 11 months (IDS-11), to her infant at 15 months (IDS-15) and the adult experimenter (ADS, collapsed over both time points). Per mother, the averages over IDS-11 and IDS-15 yielded the values for IDS. No value for IDS was computed if the value for either IDS-11 or IDS-15 was missing.

The area of each mother's vowel quadrilateral encompassing the high vowels /i/ and /u/ and the low vowels /aː/ and /ɑ/ was computed from the median formant values in Bark. The area was computed separately for ADS, IDS, IDS-11, and IDS-15.

In all analyses, ADS was compared to IDS in a first analysis, after which IDS-11 and IDS-15 were compared. As more subjects were excluded from the ADS condition than from either of the IDS conditions, the separation of the analyses on the register (IDS vs. ADS) from the analyses on the infants' age (IDS-11 vs. IDS-15) rendered the latter comparison more powerful.

A mother was excluded from the comparison between the registers, IDS vs. ADS, if she provided no useable tokens for either IDS-11, IDS-15, or ADS. A mother was excluded from the comparison between the infants' ages, IDS-11 vs. IDS-15, if she provided no useable tokens for either IDS-11 or IDS-15. Table 3 gives the number of mothers included in each of the two comparisons (Register and Infants' Age) of each of the three measured units (vowels, /s/, phrases).[5]

5 All mothers produced all vowels as well as /s/ in IDS-11, IDS-15, and ADS. The relatively large number of excluded participants in the comparison of the vowels across IDS and ADS is due to exclusion after the recordings, and to the fact that at least one useable token of all four vowels in both registers was required for a participant to be included in this comparison. A total of 10 participants is not uncommon in the research on vowel spaces in IDS, which has seen sample sizes of 10 to 14 subjects (Kuhl et al., 1997; Andruski et al., 1999; Burnham et al., 2002; Uther et al., 2007), as well as smaller (Bernstein Ratner, 1984; Englund and Behne, 2005, 2006) and considerably larger (Green et al., 2010; Cristiá, 2010) sample sizes. Three of the mothers who were excluded from the analysis of vowel quality on the basis of a lack of tokens in ADS were also the three mothers excluded from the comparison between IDS-11

	Tokens		Mothers included (max=18)	
	total	included	IDS–ADS	IDS-11–IDS-15
vowels	3263	1704	10	15
/s/s	1421	999	17	18
phrases	2816	1157	18	18

Table 3: **A summary of the content of the corpus.**The first two columns give the total number of tokens (vowels, /s/s, and phrases for analysis of pitch) in the corpus and the analysis. The second two columns give the number of mothers included in the comparison between IDS and ADS, and the comparison between IDS-11 and IDS-15.

For the statistical analyses of the formant frequencies of the vowels, repeated measures analyses of variance were performed. For the analyses of the area of the vowel space, the COG of /s/, and the pitch characteristics, paired-samples *t*-tests were performed. An alpha-level of 0.05 was adopted to evaluate the effects from these main analyses. Effects with $p < .1$ were interpreted as marginally significant in support of other effects. To further test significant effects from the ANOVAs, paired-samples *t*-tests were performed. For these comparisons a corrected alpha-level of 0.005 was adopted, but given the small sample size, effects with $p < .05$ were interpreted as marginally significant.

2.3 Results

2.3.1 *Vowel space: Area*

The auditory vowel space defined by F1 and F2 in Bark with the vowels /i/, /u/, /aː/, and/ɑ/ from IDS-11, IDS-15, and ADS is given in Figure 3.

A paired-samples *t*-test compared the area of the mothers' vowel space in IDS and ADS. The area of the Dutch mothers' vowel space was significantly and substantially *smaller* in IDS than in ADS ($t_9 = 3.22, p = .010$; IDS: $m = 12.4$, $sd = 2.35$; ADS: $m = 15.6$, $sd = 2.56$). Nine of the 10 mothers had a smaller vowel space in IDS. A second paired-samples *t*-test compared the area of the mothers' vowel space in IDS-11 and IDS-15. The areas of the vowel spaces in IDS-11 and IDS-15 did not differ significantly ($t_{14} = 0.04$, $p = .968$; IDS-11: $m = 12.5$, $sd = 2.89$; IDS-15: $m = 12.6$, $sd = 3.15$).

and IDS-15. The results may thus be slightly biased towards the mothers that were more talkative during the complete procedure.

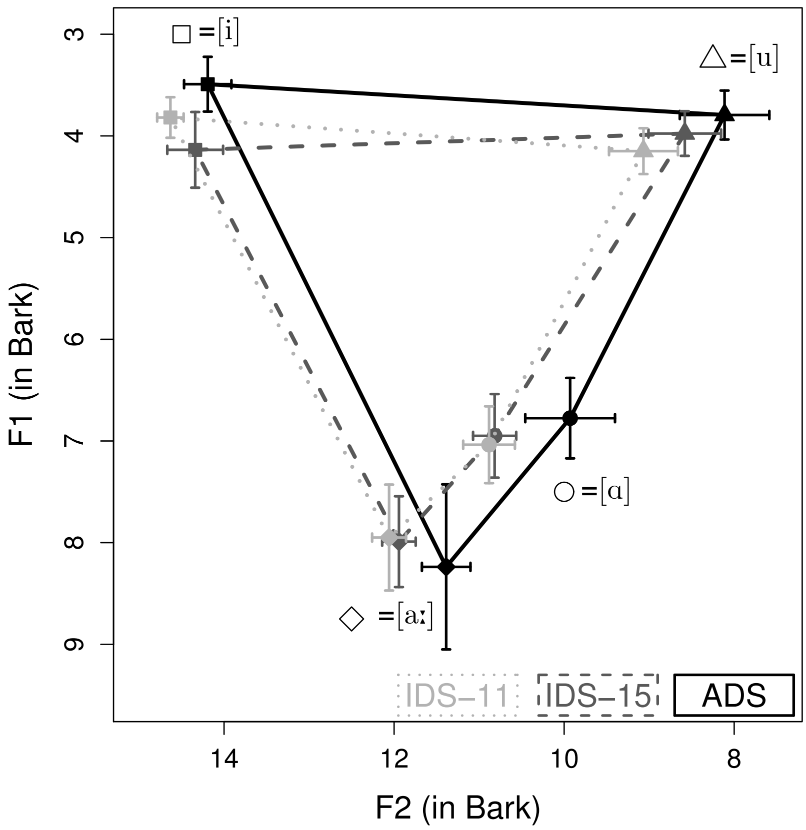

Figure 3: **The vowel space, defined by F1 and F2 in Bark, with the vowel quadruples encompassing /i/, /aː/, /ɑ/, and /u/ in IDS to 11-month-olds (light gray, dotted line), IDS to 15-month-olds (dark grey, dotted line), and ADS (black, solid line).** The filled figures represent the group means for the four vowels, with error bars showing 95% confidence intervals of the group means in F1 and F2. For ADS, the 10 participants from the comparison between ADS and IDS are included; for IDS, all 18 participants are included. See the text for further information on the measurements and computations.

2.3.2 *Vowel space: Formant frequencies*

The F1 and F2 values of the corner vowels were the dependent variables in two separate repeated measures ANOVAs with Subject as random factor, and Vowel Backness (front /i, aː/ vs. back /u, ɑ/), Vowel Height (high /i, u/ vs. low /a, ɑ/) and Register (IDS vs. ADS) as within-subject factors. In these analyses, the main effects of Vowel Backness and Vowel Height and their interaction were expected and not of interest, because it is well known that the four vowels have different formant values. The second set of analyses with Infants' Age instead of Register as within-subjects factor is reported below.

	Measure	Effect	F	df	p
IDS vs. ADS	F1	Register	0.78	1, 9	.401
		Vowel Height	701.66	1, 9	<.001
		Vowel Backness	35.23	1, 9	<.001
		R∗VH	3.42	1, 9	.098
		R∗VB	0.46	1, 9	.513
		VH∗VB	96.21	1, 9	<.001
		R∗VH∗VB	6.32	1, 9	.033
	F2	Register	49.02	1, 9	<.001
		Vowel Height	6.23	1, 9	.034
		Vowel Backness	712.69	1, 9	<.001
		R∗VH	5.89	1, 9	.038
		R∗VB	8.05	1, 9	.019
		VH∗VB	219.10	1, 9	<.001
		R∗VH∗VB	0.03	1, 9	.857
IDS-11 vs. IDS-15	F1	Infant's Age	0.03	1, 14	.862
		Vowel Height	1351.20	1, 14	<.001
		Vowel Backness	21.55	1, 14	<.001
		IA∗VH	0.15	1, 14	.705
		IA∗VB	2.00	1, 14	.180
		VH∗VB	24.47	1, 14	<.001
		IA∗VH∗VB	0.85	1, 14	.371
	F2	Infant's Age	10.27	1, 14	.006
		Vowel Height	6.74	1, 14	.021
		Vowel Backness	813.68	1, 14	<.001
		IA∗VH	4.13	1, 14	.062
		IA∗VB	0.16	1, 14	.692
		VH∗VB	377.74	1, 14	<.001
		IA∗VH∗VB	0.58	1, 14	.458

Table 4: **The results from four ANOVAs making the comparison between IDS and ADS and between IDS-11 and IDS-15 with respect to the F1 and F2 of the vowels /i/, /u/, /aː/, and /ɑ/.**

The results from the ANOVA on F1 are given in Table 4. There was a significant three-way interaction between Register, Vowel Backness, and Vowel Height ($F_{1,9} = 6.32, p = .033$). Paired-samples t-tests comparing for each vowel F1 in IDS and ADS showed a marginally significant increase of F1 for /i/ in IDS ($t_9 = 3.03, p = .014$), and this direction of the effect was found in 9 of the 10 mothers in the analysis. F1 was not significantly different between IDS and ADS for the three other vowels (all $t < 1.5$, all $p > .2$).

The ANOVA on F2 (see Table 4) revealed that the F2 difference between IDS and ADS was dependent on the backness of the vowel (significant Register∗Vowel Backness interaction: $F_{1,9} = 8.05, p = .019$), as well as on the height of the vowel (significant Register∗Vowel Height interaction: $F_{1,9} = 5.89, p = .038$).

The interactions between Register and on the one hand Vowel Backness and on the other hand Vowel Height showed that the F2 difference between IDS and ADS was not uniform across the vowel space. Because F2 differs between front and back vowels as well as between the high and low vowels investigated here, further investigations of these interactions required a measure of the F2 difference between IDS and IDS for each of the four vowels. F2-difference was computed for the four vowels separately as F2 in IDS minus the F2 in ADS. A F2-difference above 0 indicates a higher F2 of that vowel in IDS that in ADS and a F2-difference below 0 indicates a lower F2. F2-difference was above 0 for /i/ in 9 of 10 mothers, for /u/ in 9 of 10 mothers, for /aː/ in all 10 mothers, and for /ɑ/ in all 10 mothers. The average F2-difference was computed per mother for the high vowels (/i/ and /u/: $m = 0.439, sd = 0.4635$), the low vowels (/aː/ and /ɑ/: $m = 0.839, sd = 0.2963$), the front vowels (/i/ and /aː/: $m = 0.452, sd = 0.3623$), and the back vowels (/u/ and /ɑ/: $m = 0.826, sd = 0.3499$). To further assess the Register∗Vowel Backness interaction from the ANOVA, the F2-difference of the front and back vowels was compared in a paired-samples t-test. The F2-difference was found to be marginally larger in the back than in the front vowels ($t_9 = 2.84, p = .019$). Nine of the 10 mothers raised F2 more in the back vowels than in the front vowels. The auditory contrast in F2 between the front and back vowels is thus reduced in IDS. To further investigate the Register∗Vowel Height interaction, the F2-difference was compared between the high and low vowels in a paired-samples t-test. This test showed that the F2-difference was marginally larger in the low than in the high vowels ($t_9 = 2.43, p = .038$). Eight of the 10 mothers had a larger F2-difference in the low vowels than in the high vowels.

A second ANOVA on F1 with the within-subject factor Infant's Age (IDS-11 vs. IDS-15) showed no significant effects of interest (see Table 4). A second ANOVA on F2 with the within-subject factor of Infant's Age (IDS-11 vs. IDS-15) is reported in Table 4 and showed that F2 differed between speech addressed to the infants at 11 and 15

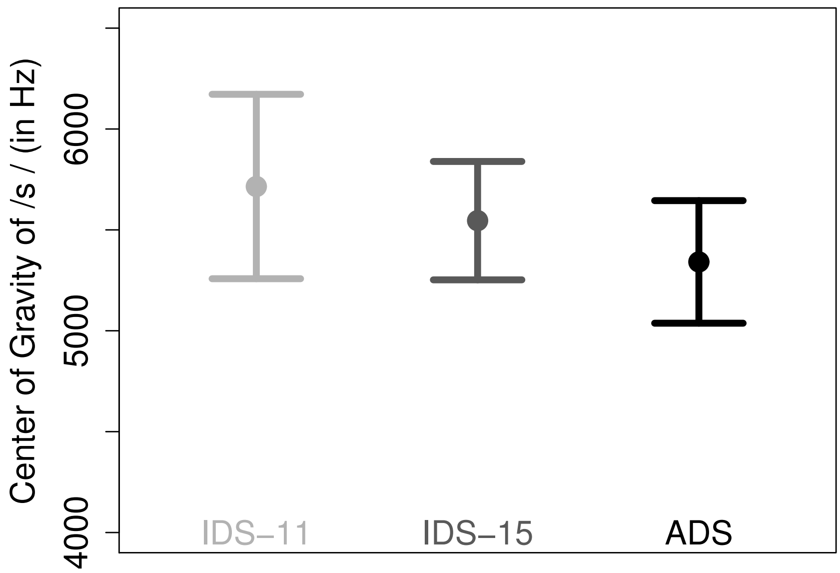

Figure 4: **The mean COG of /s/ in IDS to infants at 11 months (light grey), IDS to infants at 15 months (dark gray), and ADS (black).** The circles represent the group means, with error bars showing the 95% confidence intervals of the group means. Only participants included in the comparison between ADS and IDS are included in the figure. See the text for information on the measurements and further computations.

months (main effect of Infants' Age: $F_{1,14} = 10.27, p = .006$). From Figure 3, it can be observed that F2 was on average higher when mothers addressed their child at 11 months than when they spoke to the child at 15 months. F2 was higher in IDS-11 than in IDS-15 for /i/ in 12 mothers, for /u/ in 11 mothers, for /aː/ in 9 mothers, and for /ɑ/ in 8 mothers.

2.3.3 *The fricative* /s/

Figure 4 displays the average COG of /s/ in ADS, IDS-11, and IDS-15. Prior to the statistical analysis, the data of the COG of /s/ were square-root transformed to solve considerable skewness in the distributions. A paired-samples *t*-test comparing the square-root transformed COG of /s/ in IDS and ADS showed that the COG was marginally higher in IDS than in ADS ($t_{16} = 1.78, p = .094$). Twelve of the 17 mothers included in this analysis had a higher COG of /s/ in IDS than in ADS. There was no evidence that the COG differed between IDS-11 and IDS-15 ($t_{17} = 1.02, p = .324$).

	F0-median (Hz)	F0-excursions (ST/sec)
IDS-11	222 (14.8)	7.5 (1.51)
IDS-15	234 (20.5)	8.1 (1.84)
ADS	205 (24.0)	4.4 (1.43)

Table 5: **The average F0-median and F0-excursions in IDS to infants at 11 months, IDS to infants at 15 months, and ADS, computed over the median values for the 18 mothers.** F0-median is reported in hertz, the F0-excursions are reported in semitones per second. Averages are computed over the medians per speaker (see the text for details), and standard deviations are given in parentheses.

2.3.4 *Pitch characteristics*

Table 5 gives the average pitch measures, F0-median (in hertz) and F0-excursions (in semitones per second), in ADS, IDS-11, and IDS-15. Paired-samples *t*-test comparing F0-median and F0-excursions in IDS and ADS showed that mothers spoke at a higher F0-median and with larger F0-excursions to their infant than to an adult (F0-median: $t_{17} = 4.24, p < .001$; F0-excursions: $t_{17} = 9.85, p < .001$). Fifteen of the 18 mothers had a higher F0-median in IDS than in ADS, and 17 had larger F0-excursions in IDS than in ADS. Paired-samples *t*-tests were performed to compare F0-median and F0-excursions between IDS-11 and IDS-15. The mothers' F0-median was significantly higher to their infant at 15 months than to their infant at 11 months and their F0-excursions were not significantly larger to their infant at 15 months (F0-median: $t_{17} = 2.49, p = .023$; F0-excursions: $t_{17} = 1.29, p = .214$). Fourteen of the 18 mothers spoke at a higher F0-median to their infant at 15 months and 14 mothers spoke with larger F0-excursions to their infant at the older age.

2.4 Conclusion and Discussion

This study tested whether enhancement of the auditory contrast between the corner vowels in IDS is indeed a cross-linguistically universal pattern (Kuhl et al., 1997) that occurs independent of mothers' positive affect, as a result of their attempts to clarify the linguistic structure (Uther et al., 2007). The current results show that Dutch mothers *decrease* the size of their vowel space (as measured on an auditory scale) when they speak to their infant, and raise F2. The second question was whether the pronunciation of speech sounds and the prosodic characteristics of IDS change between the infants' age of 11 and 15 months. In Dutch IDS, mothers raise F2 more to infants at 11 months of age, while their change in pitch level is more extreme in the speech to infants at 15 months of age.

The rise of F2 in Dutch IDS can be related to mothers' smaller vowel space in that register. Dutch mothers in this study seemed to raise F2 more in back vowels than in front vowels, which effectively reduced the auditory contrast between the front and the back vowels in IDS and led to the observed smaller vowel space.[6] These patterns of change in the Dutch infant-directed vowel space are strikingly similar to those observed in Norwegian IDS (Englund and Behne, 2005, 2006). These results do not support the claim that enhancement of the vowel space is a universal property of IDS, and show that the second attested pattern of vowel changes in IDS, a rise of the formant frequencies (cf. Dodane and Al-Tamimi, 2007 and Green et al., 2010), can take place at the expense of the auditory contrast between vowels.

In the introduction, two explanations for raised formant frequencies in IDS were proposed. The first was that mothers smile more to their infant than to an adult (Englund and Behne, 2005), the second was that mothers imitate their child (Dodane and Al-Tamimi, 2007). If mothers were imitating their children in IDS, they would produce /s/ with a lower spectral peak in IDS than in ADS (Nissen and Fox, 2005). However, the current data suggested a higher spectral frequency of /s/ in Dutch IDS. A higher spectral frequency in fricatives is a property of emotional speech, such as happy speech (Kienast and Sendlmeier, 2000). Furthermore, a higher F2 is associated with lip spreading, such as occurs during smiling (Tartter, 1980; Tartter and Braun, 1994; Zacher and Niemitz, 2003; but see Aubergé and Cathiard, 2003). Since the front vowels /i/ and /aː/ are produced with some lip spreading and the back vowels are normally produced with unspread lips, smiling in IDS would mostly raise F2 in the back vowels. In addition, the formant frequencies of low vowels may be more susceptible to changes in affect than the formants of the high vowels (Waaramaa et al., 2008). The current results are in line with these vowel-specific effects of smiling, although those results were only marginally significant and require further confirmation.

It has been argued that mothers enhance different aspects of the speech signal as their child matures and in doing so provide input that at each stage in development highlights exactly those aspects that

6 A second cue to the /ɑ/–/aː/ vowel contrast in Dutch is the duration, with /ɑ/ being shorter than /aː/ (Adank et al., 2004). It could be argued that mothers especially enhance duration contrasts between vowels in IDS, as duration is considered a salient cue (Bohn, 1995) and infants are sensitive to this cue in early speech perception (Bohn and Polka, 2001). To test this claim, the median logarithm of the duration of /ɑ/ and /aː/ in IDS and ADS were the dependent variables in a repeated measures ANOVA with Register (IDS vs. ADS) and Vowel (/ɑ/ vs. /aː/) as within-subject factors. 13 mothers provided useable tokens of both low vowels in IDS and ADS and were thus included in the analysis. This resulted in a marginally significant effect of Register ($F_{1,12} = 4.3, p = .06$), but not in a significant Register$*$Vowel interaction ($F_{1,12} = 1.3, p = .27$). Both /ɑ/ and /aː/ were on average longer in IDS (/ɑ/: $m = 65.8$ ms, $sd = 6.98$ ms; /aː/: $m = 87.1$ ms, $sd = 9.79$ ms) than in ADS (/ɑ/: $m = 54.8$ ms, $sd = 5.24$ ms; /aː/: $m = 80.1$ ms, $sd = 9.51$ ms), but there was no evidence that the duration contrast between the two low vowels was enhanced in Dutch IDS.

their infant is learning about (see Malsheen, 1980, for this hypothesis on the relation between input and children's own productions). As infants at 6 months of age show the first signs of language-specific vowel perception (Polka and Werker, 1994; Kuhl et al., 1992), Dutch mothers may stop enhancing the vowel space by the time their infant becomes 11 months old. However, Dutch mothers are similar to Norwegian others in that they both raise F2 and shrink the vowel space in IDS, and Norwegian mothers do so throughout their infant's first six months (Englund and Behne, 2005, 2006). Importantly, Dutch mothers' F2 was found to be lower, more ADS-like, in the vowels spoken to their infant at 15 months than in the vowels addressed to their infant at 11 months of age. This is in agreement with the general trend that affect in mothers' speech becomes less pronounced as their infant grows older (Snow, 1977; Sherrod et al., 1978; Bornstein et al., 1992; Penman et al., 1983; Stern et al., 1983; Amano et al., 2006; Garnica, 1977; Remick, 1976; but see Jacobson et al., 1983). When we consider this age-related change in the raised F2 in Dutch IDS, this acoustic characteristic can be best regarded as a carrier of positive affect.

The raised F2 of the vowels and raised spectral frequency of /s/ in Dutch IDS can be regarded as biologically grounded acoustic carriers of positive affect. Animals tend to make low-frequency sounds, which are associated with large bodies, when being hostile, but in contrast they produce high-frequency sounds, which are more likely to stem from a small body, when they try to be friendly or appease an opponent (Morton, 1977). Ohala (1980, 1984) proposes that humans similarly make use of the relation between sound frequency and body size to signal their intentions. He argues that the smile has become the facial expression of goodwill exactly because its acoustic consequence is a rise of the formant frequencies. Whether the raised F2 and higher spectral mean of /s/ are purely acoustic side effects of smiling, or (partly) the result of other articulatory means that mothers employ to reach these friendly-sounding acoustic effects, is a subject for future research. However, it is clear that mothers' positive affect has an effect on their realization of speech sounds in IDS.

On the other hand, if mothers express less affect to older children, why was their pitch higher to their infant at 15 months than to their infant at 11 months? In the present study, most of the children had started to walk when they came to the lab for their second visit at 15 months. The mothers had to put more effort into maintaining their infants' attention throughout the whole session. At the same time, the 15-month-old children took more initiative in playing, which resulted in more interactive situations. The interactional context is known to influence mothers' expression of emotion in IDS (Fernald, 1989; Papoušek et al., 1991; Stern et al., 1982; Katz et al., 1996). During the infants' first year, the rated affect and the acoustic pitch characteristics of infant-directed intonation contours are closely related to the infants' developmental stage (Stern et al., 1983; Kitamura et al., 2002; Kitamura and Burnham, 2003). A relatively high pitch and large pitch range in IDS are for example associated with a bid for attention and playing a game (Fernald, 1989) and the larger pitch range in the speech to 9-month-olds in Australian-English IDS has been related to the more directive speech to infants of that age (Kitamura and Burnham, 2003). In the present study, the 15-month old infants' new abilities and behaviors created a different communicative setting, which enhanced their mothers' use of pitch (cf. Sherrod et al., 1978; Penman et al., 1983). Therefore, it appears that in Dutch IDS the raised F2 primarily expresses affect, whereas pitch is more strongly related to the communicative context.

Interestingly, in some of the prior studies that primarily reported on the overall enhancement of the vowel space, raised formant frequencies can be observed as well. Specifically, the vowel space enhancement in Australian-English IDS (Burnham et al., 2002) and Swedish IDS (Kuhl et al., 1992) seems due to the formants of /i/ and /a/ being raised more than those of /u/. This same pattern of vowel-specific formant changes is observed in German smiled speech (Fagel, 2010).[7] Even if mothers enhance the vowel space in IDS, this may be related to the affective characteristics of IDS. An overall lowering of formant frequencies in English IDS can be observed in the results from Lam and Kitamura (2010) and Uther et al. (2007). Formant lowering is a result of lip protrusion (Fant, 1960). Lip protrusion is the main characteristic of a comforting facial expression that is specific to interactions with infants (Stern, 1974; Chong et al., 2003). Investigating why mothers sometimes happily raise their formant frequencies in IDS and soothingly lower them in other studies, and why positive affect would lead to a raised F2 of back vowels and smaller vowel space in Dutch and Norwegian IDS, and to a raised F2 of front vowels and larger vowel space in Australian-English and Swedish IDS[8], will provide detailed insight in the impact that interactive context and affective state have cross-linguistically on the realization of speech sounds in IDS[9].

---

7 Thanks to Alex Cristiá for bringing this to my attention.

8 Anecdotical evidence from three native speakers of English in the Netherlands suggests they perceive the vowel changes in Dutch IDS as overly childish. From this, one could argue that American-English IDS is primarily happy, whereas Dutch IDS is primarily sweet.

9 With respect to the recording context, Englund and Behne (2005) specifically propose that face-to-face contact between mother and child will lead to mothers' excessive smiling, resulting in a shift of the vowel space to higher formant frequencies. Indeed, Englund and Behne (2005) and Green et al. (2010) recorded IDS in a face-to-face situation and observed a raising of the vowel formants in IDS, which is especially remarkable in the latter study on American English. As the current study employed free-play sessions with toys, as was done in (Kuhl et al., 1997; Burnham et al., 2002; Uther et al., 2007) the raised F2 in Dutch IDS cannot be regarded an artifact of the recording setting.

Ideally, such cross-linguistic comparisons of IDS would take into account more speech sounds than only vowels. An increase of the spectral energy of /s/ is also observed in American-English IDS to infants of 13 months old and results in enhancement of the contrast between /s/ and /ʃ/[10] (Cristiá, 2010). In combination with the repeatedly observed enhanced vowel space in American-English IDS, Cristiá's (2010) finding of a higher spectral energy of /s/ indicated that enhancement of speech sound contrasts is a feature of American-English IDS in addition to enhancement of the vowel space (but see Julien and Munson, 2012). In the context of the raised formant frequencies of Dutch IDS, however, the higher spectral frequency of /s/ in Dutch IDS is more readily interpreted as a consequence of affective speech. This comparison underscores that changes in the realization of speech sounds in IDS are best understood if the realization of multiple speech sounds is considered.

One aspect of the present results that requires further investigation is to what extent Dutch caregivers' raising of F2 in IDS depends on the sex of the child or the caregiver. Kitamura et al. (2002) show that the infants' sex impacts the pitch modulations in IDS, with speech to girls having more modulated pitch characteristics than speech to boys. Warren-Leubecker and Bohannon (1984) find that fathers make stronger pitch modifications than mothers to 2-year old children, but speak similarly to 5-year olds and adults. Because the present study included more female than male infants and only female caregivers, these questions cannot be addressed.

A second issue raised by the present results is the effect of a higher F2 in IDS on infants' preference for IDS[11] and their language development. Since very young infants only prefer IDS over ADS when they can listen to the F0 as well as the formant characteristics (Panneton Cooper and Aslin, 1994), there is some evidence that formant frequencies play a role in infants' preferences for certain speech types. The raised F2 in Dutch IDS resulted in a smaller vowel space. Dutch mothers thus do not promote their child's language development by enhancing the vowel space (cf. Liu et al., 2003). However, if infants recognize a raised F2 as an expression of positive affect, this characteristic of Dutch IDS may attract infants' attention to the speech sounds and promote learning. Furthermore, vowels with higher formants resemble children's own vowel productions and may provide infants with a suitable production model (for a similar suggestion regarding segmental simplifications, see Ferguson, 1977, and Lee et al., 2008). Mothers may promote their child's language development in various ways, and enhancing positive affect trough a rise of F2 can be effective via different routes.

10 E.g. /s/ as in *sand* and /ʃ/ as in *shark*.

11 Thanks to Alex Cristiá for clearly stating this question.

To conclude, the results from the present study once more confirm that IDS is a special register in many languages, but that the specific characteristics of this register differ from language to language (e.g. Fernald et al., 1989) and change with the infants' age (Kitamura et al., 2002). In Dutch IDS, mothers' positive affect is reflected in a raised F2 of the vowels. This study has brought us one step closer to understanding how mothers cross-linguistically express affect to their baby in speech.

2.5 Appendix: Details of the analysis

2.5.1 *Vowels*

Formants were automatically measured using the Burg-algorithm (Childers, 1987; Press et al., 1992) as implemented in Praat, with a window length of 25 ms. For automatic formant measurements, the number of formants and the formant ceiling must be specified. Escudero et al. (2009b) have proposed a procedure for estimating optimal ceilings for each vowel of each speaker in a corpus. Because the number of vowel tokens varied across the speakers in the present corpus, the average optimal formant ceilings for the female vowels in Escudero et al.'s (2009b) analysis of Portuguese vowels were adopted in the present analyses. The ceiling for /i/ was set at 6001 Hz, the ceiling for /u/ at 5090 Hz, and the ceiling for /aː/ and /ɑ/ at 5577 Hz, which was Escudero et al.'s optimal ceiling for /a/, the only low vowel in Portuguese.

2.5.2 *The fricative* /s/

The Center of Gravity was measured using a power of 2, to weigh the energy by the power spectrum.

2.5.3 *Pitch*

The F0 curve of each phrase was estimated in hertz using the cross-correlation method. The pitch range for the analysis was set at 120–400 Hz. If the analysis of the median F0 failed for a phrase, all three pitch measures were conducted again with a pitch floor of 75 Hz. If the analysis still failed, the criterion for voicedness was lowered from 0.45 to 0.35 (Escudero et al., 2009b, were followed in this procedure).