Chapter 1

MORE ABOUT BROWS

A cross-linguistic study via analysis-by-synthesis

Emiel Krahmer and Marc Swerts

The computer can't tell you the emotional story. It can give you the exact mathematical design, but what's missing is the eyebrows.

—Zappa (1989), The real Frank Zappa book

Abstract In a seminal paper, Ekman (1979) remarks that brows can play an accentuation role (e.g., to signal focus). However, the literature about eyebrows is inconclusive about their exact role and as a consequence there is no agreement among developers of embodied conversational agents about their precise timing and placement. In addition, it is unclear whether eyebrow movements perform the same role in different languages. In this chapter, an analysis-by-synthesis technique is used to find out what the role of eyebrow movements is for the perception of focus and to see whether this role is the same across different languages. Three experiments are performed, both for Dutch and Italian, investigating where subjects prefer eyebrow movements, whether brows influence the perceived prominence of words and whether they are used in a functional way when subjects interpret utterances. The results for Dutch and Italian are indeed different, but it is argued that these differences can be reduced to prosodic differences between the two languages. The advantages and potential limitations of studies via analysis-by-synthesis are discussed, and an approach to compensate for the limitations is offered.

Keywords: Audio-visual prosody, eyebrow movements, pitch accents, focus, prominence, perception, analysis-by-synthesis, analysis-by-observation, cross-linguistic comparisons.

1. Introduction

How can the *naturalness* of an embodied conversational agent[1] be improved? Arguably, one way is to use *variation*. An agent speaking in a monotonous way and with a static facial expression (only moving its mouth) will look unnatural and people presumably will find it unpleasant to interact with such an agent.

Variation in speech (both in humans and machines) has been the subject of many studies in the past. Some of the variation may be random, such as the smaller instabilities in pitch (jitter and shimmer) that are due to the limited capabilities of a human's vocal apparatus, and that may make synthetic speech more natural when properly implemented. In addition, research has shown that much of the variation in speech is also *functional* in that it can signal communicatively relevant information. Speakers may use pitch accents and prosodic boundaries, for instance, not to counter the monotonicity of their speech, but to give clues to the hearer about how the current utterance should be interpreted (see for instance Ladd 1996 or Cruttenden 1997). There is some psycholinguistic evidence that processing of utterances is indeed enhanced by the 'correct' placement of pitch accents and boundaries (see e.g., Cutler 1984, Terken and Nooteboom 1987, and Sanderman and Collier 1997).

But only variation in speech is not sufficient to create a natural embodied agent. Facial variation is required as well (besides visual correlates of producing the different speech sounds, i.e., movements in the mouth area). For this purpose, many current embodied agents employ some form of *Perlin noise* (Perlin 1995), i.e., small random head movements. Even though Perlin noise certainly makes animations more natural and life-like, the resulting variation is small and not functional in the linguistic sense of the word.[2] Arguably, what is needed is some form of *audio-visual prosody*, where speech cues and facial cues can be used, alone and in tandem, to enhance both the naturalness and the expressiveness of embodied agents.

Arguably, not all facial cues have speech correlates and not all speech cues have facial correlates, but for certain functional aspects of communication there is reason to assume a connection between the two (see e.g., Pelachaud et al. 1996). This implies that knowledge is required about the potential co-occurrence of auditory and visual cues. Concerning this, Pelachaud et al. (1996:32) stated that "there is a lack of empirical information on when an accent or other intonational components are accompanied by a facial action". Unfortunately, this situation has not changed much in recent years, despite a growing number of empirical studies involving embodied conversational agents. One possible

way to further this discussion is as follows. As a starting point, one can look for relevant claims made in the literature, in particular in the many descriptive (non-empirical) studies of non-verbal communication. These claims can subsequently be implemented in an embodied conversational agent. Many researchers and developers of embodied conversational agents indeed follow this strategy, but one can go even further and use the agent implementation to empirically verify, as it were, the original claims. This method could be called *analysis-by-synthesis* and is, in different disguises, applied in Granström et al. 1999, 2002 Nass et al. 2000 and Krahmer et al. 2002a, 2002b, to name but a few).

In this chapter, the analysis-by-synthesis method is used to gain insight in one aspect of audio-visual prosody, namely the signalling of important bits of information in an utterance (the *focus*), via pitch accents and eyebrow movements. It will be argued that analysis-by-synthesis is a powerful evaluation tool, but one that should be used with some caution.

2. About brows

In a seminal paper, Ekman (1979) describes the role of eyebrow movements as emotional and conversational signals. Sometimes the distinction between these two kinds of signals is difficult to make (for instance because both often occur during conversation). Still clear differences between the two exist: conversational signals typically do not occur when a person believes (s)he is unobserved, while emotional signals do. Moreover, emotional but not conversational signals are believed to be universal.

While the use of eyebrows as emotional signals has been addressed in many studies (already in Darwin 1872), the conversational use is still relatively understudied and most of the work that has been done in this area is based on intuitions and impressionistic observations. This is surprising, since eyebrow movements are according to Ekman (1979:183) "probably among the most frequent facial actions employed as conversational signals". Various authors have suggested that eyebrow movements can be used to emphasize important pieces of information (see e.g., Birdwhistell 1970, Eibl-Eibesfeldt 1972, Condon 1976, Ekman 1979). Ekman observes that eyebrows can play this accentuation role in two different ways: they can function as a *baton* (in the terminology of Efron 1941), which may be used to accentuate a particular word as it is spoken, or they can function as an *underliner* (in Ekman's own terminology), where the emphasis stretches out over more than one word.

It is well-known that speakers may use *auditory* speech signals to emphasize words as well. For instance, speakers of Germanic languages (such as Dutch, English and German) can use pitch accents to indicate the information status of words: accents tend to distinguish information that is *in focus* (since it is *new* or *contrastive*) from information which is given from the prior discourse context (see e.g., Chafe 1974, Terken 1984, Hirschberg 1993).

That both eyebrow movements and pitch accents can be used to signal focus, suggests that there is close correspondence between the two. This correspondence has indeed been noted by Morgan (1953) and Bolinger (1985:202ff). The latter formulated his *Metaphor of Up and Down* which implies, among other things, that when the pitch rises or falls, eyebrows tend to follow the same pattern. As an illustration of this metaphor, it is instructive to try and utter a two-word phrase, say "blue square," with a pitch accent (and no corresponding eyebrow movement) on the word "blue" and an eyebrow movement (but no pitch accent) on the word "square". Most people find this a difficult exercise. Yet, speakers have no problems whatsoever to produce the utterance with pitch accent and eyebrow movement on the same word.

One of the few empirical studies devoted to the connection between pitch accents and eyebrow movements is Cavé et al. (1996), who conducted a small production experiment (i.e., they recorded speakers). They found a significant correlation between the two (in particular, and surprisingly, for the *left* eyebrow). This implies that eyebrow movements often co-occur with pitch accents. It is important to realize that the opposite is not the case. Ekman (1979:184): "There are many occasions when people mark emphasis in their speech without either a baton or an underliner." People do *more* with their pitch than with their eyebrows, as the reader can easily verify by looking at an arbitrary speaker.

If not all emphasized words are accompanied by an eyebrow movement, which words are? This is still an open question. Ekman (1979:184) is "not optimistic about being able to predict when a baton or underliner will be used and when emphasis will be carried just by voice, although perhaps there might be some weak relationship with overall involvement in what is said."

It thus appears that the literature on non-verbal behavior is inconclusive about the role of eyebrow movements for communication. As a result, it is no surprise that among developers of embodied conversational agents there is no consensus about the timing and placement of eyebrow movements. Pelachaud et al. (1996) assume that the conversational use of eyebrow movements is affect dependent (e.g., it is assumed that a disgusted person uses more eyebrow movements than, say, a sad

one). In response to the question *I know that Harry prefers* POTATO *chips, but what does* JULIA *prefer?*, a disgusted agent would respond with:

$$(\overline{\text{JULIA}}\ \text{prefers})_{\text{theme}}\ (\overline{\text{POPCORN}})_{\text{rheme}}$$

(Here and elsewhere, small caps indicate a pitch accent, and over-lined words are accompanied by an eyebrow movement.) Cassell et al. (2001: 482) use eyebrow movements (or *flashes* as they call them[3]) more sparingly. The eyebrows are raised only when an *object* is introduced in the rheme. So, in response to the question above, the *beat* algorithm of Cassell and co-workers would not produce an eyebrow movement on Julia. It is worth noting that neither Pelachaud et al. (1996) nor Cassell et al. (2001) report on empirical evaluation. As a result we get no insight in the effectiveness of the animations; it is unknown, for instance, whether eyebrow movements influence the way human listeners process the information.

The general picture that emerges is that both pitch accents and eyebrow movements may be used to signal focus. Eyebrow movements tend to accompany pitch accents, but the opposite is not the case; often words may be emphasized in speech, but not accompanied by an eyebrow movement. On the basis of such observations, Cavé and co-workers suggest that eyebrow movements and pitch do not link up automatically (e.g., due to muscular synergy), but rather coincide for *communicative* reasons. Naturally, one wonders what these communicative reasons might be. In general, it is uncertain what the function of eyebrow movements for the perception of focus is. Do they help in emphasizing a particular word as it is spoken? Do they influence the way human listeners process information in a functional way?

There is still another complication. Various researchers have stressed the functional link between eyebrow movements and pitch accents. However, pitch accents have different functions in different languages; they play an important role in Germanic languages as signallers of information status, but this is not a linguistic universal. In Romance languages (such as Italian and Spanish), for instance, less use is made of pitch accents (and prosody in general) to mark information status (certainly within syntactic constituents, Ladd 1996:177ff). Instead, word order variation may be used for this purpose. This raises the question what the function of eyebrow movements is for Romance languages. It is not obvious that eyebrow movements perform the same function for focus perception in Romance languages as they do in Germanic ones. In sum, the general picture raises (at least) two questions:

Question one What *is* the role of eyebrow movements for the perception of focus?

Question two Is this role the same across languages?

Below these questions are addressed via an analysis-by-synthesis method, applying it to both Dutch (a Germanic language) and Italian (a Romance one). In section 3 the stimuli used in the three experiments are described. The first experiment (section 4) is about *subjective preferences*, asking both Dutch and Italian subjects where they prefer to see eyebrow movements in relation to pitch accents. In the second experiment (section 5) it is investigated what the contribution of eyebrow movements is for the *perceived prominence* of words in Dutch and Italian. The third experiment (section 6) is a *functional study*, investigating to what extent Dutch and Italian subjects use pitch accents and eyebrow movements to interpret incoming utterances. We end with a general discussion, in which we attempt to answer the two general questions introduced above. In addition, we discuss the pros and cons of the analysis-by-synthesis method, and offer a general remedy to alleviate some of the cons of this method.

3. Materials

In all three experiments the stimuli consisted of animations of a male Talking Head uttering the Dutch phrase "blauw vierkant" (*blue square*) or the Italian phrase "triangolo nero" (*black triangle*).

3.1 Speech

The Dutch and Italian speech materials were collected in a (semi-) spontaneous way in two earlier production experiments (for more details see Krahmer and Swerts 2001 or Swerts et al. 2002). This was done using a simple dialogue game, played by four Dutch pairs and four Italian pairs of speakers, thus giving eight speakers per language. All Dutch subjects were students and colleagues working in the south of the Netherlands and speaking standard Dutch. The Italian speakers were all living in Italy and were native speakers of the Tuscan variety of Italian.

The dialogue game is essentially an alignment task of figures played by two subjects, call them A and B, who are separated from each other by a screen. In each game, both players have an identical set of eight cards at their disposal, each card displaying a geometrical figure in a particular color (such as a blue square or a black triangle). Four of these cards are put on a stack in front of the subjects, the remaining four are in a row before them. The four cards in the *stack* of A are the same as

Table 1.1. Example contexts for collection of target utterances in Dutch ("blauwe vierkant", *blue square*) and Italian ("triangolo nero", *black triangle*).

Context	*Dutch*	*Italian*
CC	A: rode driehoek B: blauwe vierkant	A: rettangolo rosa B: triangolo nero
GC	A: blauwe driehoek B: blauwe vierkant	A: triangolo rosa B: triangolo nero
CG	A: rode vierkant B: blauwe vierkant	A: rettangolo nero B: triangolo nero

those in the *row* of B, and *vice versa*. The task for both subjects is to create an identical ordered list of geometrical figures. The game consists of a series of turns in which one participant describes the figure on top of his or her stack and instructs the other participant to select this card. Once a card has been described, both players discard it by placing it in the ordered list. After each turn the subjects change roles, so that the instruction-giver in one turn is the instruction-follower in the next turn. The game is over when both players are out of cards. There are no winners or losers. Each pair of subjects plays eight games. There is always a two minute break between games. Speakers found it an easy game to play.

The data thus obtained allow for an unambiguous operationalization of the relevant contexts. A property is defined to be *given* (**G**) if it was mentioned in the previous turn, and it is *contrastive* (**C**) if the figure described in the previous turn had a different value for the relevant attribute. Here we ignore initial dialogue contributions, so all properties are either given of contrastive. We say that a phrase is *in focus* if it is contrastive.

By systematically varying the order of the cards in the stack, we collected target utterances ("blauw vierkant" for Dutch and "triangolo nero" for Italian) in three different contexts: all contrast (**CC**), contrast in the final word (**GC**) and contrast in the pre-final word (**CG**). Note that in the two-letter abbreviations of the contexts the first letter represents the information status of the first word and the second letter that of the second word in the utterance. Table 1.1 summarizes the three contexts of interest and illustrates them with Dutch and Italian examples.

A distributional analysis was performed for all target utterances by three independent labellers for Dutch and three independent ones for

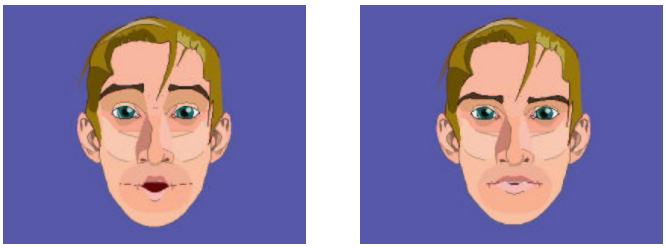

Figure 1.1. Two stills from the Talking Head uttering "blauw vierkant" (blue square) with a raised eyebrow on the first word (left) and no eyebrow action on the second word (right).

Italian. All labellers were intonation experts and did not know the discourse context of the utterances while labelling them. For the utterances used in the three experiments below, the results are unequivocal. In Dutch, words receive a pitch accent when they are in focus (here: contrastive). In Italian, every word is always accented, irrespective of the discourse context. All Italian speakers produce the same intonation contour in all contexts: a double accent (the pitch contour may be likened to a "flat hat"), with the second accent downstepped with respect to the first (the hat is dented). Thus, the first accent is stronger and more prominent than the second, which is reflected, among other things, in that it has a larger pitch excursion size (i.e., a larger difference between the minimum and maximum F_0, Swerts et al. 2002:643).

In sum, the distribution of pitch accents is context-dependent in the Dutch data and clearly reflects the information status of words; a focussed (contrastive) word carries a pitch accent, an unfocussed (given) word does not. The distribution in the Italian data is different in that it is always the same irrespective of the context; it provides no clues about the focus of the utterance.

3.2 Animations

The animations used in the experiments were made with the *CharToon* environment (e.g., Ruttkay and Noot 2000), and take a 2D head of a male character as their basis. Chartoon animations are based on constraints over control points (Ruttkay 2001). As speech materials we used the utterances of "blauw vierkant" and "triangolo nero" that our male speakers produced in the relevant contexts. Visual speech is generated on the basis of a set of 48 visemes. Phonemes from the input speech are mapped to corresponding visemes with a sampling rate of 100ms, while intermediate stages are computed using linear interpolation. Rapid eye-

brow movements coincide with the stressed syllable of either the first or the second word in the relevant utterances. Notice that these are eyebrow counterparts of focus on the first word and focus on the second word respectively. We mark the presence of an eyebrow movement by placing a line over the relevant character; thus, for instance, a $\overline{\mathbf{C}}\mathbf{G}$ animation uses speech that was collected in a CG context (the first word is contrastive, the second given) and the first word is associated with an eyebrow movement.

The eyebrow movements always had the same pattern: first, a 100ms dynamic raising part, then a static raised part of 100ms, and finally a dynamic lowering part of 100ms. The overall length of the movement is comparable with the average duration of rapid eyebrow movements of human speakers (±375ms, Cavé et al. 1996). We opted for slightly shorter movements due to the overall short duration of the spoken utterances. The 300ms long movement also aligned nicely with the onset and offset of syllables in the Dutch and Italian words used in our stimuli. The brow movement always corresponds with Action Unit AU 1+2 (Ekman and Friesen 1978).

4. Experiment 1: Subjective preference

4.1 Method

In the first experiment, subjects were presented with minimal pairs of stimuli. The members of these pairs were always identical in terms of their sound properties, including the pitch accent distribution. They only differed in that one member had an eyebrow movement on the first word while the other had an eyebrow movement on the second word. Subjects were asked in which of the two sound and image were best synchronized.

Subjects were 25 native speakers of Dutch for the Dutch experiment and 25 native speakers of Italian for the Italian experiment.[4] They watched and listened to the Talking Head uttering the different pairs of two-word phrases “blauw vierkant” (Dutch) and “triangolo nero” (Italian). Two male voices were used for each language. All pairs of stimuli, in both AB and BA order, were presented randomly. Subjects could watch and listen to each pair twice, and were encouraged to select, by forced choice, the most natural animation from the pair after the first presentation, and then verify their initial choice during the second showing. Before the actual experiment subjects entered a brief training session (consisting of three pairs of stimuli) to make them acquainted with the experimental setting and the kind of stimuli. No feedback was given on the ‘correctness’ of their answers and there was no further commu-

Table 1.2. Preference judgements (Dutch) for eyebrow movement on first or second word as a function of context ($N = 300$; 12 stimuli $\times$ 25 subjects).

	Eyebrow preferred on	
Context	*First word*	*Second word*
CC	.60	.40
GC	.38	.62
CG	.75	.25

nication with the experimenter. The experiment itself consisted of 12 stimuli per language: 3 different contexts (CC, GC, CG) $\times$ 2 voices $\times$ 2 orders (AB and BA). Subjects were not informed about the kinds of cues they could pay attention to while making their selection. The experiment lasted approximately 5 minutes.

4.2 Results (Dutch)

The Dutch results are given in Table 1.2. The overall distribution is significantly different from chance ($\chi^2(2) = 34.8$, $p < 0.001$). Looking at the top row in this table, it can be seen that there is a mild preference for the eyebrow movement to be aligned with the first word in the all-focus (double contrast) case, which is realized in our Dutch speech data with a double accent. However, the next two lines with results on utterances with a single accent, clearly indicate that Dutch subjects disprefer cases where the eyebrow movement and the pitch accent do not coincide. Arguably, such stimuli are 'inconsistent' in that the speech cues indicate that one word is in focus, while the eyebrows suggest that the other word is in focus. Apparently, Dutch subjects prefer cases where pitch and eyebrows are synchronized. This preference is clearest in the case where the pitch accent falls on the first word (CG); in 75% of these cases, the Dutch listeners prefer the eyebrow movement on the first word as well. This is in accordance with our earlier speech-only results (Krahmer and Swerts 2001). In Dutch the default position for the nuclear accent (basically, the most prominent accent in a phrase) is the final word. When the pre-final word is in focus (and the final word is given), the nuclear accent shifts to a non-nuclear position and as a consequence it is somewhat more 'conspicuous' than when the nuclear accent appears in default position.

Table 1.3. Preference judgements (Italian) for eyebrow movement on first or second word as a function of context ($N = 300$; 12 stimuli $\times$ 25 subjects).

	Eyebrow preferred on	
Context	*First word*	*Second word*
CC	.84	.16
GC	.76	.24
CG	.79	.21

4.3 Results (Italian)

The Italian results can be found in Table 1.3. Again, the overall distribution is significantly different from chance ($\chi^2(2) = 106.92$, $p < 0.001$). Inspection of the table reveals that Italian subjects have a clear preference for the eyebrow movement to coincide with the first word, irrespective of the context. This is in line with the earlier observation that even though both words always receive an accent, the accent on "triangolo" is more prominent than the one on "nero".

4.4 Discussion

The Dutch and Italian results are significantly different (Pearson $\chi^2(5) = 49$, $p < 0.001$). Interestingly, these differences can be reduced entirely to prosodic differences between the two languages. The Italian subjects prefer the animations with the eyebrow movement on the first word, irrespective of the context. This can be explained through the fact that the first accent is the most prominent one. In the cases where our two Dutch speakers produced a single accent (CG and GC) the Dutch subjects prefer the animation in which eyebrow movement and pitch accent fall on the same word. So, in general, if an animation contains an eyebrow movement, *both* Dutch and Italian subjects prefer the eyebrow movement on the most prominent word, the difference being that in Italian the most prominent word is always the first one while in Dutch this depends on the context (see Swerts et al 2002 for more details).[5] The fact that in Dutch pitch accents and eyebrow movements are preferred to be aligned, suggests that they may serve the same purpose, namely to render a word more prominent. This issue is investigated further in study 2.

5. Experiment 2: Perceived prominence

5.1 Method

Subjects were again confronted with pairs of animations that have identical speech properties (including pitch accent distribution), but differ in the presence and placement of eyebrow movements. Unlike in the previous study, however, for experiment 2 each pair of stimuli consisted of one animation without any eyebrow movements and one animation with an eyebrow movement on either the first or the second word of the utterance. Given the finding of the previous experiment that listeners disprefer cases where pitch accents and eyebrows do not coincide (a situation that can only arise in Dutch), the eyebrow movements in the second study always accompanied a pitch accent. This implies that Dutch subjects had to make less pairwise comparisons than Italian ones, since the two 'inconsistent' kinds of Dutch stimuli (i.e., $\overline{\text{G}}$C and C$\overline{\text{G}}$) are left out of consideration. For both the Dutch and the Italian study, the same 25 subjects from study 1 participated. Moreover, the same two male voices for each language were used.

The second experiment consisted of four different sessions. In two sessions subjects had to focus on the first word (the adjective "blauw" for Dutch and the noun "triangolo" for Italian), once for each male voice. In the two other sessions, subjects had to focus on the second word (the noun "vierkant" for Dutch and the adjective "nero" for Italian). In all four sessions, subjects had to determine by forced choice which of the two animations contained the most prominent realization of the word of interest. The stimuli were presented in two different random orders to compensate for any learning effects. For both Dutch and Italian, half of the stimulus pairs in each session were distractors. These consisted of utterance pairs that were not only different in terms of eyebrow movements, but also used different speech realizations (taken from different contexts), in an attempt to deliberately confuse subjects about the purpose of the experiment. Before a session started, subjects again entered a brief training session (one stimulus pair per session) to make them acquainted with the material and the task. Again, no feedback was given on the 'correctness' of their answers and there was no further communication with the experimenter. Apart from the distractors, each Dutch session consisted of 4 pairs of stimuli and each Italian session of 6 pairs. Subjects were not informed about the kinds of cues they could pay attention to while making their selection. The second experiment lasted approximately 15 minutes.

Table 1.4. Prominence judgements (Dutch) for the first word ("blauw") and the second word ("vierkant") in animations with an eyebrow movement either on the first or second word (indicated by a line on top of the relevant word). ($N = 400$; 2 words $\times$ 4 comparisons $\times$ 2 voices $\times$ 25 subjects).

Word	*Pairwise comparisons*	
Blauw	$\overline{\text{blauw}}$ vierkant .95	blauw vierkant .05
	blauw $\overline{\text{vierkant}}$ .10	blauw vierkant .90
Vierkant	$\overline{\text{blauw}}$ vierkant .14	blauw vierkant .86
	blauw $\overline{\text{vierkant}}$ .90	blauw vierkant .10

5.2 Results (Dutch)

Table 1.4 gives a summary of the results obtained for Dutch. To keep the table readable, we do not present separate results for each individual pairwise comparison. In fact, the different speech conditions gave rise to very little variation anyway; all pairwise comparisons yielded significant differences, with χ^2 scores in the range of 28.8 and 42.3, $df = 1$, $p < 0.001$) (the interested reader may consult Krahmer et al. 2002b for the detailed tables).

Looking at the results for the first word, "blauw", it is clear that the presence of an eyebrow movement on this word (marked by an over-line) has an effect on the perceived prominence; in 95% of the cases, subjects consider "blauw" more prominent in the animation where this word is associated with an eyebrow movement. Whether the phrase contains only one accent (CG) or two accents (CC) does not influence the result. This suggests that pitch and eyebrows have an additive effect for prominence ratings. Interestingly, eyebrow movements also appear to downscale the perceived prominence of words that appear in the immediate context of a word which is accompanied by an eyebrow movement. This can be seen from the second row of Table 1.4; when the word "vierkant" is associated with an eyebrow movement, subjects consider the utterance of "blauw" in the animation *without* eyebrow movements the most prominent one (and recall that the speech in the two animations is identical).

Table 1.5. Prominence judgements (Italian) for the first word ("triangolo") and the second word ("nero") in animations with an eyebrow movement either on the first or second word (indicated by a line on top of the relevant word). ($N = 600$; 2 words × 6 comparisons × 2 voices × 25 subjects).

Word	*Pairwise comparisons*	
Triangolo	$\overline{\text{triangolo}}$ nero .85	triangolo nero .15
	triangolo $\overline{\text{nero}}$ .35	triangolo nero .65
Nero	$\overline{\text{triangolo}}$ nero .29	triangolo nero .71
	triangolo $\overline{\text{nero}}$ .71	triangolo nero .29

The results for "vierkant" in the lower half of Table 1.4 mirror those for "blauw" in the upper half: when an eyebrow movement accompanies "vierkant", this boosts the perceived prominence of this word (fourth row), but when the eyebrow movement is associated with "blauw", the word "vierkant" is perceived as less prominent (third row).

5.3 Results (Italian)

The overall Italian results are summarized in Table 1.5. All but one of the pairwise comparisons are statistically significant, with χ^2 values in the range of 3.92 ($df = 1$, $p < 0.05$) and 35.28 ($df = 1$, $p < 0.001$). The only non-significant comparison is one in which subjects had to focus on "triangolo" in a GC context.

The general picture that emerges from Table 1.5 is the following. If "triangolo" is accompanied by an eyebrow movement, subjects rate its prominence higher than when it is not accompanied by such a movement. Alternatively, if the eyebrow movement occurs on the word "nero", in 65% of the cases, subjects consider "triangolo" more prominent in the animation without eyebrows. The basic picture for the second word ("nero") is essentially the same; the presence of an eyebrow movement on "nero" increases its perceived prominence, but when the eyebrow movement is associated with "triangolo" this reduces the perceived prominence of "nero".

5.4 Discussion

The results for Dutch and Italian are very similar: the presence of an eyebrow movement boosts the perceived prominence of the associated word and downscales the prominence of the preceding or following word. This effect holds for both the first and the second word, and is independent of the context in which the speech was uttered. This is in line with earlier observations from Krahmer and Swerts (2001) that prominence judgements are very much dependent on the prosodic context, in that an isolated pitch peak is perceived as more prominent than the same peak presented in the context of an intonationally comparable pitch peak. The results for Italian are somewhat less pronounced than the Dutch ones, in particular when the eyebrow movement occurs on "nero" or when "nero" is the word of interest. This is might be due to the inherent prominence of "triangolo" in these utterances.

So far, the results for both languages are consistent with claims that eyebrow movements are relevant for prominence perception. In the next experiment it is examined to what extent subjects *use* information from eyebrow movements when processing utterances.

6. Experiment 3: Functional analysis

6.1 Method

In the third study it is investigated to what extent Dutch and Italian subjects use audio-visual cues when interpreting utterances. For this purpose a "dialogue reconstruction" experiment is used (Swerts et al. 2002). Subjects watch and listen to the Talking Head uttering (the Dutch and Italian) counterparts of "blue square" (i.e., "blauw vierkant" or "triangolo nero"), with a certain intonation contour (taken from its original context) and an eyebrow movement on either the first or the second word. This gives rise to six different kinds of stimuli ($\overline{\text{C}}$C, C$\overline{\text{C}}$, $\overline{\text{G}}$C, G$\overline{\text{C}}$, $\overline{\text{C}}$G, and C$\overline{\text{G}}$). For Italian, four male voices were used. For Dutch, six male voices were used; four human speakers recorded in the earlier dialogue game experiment and in addition, two synthetic speakers (copying intonation contours of two human speakers).[6]

The task for the subjects is to decide by forced choice what the *preceding* utterance would have described: (1) a red square, (2) a blue triangle or (3) a red triangle. To perform this task subjects have to determine what the focus of the *current* utterance is: (1) the first word ("blue"), (2) the second word ("square") or (3) both. See Table 1.1 for the actual Dutch and Italian phrases used in the third experiment.

Table 1.6. The perception of focus in Dutch as a function of context ($N = 900$; 6 conditions $\times$ 6 voices $\times$ 25 subjects).

Context	*Blauw*	*Focus perceived on* *Vierkant*	*Both*
$\overline{\text{C}}$C	.30	.27	.43
C$\overline{\text{C}}$	.14	.47	.39
$\overline{\text{G}}$C	.17	.61	.22
G$\overline{\text{C}}$	.18	.60	.22
$\overline{\text{C}}$G	.75	.15	.10
C$\overline{\text{G}}$	.70	.20	.10

Subjects were 25 native speakers of Dutch (different from those used for studies 1 and 2) and 25 native speakers of Italian (the same as those for studies 1 and 2).[7] Before the actual experiment started, subjects entered a brief training session (3 stimuli), to make them acquainted with the experimental setting and the kind of stimuli. No feedback was given about the 'correctness' of their answers, and there was no further communication with the conductor of the experiment. The experiment consisted of 36 stimuli for Dutch (6 voices $\times$ 6 conditions) and 24 for Italian (4 voices $\times$ 6 conditions). The experiment lasted approximately 10 minutes.

6.2 Results (Dutch)

In Table 1.6 the results of the dialogue reconstruction experiment for Dutch are given. The overall distribution is significantly different from chance ($\chi^2(10) = 292.2$, $p < 0.001$). First consider the cases where the speech has a single pitch accent, either on the adjective (CG) or the noun (GC). In the first case, the majority of the subjects perceives the focus on the word "blauw", while in the second case, the majority of subjects perceives the focus on the word "vierkant". Hence, in both cases subjects perceive the focus on the accented word, irrespective of the position of the eyebrow movement. Nevertheless, if we compare the distribution obtained with an eyebrow movement on the first word with the distribution obtained with such a movement on the second word, a significant difference is found (Pearson $\chi^2(8) = 19$, $p < 0.025$). This difference is primarily due to the cases where both words receive a pitch accent (CC). In those cases, a word which is associated with an eyebrow movement is perceived to be in focus roughly twice more often than when the word is *not* accompanied by a brow movement. Thus,

Table 1.7. The perception of focus in Italian as a function of context ($N = 600$; 6 conditions $\times$ 4 voices $\times$ 25 subjects).

	Focus perceived on		
Context	*Triangolo*	*Nero*	*Both*
$\overline{\text{C}}$C	.36	.31	.33
C$\overline{\text{C}}$	.35	.28	.37
$\overline{\text{G}}$C	.37	.29	.34
G$\overline{\text{C}}$	.26	.49	.25
$\overline{\text{C}}$G	.25	.36	.39
C$\overline{\text{G}}$	.32	.38	.30

if the eyebrow movement coincides with "blauw", subjects perceive the focus on this word in 30% of the cases (as opposed to 14% of the cases when no eyebrow movement accompanies "blauw"). And, if the eyebrow movement is aligned with "vierkant", this word is perceived to be the focussed one in 47% of the cases (as opposed to 27% of the cases when no eyebrow movement accompanies "vierkant"). So, for Dutch both pitch accents and eyebrow movements can influence the perception of focus, albeit that the effect is much larger for pitch.

6.3 Results (Italian)

The Italian results are rather different, as the reader can observe in Table 1.7. The overall distribution is not significantly different from chance ($\chi^2(10) = 16.8$, n.s.). Moreover, the distribution obtained with the eyebrow movement on the first word is not significantly different from that with the movement on the second word (Pearson $\chi^2(8) = 10.84$, n.s.) This indicates that Italian subjects can not reconstruct the dialogue history on the basis of the audio-visual properties of the stimuli. Put differently, the placement of pitch accents and eyebrow movements does not provide any clues for our Italian subjects about the context.

6.4 Discussion

The results show that Dutch subjects are capable of "reconstructing the dialogue history" in the current experiment, while Italian subjects are not. The results for both languages confirm the earlier speech-only results of Swerts et al. 2002. In the Dutch speech-only results, subjects could reconstruct the dialogue history best in the CG case (because the nuclear accent falls on a non-default position) and least in the CC case.

In the current experiment we can basically observe the same picture. Interestingly, the eyebrow movements contribute only in the all contrast (CC) case, which is the one where the speech cues are least informative. Overall we see somewhat more confusion in the current experiment than was found in the speech-only experiment. This might indicate that the presence of the face is somewhat distracting for subjects. Similar observations have been made for 'real' face-to-face communication (Doherty-Sneddon et al. 2001).

In our earlier speech-only experiment for Italian we found that subjects are incapable to reconstruct the dialogue history on the basis of prosodic cues. This was not surprising, since our Italian speakers always pronounced "triangolo nero" with the same contour irrespective of the context. On the basis of this, and in analogy with the Dutch CC case, one might hypothesize that eyebrow movements would contribute more for Italian than they did for Dutch. This would also be in line with observations from Rimé and Schiarature (1991) that gestures occur more when speech cues are underspecified. But in fact, the opposite of our expectation turned out to be true: eyebrows contributed *less* for Italian than they did for Dutch.

Thus, again we find differences between Dutch and Italian (eyebrows do something for focus perception in Dutch and nothing in Italian) and again these differences seem related to prosodic differences between the two languages (prosodic cues contribute to focus perception for our Dutch but not for our Italian speech materials).

7. General Discussion

Eyebrows in Dutch and Italian. This chapter has reported on three experiments with an embodied agent, in an attempt to gain more insight into the cue value of eyebrow movements for the perception of focus in Dutch and Italian.

The first experiment tested how Dutch and Italian listeners react to two-word stimuli with an eyebrow movement either on the first or on the second word. Results showed that our Dutch subjects prefer those animations in which the eyebrow movement is synchronized with a word that carries a pitch accent (due to contrastiveness) rather than with an unaccented word. Our Italian subjects preferred the eyebrow movement to occur on the first word, irrespective of its information status. So Dutch and Italian subjects appear to have different preferences, but these differences can be explained entirely by the prosodic differences between the two languages. Essentially, both Dutch and Italian subjects prefer the eyebrow movement to coincide with the most prominent word in the

utterance, which is determined by context in Dutch and always is the first word in our Italian speech data.

The second experiment investigated whether listeners are sensitive to eyebrow movements when they have to rate the prominence of particular words. This experiment showed that for both the Dutch and the Italian stimuli, eyebrow movements boost the perceived prominence of the word they are associated with and simultaneously downscale the prominence of words in the immediate preceding or following context. The situation was somewhat more clear for Dutch than for Italian, which again can be ascribed to prosodic differences between the languages (in particular to the inherent prominence of the first pitch accent with respect to the second, downstepped one in the Italian stimuli).

The third experiment tried to found out the relative contributions of pitch accents and eyebrow movements for the perception of focus in Dutch and Italian. Our Dutch listeners use both cues to determine the focus of an utterance, albeit that the effect of pitch accents is much larger than that of eyebrow movements. The latter only contribute when speech cues are relatively unclear (i.e., the double contrast, CC case). Our Italian subjects, however, were unable to determine the focus of the utterances. The differences between the Dutch and Italian results once again mirror prosodic differences between the two languages. In earlier work (Swerts et al. 2002) we have found that Dutch listeners can and Italian listeners cannot determine the focus of utterances on the basis of auditory cues alone.

This suggests that the two questions from the introduction may be answered as follows. About question one: eyebrow movements seem to play only a secondary role for the perception of focus; they follow pitch accents and mainly enhance the perceived prominence of words. And concerning question two: the proper placement of eyebrow movements is language dependent and their functional contribution may differ per language. Interestingly, however, to the extent that eyebrow movements have different functions in the languages under consideration here, these differences can be fully explained from the prosodic differences between the languages.

The first two experiments confirm the earlier claims that eyebrow movements and pitch accents are related for communicative purposes; both the Dutch and Italian subjects prefer the eyebrow movement to coincide with the most prominent word, and the brow movement indeed seems to perform some accentuation function. Still, it remains puzzling that eyebrow movements play only a small (Dutch) or no (Italian) role for the perception of focus. It might be that eyebrow movements are exploited more consistently as a cue to different kinds of conversational

phenomena.[8] Another explanation might be that listeners are simply more biased to auditory cues than to visual cues for focus perception. This would be in line with the earlier observation that speakers do more with pitch than with eyebrows. Many accented words are not accompanied by a baton or an underliner, so it is not unlikely that we are most sensitive to verbal prosody.

About analysis-by-synthesis. Analysis-by-synthesis is a powerful evaluation method, which may provide useful empirical data about the relation between verbal and visual prosody. The two main advantages of the method are that (a) one has direct control over all the relevant parameters, and (b) once a theory has been implemented (and evaluated positively) it can be applied directly in an embodied conversational agent. One can think of many variations on the three experiments discussed above that could be pursued using the analysis-by-synthesis method. For instance, we have only looked at one eyebrow movement (AU 1+2). Ekman and Friesen (1978) also describe another brow movement that may serve as a baton or an underliner, namely AU 4 (in which the brows are lowered and drawn). According to Ekman (1979) this movement can have a similar function as AU 1+2, but seems to contain an element of doubt as well. It would be interesting to test this.[9] Other variations involve manipulating the duration and the strength of the brow movement. What happens if we would use shorter/longer movements, where the eyebrows move upwards to a lesser extent? Would they still increase the perceived prominence of words? For such research questions, the analysis-by-synthesis method seems very useful.[10]

There is also a potential disadvantage of the analysis-by-synthesis method, however, in that the results may be incomplete. In the three experiments described above, we only manipulated one parameter (brows) and measured the results. Still, it might be that some other visual factor or combination of factors is more relevant for focus perception. Since no other parameters were manipulated in the experiments, such an alternative explanation cannot be ruled out. Of course, we can redo the experiments with, say, head nods, in combination with or instead of eyebrow movements. It might, for instance, be the case that head nods are more convincing visual cues than eyebrow movements for reconstructing the dialogue history (experiment 3). But even if that were the case, it would not solve the general problem. After all, it might be that there still is another cue or combination of cues which more accurately corresponds with focus signalling. The number of potential cue combination grows explosively and it does not seem feasible to try out all of them via analysis-by-synthesis experiment. In our opinion, the best way to ad-

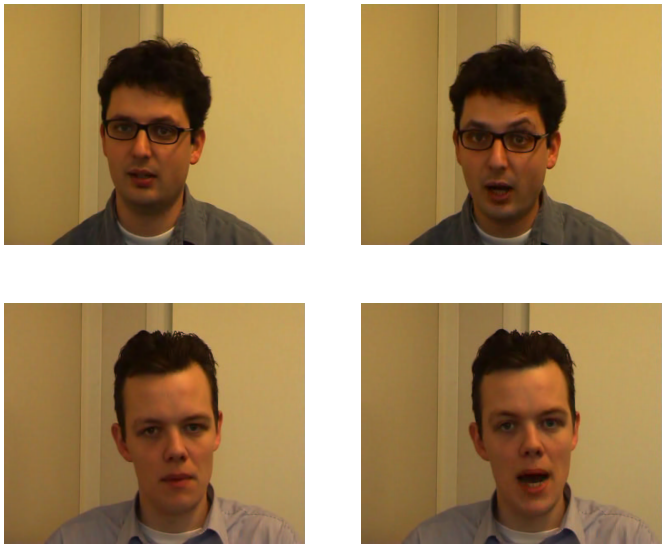

Figure 1.2. Representative stills of two subjects uttering unstressed (left) and stressed (right) syllables.

dress this potential problem is by combining analysis-by-synthesis with *analysis-by-observation*. Below we describe such an analysis.

Analysis-by-observation. To gain insight into which audio-visual prominence cues human speaker actually use, an analysis-by-observation test was conducted. Twenty (Dutch) subjects were asked to pronounce nonsense words consisting of three CV (consonant vowel) syllables: /ma ma ma/ and /ga ga ga/.[11] In each utterance, subjects had to emphasize one syllable. To achieve this, they were given cards with the three syllables, one of which was printed in upper case. The text on a card could be, for instance, "ma MA ma," which indicated that the second syllable should be pronounced with more emphasis than the other two. Subjects were not instructed about the kinds of cues they could use for emphasizing a syllable. They were given six cards in total (2 words $\times$ 3 stressed syllables). After looking at the top card, they were asked to pronounce the word printed on this card while looking into the camera. They did so in two different conditions: *neutral* and *exaggerated*. This process was repeated for each of the six cards, which resulted in 12 utterances per speaker (240 utterances in total).

As expected, almost all speakers used verbal cues to stress the designated syllable, but many speakers used visual cues as well. See figure

1.2 for some illustrative screen shots. Two findings are particularly noteworthy for the purposes of this chapter. First, speakers clearly *differ* in the kind of visual cues they use. Nine out of the 20 speakers indeed raise their eyebrows when uttering the stressed syllable (at least occasionally), while four speakers would use head movements. Interestingly, a recent production study by Keating et al. (2003) showed clear correlations between phrasal stress (the kind of prominence related to focus) and *both* head and eyebrow movements. Second, the most obvious audio-visual cue in the exaggerated condition is that speakers articulate more clearly when pronouncing the stressed syllable. This could be observed for 18 out of the 20 speakers. See Keating et al. (2003) for interesting related results on perceptual relevance of visual cues in the mouth area and Erickson et al. (1998) for correlations between jaw opening and accent.

To find out the relative contributions of the visual and the auditory cues for prominence, a perception test was conducted. Five speakers from the 20 were selected (we used those speakers who always looked in the camera and always produced the utterances with emphasis on the designated syllable). Their utterances were offered to three groups of 15 subjects in three different experimental conditions: one group saw the utterances as they were recorded (audio+vision), one group only heard the speech (audio) and the last group only saw the speakers (vision). All subjects were asked to determine which of the three syllables was the emphasized one. As expected, in both the audio+vision and in the audio condition, subjects were very good at determining the stressed syllable (97.1% and 97.3% correct, respectively). In the vision condition subjects scored significantly less good, confirming our earlier observation that auditory cues are more important for the perception of prominence than visual ones. Nevertheless, subjects in this condition performed still surprisingly good, with overall 92.89% correct guesses. What this indicates is that there are clear visual cues for prominence besides the well-known auditory ones.[12]

The analysis-by-observation approach does not suffer from the potential problems that may plague analysis-by-synthesis. Still, the approach cannot give us all the information that we would like to have. In particular, while the perception test clearly shows that there are audio-visual cues that Dutch people may use when interpreting an utterance (e.g., to detect prominence), we do not know *which* cues people actually use. In fact, one way to find out would be using an analysis-by-synthesis experiment. This illustrates that the combination of analysis-by-observation with analysis-by-synthesis is a good way to gain insight in functions of audio-visual prosody, because it enables us to get insight in which

cues human speakers employ, but also in how human listeners interpret stimuli which include these cues.

Acknowledgments

This research was partly conducted within the VIDI-project "Functions of audio-visual prosody (FOAP)", sponsored by the Netherlands Organization for Scientific Research (NWO). Parts of the Dutch results have been presented at the Workshop on Coordination and Fusion in Multimodal Interaction (Dagstuhl, November 2001), Speech Prosody (Aix-en-Provence, April 2002), and ICSLP (Denver, September 2002). The Italian results appear here for the first time. We would like to thank Han Noot, Zsofi Ruttkay and Wieger Wesselink for their help in making the animations, and Cinzia Avesani and Jeannine de Raad for their help in carrying out the Italian experiments. Iris Boshouwers has been a great help in the analysis-by-observation study. We have benefitted from comments by Matthew Stone, Mariët Theune, Loredana Cerrato and an anonynous reviewer.

Notes

1. Embodied Conversational Agents (e.g., Cassell et al. 2000) are also referred to as Virtual Humans (e.g., Gratch et al. 2002) or Talking Heads (e.g., Rubin and Vatikiotis-Bateson 1998). In this chapter we mainly concentrate on Talking Heads although the methodological part of the story is applicable to any kind of embodied agent.

2. There is even some evidence that the presence of Perlin noise results in animations which are slightly *less* functional than animations without Perlin noise, since subjects are somewhat more likely to miss potentially informative facial cues when random movements are present (van de Laar 2003).

3. Their usage of the term *flash* for an eyebrow movement does not coincide with Ekman's usage of the term. The flashes of Cassell et al. (2001) are really batons or underliners, while Ekman's flashes refer to repeated brow raises which do not coincide with speech (i.e., emblems).

4. The Dutch and Italian subjects came from different parts of The Netherlands and Italy, respectively. For methodological reasons, it would have been better to have Italian subjects from Tuscany only (the dialect of the speakers), as Italian dialects are known to vary regarding their intonation structures. Unfortunately, we were unable to find enough Tuscanian subjects. However, since the Italian results are so unequivocal, we suspect that the results would not have been dramatically different from the ones reported here.

5. Of course, it might be that our Dutch and Italian subjects would have preferred an animation *without* eyebrow movements, but the experiment was not designed to test this. It would be interesting, however, to redo the first experiment including animations without eyebrow movements, and ask subjects for their preference.

6. The synthetic voices were added to see to what extent naturalness of the voice influences the perception of focus. Arguably, a human voice has more natural and better sounding prosody, but a synthetic voice might be more suitable as the auditory counterpart of a synthetic character. It turned out that this was not the case: the results for the 4 human voices did not differ significantly from the results for the 2 synthetic voices.

7. The order of presentation in this chapter is a historical falsification. The Dutch experiments were carried out first. This was done in two steps: first, we performed the functional analysis, and in a later stage we did the subjective preferences and perceived prominence tests to get a better understanding of the results of the functional analysis. The Italian experiments were done at a later date, but in the same order as the Dutch experiments (i.e., 3, 1, 2).

8. Ekman (1979) also mentions other conversational functions of brows besides accenting, in particular they may cue punctuation, question marks, word search, and agreement between dialogue participants. None of these functions seems intuitively right for the stimuli used in the three experiments. It is interesting to observe that all of the functions Ekman mentions are also typical functions of verbal prosody.

9. It seems that a study along these lines has been carried out by O'Sullivan and Eyman (1978). We have not been able to consult this paper, but O'Sullivan (p.c.) informed us that they compared AU 4, AU 1+4 and AU 1+2+4, combined with neutral statements, and found that different brows affected the interpretation.

10. Along the same lines, it might be worth investigating subtle interactions of visual cues with other auditory cues to prominence, such as different pitch accent types and voice intensity.

11. The motivation to select /m/ and /g/ was that the former phoneme is pronounced in the front of the articulatory channel, while the latter is pronounced in the back. It was hypothesized that the /m/ is visually easier to perceive than the /g/. This turned out not to play a role for prominence perception.

12. More details about this and some related experiments will be given in a sequel to this paper.

References

Bolinger, D. (1985). *Intonation and its parts*, London: Edward Arnold.

Birdwhistell, R. (1970). *Kinesics and context*, University of Pennsylvania Press.

Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (2000). *Embodied Conversational Agents*, Cambridge, MA: The MIT Press.

Cassell, J., Vihjálmsson, H., Bickmore, T. (2001). BEAT: The Behavior Expression Animation Toolkit, *Proceedings of SIGGRAPH'01*, Los Angeles, pp. 477–486.

Cavé, C., Guaïtella, I., Bertrand, R., Santi, S., Harlay, F., Espesser, R. (1996). About the relationship between eyebrow movements and F_0 variations, *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, Philadelphia, pp. 2175–2179.

Chafe, W. (1974). Language and consciousness, *Language* 50: 111–133.

Condon, W. (1976). An analysis of behavioral organization, *Sign Language Studies* 13:285–318.

Cutler, A. (1984). Stress and accent in language production and understanding, in: *Intonation, accent and rhythm. Studies in Discourse Phonology*, D. Gibbon and H. Richter (eds.), Berlin: de Gruyter, pp. 77–90.

Cruttenden, A. (1997). *Intonation*, 2nd edition, Cambridge: Cambridge University Press.

Darwin, Ch. (1872). *The Expression of the emotions in man and animals*, New York: Philosophical Library.

Doherty-Sneddon, G., Bonner, L., Bruce, V. (2001). Cognitive demands of face monitoring: Evidence for visuospatial overload, *Memory and Cognition* 29(7): 909–919.

Efron, D. (1941). *Gesture and environment*, New York: King's Crown Press.

Eibl-Eibesfelt, I. (1972). Similarities and differences between cultures in expressive movements, in: *Non-verbal communication*, R. Hinde (ed.), Cambridge: Cambridge University Press.

Ekman, P. (1979). About brows: Emotional and conversational signals, in: *Human ethology: Claims and limits of a new discipline*, M. von Cranach, K. Foppa, W. Lepenies, D. Ploog (eds.), Cambridge: Cambridge University Press, pp. 169–202.

Ekman, P., Friesen, W. (1978). *Facial Action Coding System*, Palo Alto: Consulting Psychologists Press, Inc.

Erickson, D., Fujimura, O., Pardo, B. (1998). Articulatory correlates of prosodic control: Emotion and emphasis, *Language and Speech* 41 (3-4): 399–417.

Granström, B., House, D., Lundeberg, M. (1999). Prosodic cues to multimodal speech perception, in: *Proceedings 14th International Conference of the Phonetic Sciences (ICPhS)*, San Francisco.

Granström, B., House, D., Swerts, M. (2002). Multimodal feedback cues in human-machine interactions, in: *Proceedings of Speech Prosody 2002*, Aix en Provence, France, pp. 347–350.

Gratch, J., Rickel, J., André, E., Badler, N., Cassell, J., Petajan, E. (2002). Creating interactive virtual humans: Some assembly required, *IEEE Intelligent Systems*, 17(4):54–63.

Hirschberg, J. (1993). Pitch accents in context: predicting intonational prominence from text, *Artificial Intelligence*, 63: 305–340.

Keating, P., Baroni, M., Mattys, S., Scarborough, R., Alwan, A., Auer, E., Berstein, L. (2003). Optical phonetics and visual perception of lexical and phrasal stress in English, in: *Proceedings 16th International Conference of the Phonetic Sciences (ICPhS)*, Barcelona, Spain, pp. 2071–2074.

Krahmer, E., Swerts, M. (2001). On the alleged existence of contrastive accents, *Speech Communication* 34:391–405.

Krahmer, E., Ruttkay, Zs., Swerts, M., Wesselink, W. (2002a). Pitch, eyebrows and the perception of focus, in: *Proceedings of Speech Prosody 2002*, Aix en Provence, France, pp. 443–446.

Krahmer, E., Ruttkay, Zs., Swerts, M., Wesselink, W. (2002b). Perceptual evaluation of audio-visual cues to prominence, in: *Proceed-*

ings of International Conference on Spoken Language Processing (ICSLP'02), Denver, CO.

Ladd, D. (1996). *Intonational phonology*, Cambridge: Cambridge University Press.

van de Laar, L. (2003). *Influence of eyes on the interpretation of utterances of embodied conversational agents: an experimental inquiry*, MA thesis, Tilburg University.

Morgan, B. (1953). Question melodies in American English, *American Speech* 2:181–191.

Nass, C., Isbister, K., Lee, E. (2000). Truth is beauty: Researching embodied conversational agents, in: *Embodied Conversational Agents*, J. Cassell, J. Sullivan, S. Prevost, E. Churchill (eds.), Cambridge, MA: The MIT Press.

O'Sullivan, M., Eyman, J. (1978). The signal value of eyebrow movements in conversation, in: *J. Western Psychological Association Convention*, San Francisco.

Pelachaud, C., Badler, N., Steedman, M. (1996). Generating facial expressions for speech, *Cognitive Science* 20:1–46.

Perlin, K. (1995). Real time responsive animation with personality, *IEEE Transactions on Visualization and Computer Graphics*, 1(1):5–15.

Rimé, B., Schiaratura, L. (1991). Gesture and speech, in: *Fundamentals of nonverbal behavior*, R. Feldman, B. Rimé (eds.), Cambridge: Cambridge University Press, pp. 239–281.

Rubin, P., Vatikiotis-Bateson, E. (1998). Talking heads, in: *International Conference on Auditory-Visual Speech Processing (AVSP'98)*, D. Burnham, J. Robert-Ribes, E. Vatikiotis-Bateson (eds.), pp. 233–238.

Ruttkay, Zs. (2001). Constraint-based facial animation, *Journal of Constraints* 6:85–113

Ruttkay, Zs., Noot, H. (2000). Animated CharToon Faces, *Proceedings of NPAR 2000 - First International Symposium on Non Photorealistic Animation and Rendering*, pp. 91–100.

Sanderman, A., Collier, R. (1997). Prosodic phrasing and comprehension, *Language and Speech* 40(4):391–409.

Swerts, M., Krahmer, E., Avesani, C. (2002). Prosodic marking of information status in Dutch and Italian: A comparative analysis, *Journal of Phonetics* 30(4): 629–654.

Terken, J. (1984). The distribution of pitch accents in instructions as a function of discourse structure, *Language and Speech* 27:269–289.

Terken, J., Nooteboom, S. (1987). Opposite effects of accentuation and deaccentuation on verification latencies for Given and New information, *Language and Cognitive Processes* 2 (3/4):145–163.

Zappa, F. (1989). *The Real Frank Zappa Book*, New York: Poseidon Press.