Listening Heads

Dirk Heylen

University of Twente
The Netherlands
d.k.j.heylen@ewi.utwente.nl

Abstract. In [1] we discussed functions of head movements and gaze. In this paper, we will go into more depth in the classification of various head movements: how they are distinguished in both formal and functional terms. We look at the distribution of a selection of primitive head movements and their related meanings and the way they are composed out of smaller units. This catalogue is not intended to be exhaustive and does not take into account the component items from other modalities such as speech, facial expressions, gestures and posture. The research is motivated by our desire to build systems that can interpret and mimick expressive human behaviour.

Keywords: head movements, corpus analysis, embodied agents.

1 Introduction

The way people move their head when they talk and listen to each other is one of the most fascinating forms of nonverbal communication to study because the apparent restriction in parametric variation does not seem to put a bound on the number of different expressive gestures that can be made and the diversity of functions that are served by them. The repertoire of behaviours is defined by only a few parameters but movements can be combined into one or can differ in velocity and amplitude. Movements may also be concatenated in different ways giving rise to configurations with different meanings. Depending on the context the meanings of a movement may change as well. There are clearly more head movements than simple nods or shakes. Also the meanings conveyed may be complex. Even if all the shakes in every context investigated by [2] share a core meaning of *negativity*, this is not to say, that this core is the only meaningful component to them nor that various forms of shaking can be exchanged from one context to the other. As Kendon writes:

> The head shake is used in many different discourse contexts where [...] it yet comes to have a very different force, depending upon how this theme of negation combines with the other semantic themes that are also being expressed. [p. 148]

In this paper, Kendon goes to some effort to argue that the general theme of negation is present in each of his examples, while at the same time showing the subtle variations in meanings that the contexts gives rise to.

I. Wachsmuth and G. Knoblich (Eds.): Modeling Communication, LNAI 4930, pp. 241–259, 2008.

> There is no doubt that these variations in performance intersect with and modify the meaning of the gesture. Furthermore, the head shake may be combined in various ways with other movements, such as movements of the eyelids, mouth expressions, and gestures of the shoulders, such as the raising of the shoulder in a 'shrug' or with manual gestures of various kinds. [p. 149]

But this analysis only scratches the surface, as Kendon points out himself. In general, research into the 'morphology', 'syntax', 'semantics', and 'pragmatics' of head movements is very incomplete. Compared to the analysis of other communicative signals such as spoken language, facial expressions and gestures, the interest in head movements has been marginal. Our aim is to give these humble movements the attention that they deserve as we believe they are of interest in both applications that analyse and interpret human behaviours as well as in applications of virtual humans where they need to be generated. They may be easier to recognize by computer vision techniques than facial expressions, but are still highly informative. This means that the ratio between detection and interpretation quality and thus the knowledge gain might be much higher than with other verbal and nonverbal cues.

With respect to the generation side, the proper selection and execution of head movements is important as they carry different kinds of information both while an agent is talking and listening. While talking, head movements play an important role in providing cues as to the information structure of the sentence. They are used to put emphasis on important parts of the utterance and can thereby signal higher or lower attitudinal involvement of speakers with the things they are saying. Besides their information-bearing impact, head movements play an important role in visual prosody improving the auditory speech perception ([3]). We will discuss the functions and determinants of head movements in Section 4.

Nodding and shaking are typical actions that listeners are commonly known to engage in. Head movements are part of a small feedback loop between speakers and listeners. They provide on-line, real-time feedback to speakers on all kinds of matters related to the speech: showing recognition of perception, understanding, agreement, attitude and engagement in general. They are also known to enter in behaviours of listeners that fall under the heading of interactional synchrony. Hadar and colleagues [4] report that approximately a quarter of all the head movements of the listeners in the conversations they looked at occurred in sync with the speech of the interlocutor.

Because of the way head movements function in interaction, we have decided to pay special attention to them in the Sensitive Artificial Listener project [5]. In the next section we will outline the goals and challenges of this project. We summarise the results of a typical experiment that we have carried out. This is to motivate the kind of research on head movements we present in the rest of the paper. This concerns ongoing work on the classification and interpretation of head movements. We first describe the data that we have used for this paper (Section 3). Next, in Section 4, we review the literature on head movements

and their functional determinants. This leads to the presentation of some types of head movements that we distinguish and of their distributional properties (Section 5).

By the end of this paper, we hope to have shown the richness of head movements in face-to-face conversations and the importance of studying them for developing responsive embodied conversational agents.

2 The Sensitive Artificial Listener

Gumpertz [6] points out the reciprocal nature of the actions of speakers and listeners as a defining characteristic of communication.

> Communication is a social activity requiring the coordinated efforts of two or more individuals. Mere talk to produce sentences, no matter how well formed or elegant the outcome, does not by itself consitute communication. Only when a move has elicited a response can we say communication is taking place.

Responses to talk by recipients can take the form of a subsequent move in the next turn, but also the behaviors of the non-talking participant in the conversation displayed during the turn of the speaker can count as some kind of "response". They provide the speaker with feedback on whether and how the listener perceives, attends to, or understands the speaker's utterance and on the way in which the message is received in general, i.e. how the beliefs, attitudes and the affective state of the recipient is changed. These cues and signals enable communicators to synchronize the communicative actions of turn-taking. They are also involved in grounding processes and in the building of rapport.

The bulk of work on the generation of communicative behaviors of embodied conversational agents is on generating the appropriate non-verbal behaviors that accompany the speech of the embodied agent. The generation of the verbal and non-verbal behaviors to display during the production of speech by another actor, that is the behavior of a listening agent, has received less attention. A major reason for this neglect may be the inability of the interpretation modules to construct representations of meaning incrementally and in real-time, that is contingent with the production of the speech of the interlocutor. Another reason may be that the production of language and speech by speakers seems more important than the way it is perceived, which can be seen as a rather passive and reactive kind of behavior.

When one takes a look at the various modalities through which embodied agents can communicate, it appears that language, facial expressions, gestures and gaze are the main kinds of expressive behaviors that have been studied so far. Posture and head movements form another group of nonverbal behaviours, that are very informative about the intentions, attitudes, emotions and the mental state of interlocutors, in particular, "auditors", but these have been less widely studied. In the Sensitive Artificial Listener project we specifically focus on these behaviours.

A peculiarity of the SAL project is that a person can talk to four different characters with individual personality profiles. The character Poppy, for instance, is cheerful and optimistic and will try to cheer up the interlocutors when they are in a negative state and be happy for them when they are in a positive state. Obadiah, on the other hand, is gloomy and passive and will say things with the opposite effect. The voices, created by (amateur) actors are also quite expressive. The choice of talking head should match this, as should their nonverbal behaviors. In order to build up a repertoire of feedback and other expressions that match the personalities, we have carried out several tests in which we have asked participants to judge small video clips of an embodied conversational agent displaying various kinds of feedback behaviour ([5]).

In one of the studies ([7]) we varied the gaze behaviour and the head movements of a character to create differences in the judgment of personality. Earlier experiments carried out by Fukayama et al. [8] for gaze and Mignault and Chauduri [9] for head movements formed the basis for this study, complemented by many other works on gaze[1] and head movements.

Similar to the study in [8], a probabilistic model of the behaviours was implemented that determined the gaze of the RUTH talking head ([14]). We limited the variation in movements by fixing the head tilt. We tried to combine some of the results of the studies by Fukayama et al. and Mignault & Chauduri in order to model the behaviours for a happy, friendly, unobtrusive, extrovert agent (A) and for an unhappy, unfriendly and rather dominant agent (B).

The head tilt for A was set to $+10^\circ$ (raised). According to the study by Mignault and Chauduri a head tilted upwards can be perceived as more dominant which is not exactly what we wanted, but it also has an effect on the impression of happiness, which is what we aimed for. The amount of gaze was set at 75% and the mean gaze duration was kept short (500ms). We expected this would create the impression of engagement, friendliness and liking. In the [8] experiment short gaze durations were associated with friendly characters. Gaze aversion for A was downwards, which is associated with submissive rather than dominant personalities.

For B the head tilt was 0°. According to Mignault and Chauduri this may lead to low scores on happiness. With respect to gaze, the amount of gaze was kept at 75% just as with A, but the mean gaze duration was set at 2000ms, resulting in longer periods of gaze. We expected that this would create a rather dominant, unfriendly impression. Gaze aversion for B was to the right.

For both A and B two animations were created, one with smaller (A1, B1) and one with larger movements (A2, B2). Each animation lasted 40 seconds. The four movies were shown to 21 participants, divided into three groups that were each shown all the movies but in a different order: (A1 B1 A2 B2; B1 A1 B2 A2; A2 B2 A1 B1). The ordering did not have an effect. To rate the impressions we had the participants fill out a questionnaire for each movie consisting of a rating on a 7-point scale for 39 dutch adjective pairs with the following translations.

[1] In particular, [10],[11], [12], and our own work [13].

extrovert - introvert, stiff - smooth, static - dynamic, agitated - calm, closed - open, tense - relaxed, sensitive - insensitive, polite - rude, suspicious - trusting, interersted - uninterested, credbile - incredible, sympathetic - unsympathetic, self-confident - uncertain, cold - warm, weak - strong, selfish - compassionate, formal - informal, winner - loser, thougthful - reckless, unattractive - attractive, organized - disorganized, unfriendly - friendly, reliable - unreliable, refined - rude, involved - distant, flexible - linear, amusing - boring, attentive - absent, lazy - industrious, inactivy - lively, optimistic - pessimistic, happy - depressed, loving - unloving, empathetic - unempathetic, dominant - submissive, aggressive - timid, stubborn - willing, enterprising - passive, realistic - artificial.

Factor analysis reduced the number of dimensions to the following 8 factors.

1 absence, unfriendliness, rudeness
2 submissive, weak, sensitive
3 warm, energetic
4 dull, drained
5 unreliable
6 rigid, static, linear
7 informal
8 attractive

A scores higher on Factors 2 (submissiveness) and 5 (unreliable). B scores significantly higher on Factors 1 (absence, unfriendliness...) and 4 (dullness). Large movements create a more unfriendly impression (Factor 1) whereas small movements score significantly higher on Factor 2 and 8, that is, the smaller movement animations are considered more submissive, but are also more attractive.

All in all the differences in the generated behaviours resulted in several impression values that fitted the intended personality of the agents. However, this way of designing behaviours by combining functions associated with behaviours as mentioned in the literature poses many problems. If one tries to achieve an effect on various impression variables, a particular behaviour may be very well suited for yielding good scores on one variable but mediocre scores on another Also the combination of two behaviours may yield a combined effect that is different from what might be expected from the descriptions of the behaviours considered independently. Adding yet more behaviours - such as speech, for instance - may change the results again. Context plays an important role as well. The literature reports on functions and impressions, derived from data in a particular context which can be very different from the context of use that we are considering. Furthermore, the precise set of impression categories that one is aiming for may not correspond exactly to the categories used in the studies in the literature. Expressions are ambiguous and fit more than one category.

In a study related to the one described in the previous section, we looked at how people would interpret a combination of facial expressions, gaze shifts and

head movements by an embodied agent ([15]). We asked people to think of the small video clips that were shown as the behaviour of an agent that was listening to someone speaking. In some cases there was a clear tendency to associate a label with a particular head movement across the various contexts in which it was put. For instance, the nods were associated with *agree*, *understand*, labels, in various constellations. However, when combined with a smile it was mostly associated with *like* and *accept*. In this case the smile added a nuance. The case of shakes was similar, with dislike popping up when the nods combined with a frown. With the head tilt we found mostly associations with *disbelief* or *not understand* labels, which might be combined with *interest*, *surprised*, and *bored* depending on the way the tilt was executed and the other expressions it combined with. These judgement studies are one way to study the contribution of head movements to the communicative repertoire. Another way to get more insight is by looking at real data of real people as we will do here.

3 Head Movements in Context

To study why people move their heads in conversations the way they do, we are gathering together a video corpus of dyadic and multi-party conversations. The corpus consists of available material from the Humaine and AMI data sets, both EU projects, complemented by new recordings. For the presentation in this paper, the examples are limited to a Dutch television program[2].

A central problem with head movements, just as with many other movements it is not a priori clear how to segment the continuous stream into meaningful segments. It is possible that two meaningfull stretches overlap (partially). To take an example from our data set, there is a moment where a person turns his head to the left to look from one person to another and while turning also displays tiny shakes. This combines two kinds of movements with different functions that will often occur on their own.

One way to characterise units of analysis is by following Birdwhistell who analysed body movement into different hierarchically organized units analogous to the way in which in speech and language phonemes combine into syllables or morphemes, morphemes combine into words and words into phrases.

> [K]inemes combine to form *kinemorphs*, which are further analyzable into *kinemorphemic* classes which behave like linguistic morphemes. These, analyzed, abstracted, and combined in the full body behavioral stream, prove to form *complex kinemorphs* which may be analogically related to words. Finally, these are combined by syntactic arrangements, still only partially understood, into extended linked behavioral organizations, the *complex kinemorphic constructions*, which have many of the properties of the spoken syntactic sentence. [16, p. 101]

[2] The talkshow B&W that was broadcasted October 3, 2002.

Fig. 1. B nose-tracked

Birdwhistell distinghuishes three kinemes of head nod (the "one nod", the "two nod", and the "three nod"), two kinemes of lateral head sweeps "one" and "two" sweep), one kineme of head cock and one of head tilt. Furthermore, he distinguishes several connective kinemes that use the entire head. Kinemes such as these have allokines that differ in intensity, extent and duration. In Section 5 we distinguish some further kinemes.

Several movements may combine to form complex units. A head nod, typically consists of multiple repeated up and down movements. But there are other cases as well. Let us consider a specific example and determine some options to cut up the movement stream in meaningful units.

In the picture above we see a frame of a video in which the person's nose has been tracked. The white line shows the trace of the tip of the nose as it moved from the start of the line (now just left from his mouth) to the position the nose is currently in. This complex movement was executed in about 3 seconds.

The snapshot of the annotation window in Figure 3 shows some of the annotations for this episode(the darker shaded part in the picture). It shows four levels of annotation. The top level indicates the camera position. It shows the start of the camera position on the restaurant chef BLAAUW just before the 23th second of this fragment. The video continues to show BLAAUW beyond the segment shown here. The second line ("ja maar er zij(n)" // "yes but there are") is part of the transcription for the host of the show. The third line shows the annotations for the head movements that we have visualised in the picture. These are presented in the table below. Finally, the line below this provides the annotations of what the chef is saying. The text is not completely shown. What he says is the following, with each line starting a new (possibly incomplete) sentence.

en, ik ben restaurant (sic)	and, I am restaurant (sic)
ik werk met smaken	I work with flavours
ik werk met	I work with
mijn gerechten	my dishes

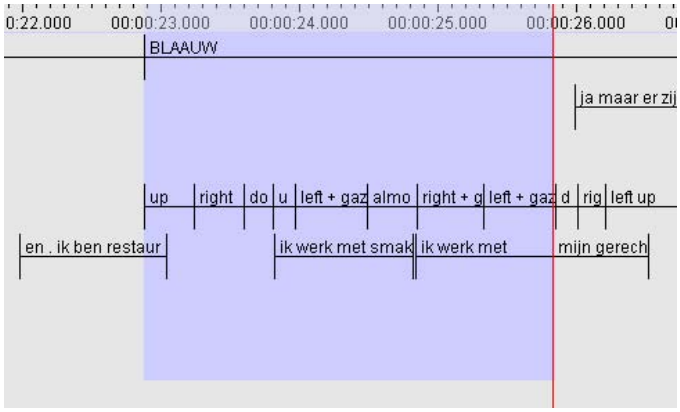

Fig. 2. Annotation of nose-tracked fragment

As one can see, the head movement in this analysis is annotated in 8 parts. We present the same information in tabular form below with the timing information (start time, end time and duration) given in milliseconds .

44 up	22.882	23.240	.358
45 right + gaze (towards person opposite)	23.240	23.600	.360
46 down right	23.600	23.810	.210
47 up	23.810	23.970	.160
48 left + gaze (towards host)	23.970	24.490	.520
49 almost stationary	24.490	24.850	.360
50 right + gaze (towards person opposite)	24.850	25.330	.480
51 left + gaze (towards host)	25.330	25.850	.520

The mean duration of an annotation is .371 seconds; the shortest movement lasts .160 and the longest .520 seconds. The general guideline for these segmentations was that a new unit starts as soon as the head movement changes direction or the movement stops. We call this an *elementary movement*. The way the movement is segmented by elementary movements in this case is mechanical and does not necessarily identify the meaning-bearing segments.

The labels given for the movements are simple and quite straightforward. First the head moves upwards (44) than sideways to the right if we take the perspective of the viewer (45). This is followed by a small movement further to the right (46) which is slightly more downwards. Next the head moves up again a little bit (47) and then it turns to the left (48). For about 360 milliseconds the head remains almost stationary followed by a turn right (50) and a turn to the left (51). At the start of this last movement the head moves up first for a couple of frames before it turns left. This pattern of movements occurs several times in the video, both by the same speaker but also by others.

There are gaze shifts associated with movements 45, 48, 50, 51. These are all either right or left movements in which BLAAUW alternates between looking at the host and the person sitting opposite. The lateral movements may thus be explained by the desire to alternate the gaze between these two persons indicating that they are intended addressees of the utterance and possibly also to elicit feedback from them. The first movement to the right follows the end of the first sentence. The movement back is accompanied by the second utterance, after which there is a small pauze in the speech and the head remains almost stationary (49) with the gaze fixated on the host of the show. There is thus a close correspondence between the head movements and the production of the utterances that we can explain by mechanisms of addressing and requests for feedback.

In Section 5 we will identify more typical head movement patterns and the functions with which they are associated. We will introduce the functional specifications in the next section.

4 Functional Schemes

What explains all the head movements that people make when they are in a conversation? Of course there is not one single function that underlies all the kinds of movements that are made. On the functional level we combined a number of information sources in order to arrive at a list of determinants that was as exhaustive as possible. The most important sources of inspiration that we will refer to in the following sections were [2], [17], and [18]. Other works that we have relied on are amongst others [4] and [19], [20] and [21]. In the next section we introduce the kinds of labels that have been introduced in the literature to describe the functions and meanings of head movements. We conclude by the taxonomy we are using in most of our annotations.

When one looks at classifications of determinants of head movements during conversations, there are three general classes one can distinguish. The first involves determinants that are related to other kinetic behaviours, mostly induced by the rhythm of the speech. The second, and probably major determinant of head movements involves movements made in order to properly attend to an object or a person. The third class tells something about the way the movement might perceived by others or is intended to be perceived by others: the signalling function of a movement. Movements can tell something about stress (prominence, attitude, the structure of the interaction (turn taking, for instance), etc.

These classes are not mutually exclusive. Any behaviour can function as a signal when it is perceived by another observer, of course. Also the way in which the signal works can be different. If someone turns his head to look at someone the intention behind it may be caught by an observer and the movement thus works as a signal. But the way the head is turned may also have an expressive value. Through its form of excecution it may reveal anxiety, desire, or anger. Other movements may act as symbols, conventionally agreed upon signals.

It is important to note that a single movement may be determined by a combination of functions from different types and also that there are some systematic

relations between the various functions. For instance, a movement with the head and a shift in eye gaze to pay attention to a particular person or object may - by conventional signalling procedures - also have the effect of a deictic pointing gesture.

As a first way to get an idea of what kinds of functions a head movement can serve we can turn to Kendon ([2]) whose study focusses on the function and placement of head shakes. He distinguishes various types of elements to which they may be related.

1. *Semantic*: The referential content of the utterance. A shake may be equivalent to a statement ("no"), or to some kind of modifier: expressing negation or intensification
2. *Pragmatic*: Its illocutionary force. A shake may express denial, for instance.
3. *Stance*: The speaker's own stance with respect to what is spoken. In the case of shakes these can be attitudes such as lack of knowledge, uncertainty, doubt, or disapproval.
4. *Role*: The dialogic role of the utterance within the interactional sequence in which it is occurring.

These functions are all largely equivalent to linguistic functions. Similar and other functions are also presented in the paper by McClave ([17]), which we now turn to to get a more complete picture of the functions. In discussing the literature on head movements McClave mentions the following functions:

a. Head nods function as stress markers ([16])
b. Turning the head can have a deictic function ([16])
c. Head movement characteristics (speed and amplitude) correlate with the speech characteristics and can be said to serve a motoric functions ([19]); including the movements made at dysfluencies in the speech.
d. Postural shifts of the head mark boundaries: phonological, syntactic, semantic and interactional like turns ([2], [20]); listeners can signal the start of a response ([20])
e. Head movements can control interpersonal interaction (approach, withdrawal) ([22])
f. Particular patterns of movements can vary according to the discourse function of an utterance ([23])
g. Listeners can use head movements to provide feedback ([24], [21])

Again, these classes are not mutually exclusive. Furthermore, in the study she reports on in the paper she elaborates on the following functions of head movements, which partly overlap with the functions described before in the literature.

h. Shakes and sweeps can signal *semantic concepts* such as negation, inclusivity, intensification, and uncertainty.
i. Head movements can have *narrative functions*:
    1. marking the switch from indirect to direct discourse
    2. expressing mental images of the characters

   3. for deixis and the referential use of space
   4. marking alternatives or elements from a list

j. Small lateral shakes occur when doing a lexical repair as if erasing or wiping away something are related to *cognitive processing*
k. Head nods of speakers may serve the *interactional function* to elicit feedback from listeners.

McClave notes that head movements can play a role in interpersonal interaction ([e.]). This effect that different movements or positions of the head on the person who perceives them has been taken into account in several studies on perception of facial displays and other forms of body movement ([25], [26], [27], [28], [29], [30], [31], [9]). Head positions and movements can give the impression of shyness, dominance, approachability or distance. Some of these variables we used in the study reported on in the previous section.

In our annotations we group the various functions as follows (with a reference to the list of functions above). Note that we do not provide an exhaustive list of all the possible meanings but indicate what we mean by giving a few examples.

1. Motoric / speech [c.]
   (a) movements resulting from speech production
   (b) movements resulting from acts like sneezing
2. Markers (linguistic, conversational) [d.]
   (a) Information structure
      i. given/new, topic/comment...
      ii. stress, prominence [a.]
   (b) Discourse structure [f.]
      i. punctuation: question markers, end of sentence
      ii. discourse relations: direct/indirect speech, listing of alternatives... [i.]
   (c) Interaction structure [g.], [k.]
      i. grounding: backchannel requests, backchannels
      ii. turn-taking
      iii. addressing
3. Symbolic, semantic, pragmatic
   (a) Semantic: propositional, adjectival, adverbial [h.]
   (b) Pragmatic: illocutionary force, deixis [b.]
4. Cognitive processes [j.]
   (a) thinking, memory search
   (b) attention (focus of)
5. Stance, attitude, emotion [e.]
   (a) epistemic markers (doubt, certainty...)
   (b) emotion related states: nervousness, anxiety
   (c) interpersonal attitudes: dominance, shyness

In the example of the lateral movements we gave in the previous section, the major determinant of the head movement seems to lie in the interactive function: turning the head from one participant to the other to indicate they are considered addressees of the utterance and to elicit a grounding act. This builds on the focus

of attention function. Its relation with speech indicates also the boundaries of segments. More illustrations of this classification scheme will be presented in the section to come where we identify some typical head movements. One of the questions we address is whether functions are articulated through different forms of movements. To a certain extent this is obvious: nods signal yes, shakes signal no; but for other movements and functions the relation may be more subtle.

5 Form and Function

Several researchers have observed that sometimes differences in the formal execution of a head movement may distinguish different functions. In [4], an attempt to relate form and function has been made for a limited number of head movements. They show that kinematic properties such as amplitude, frequency and cyclicity distinguish between signals of 'yes and no' (symmetrical, cyclic movements), anticipated claims for speaking (linear, wide movements), synchrony movements occurring in phase with stressed syllables in the other's speech (narrow, linear) and movements during pauses (wide, linear). Kendon notes somethings similar.

> Head shakes vary in terms of the amplitude of the head rotations employed, in the number of rotations and in the speed with which these are performed. There is no doubt that these variations in performance intersect with and modify the meaning of the gesture. ([2, p. 149])

In the following section we will analyse the data from the B&W corpus and discuss some of the mappings of head movement forms with functions; not just head shakes but also other kinds of movements. We hope this adds to a better understanding of how the variations in form give rise to variations in meaning.

5.1 Basic Movements

Our analysis of the nose-tracked data in the B&W corpus allowed us to identify some seven classes of basic movements that differ in form and function[3].

1. Tiny nods
2. Tiny shakes
3. Lateral sweeps (medium sized)
4. Moves lateral or horizontal (medium sized)
5. Waggle
6. Various head repositionings
7. No motion

These basic movements can occur in different contexts with different functions. They can also combine into larger meaningful composite movements. We will now provide some details about form and functions for each of them.

[3] Not surprisingly, such a classification resembles classifications by others, at least with respect to form, rather closely [32]. Here we provide some more details on execution and function. A more detailed comparison between the various classifications made in the literature and the one made here is important as this might reveal subtle but important differences.

Fig. 3. Fragment showing a downward segment of a nod

Nods. Almost all nods in the corpus are tiny movements only a few are slightly bigger but these also remain small. The elementary movements that make up nods like these typically last about one to three frames and cover about 15 pixels. To get an idea about the size and duration, the picture below shows the first elementary downward movement from a nod made by Witteman (the host of the show). This movement took 220 milliseconds.

A couple of times the size of one nod movement down was twice this length. Typically nods occur as a kind of beats, in synchrony with the speech and they express some kind of insistence. The small up and down movements often repeat a couple of times, as one would expect.

(Tiny) Shakes. The next snapshot from the annotation window shows a segment with several shakes.

The fragment first shows some longer movements (the first three movements annotated) followed by many small movements left and right. At first they correspond with a hesitation in the speech, but also the semantic context of the utterance before and after is negative "there is no alternative".

What we call shakes are similar in size and number of repetitions to the nods but instead of vertical movements these are horizontal. Other lateral movements, with different formal properties, also occur in the corpus but they will be called differently; see sweeps and moves below. Shakes occur in a number of contexts (with explicit or implicit negations or at hesitation points) and express several of the semantic functions mentioned in the literature: negation, inclusivity, intensification and uncertainty. Besides this semantic dimension the context in which they occur in our corpus is often one in which the speaker is agitated or insistent. The corpus also showed one or two instances of listeners displaying their head in synchrony with the nods and shakes of the speaker.

Sweeps. The head shake forms a particular kind of excursionary, horizontal movement to the left or the right and back again, after perhaps several cycles to the initial position. Kendon ([2]) points out that there are different kinds of lateral movements.

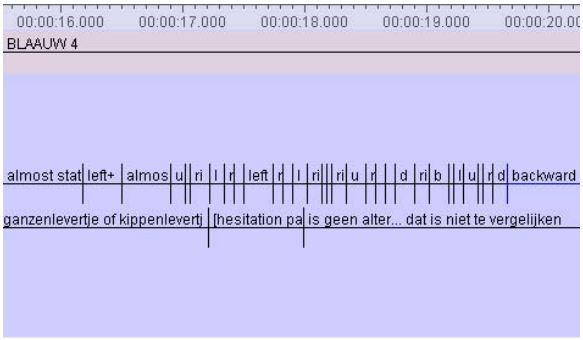

Fig. 4. Annotation of several movements including several shakes

> Head shakes can be distinguished from head turns of various other kinds (such as turning the head momentarily to the left or to the right to glance at something, or to accomplish the 'look-away' that is common as a speaker initiates a turn at talk or initiates a new unit of discourse within a turn) because they are always performed within the 'frame' of the orientation the person is using, which may be either the orientation in use for a current turn at talk, or the orientation that a person is using as a recipient of another's utterance.

What is not mentioned in this characterisation is that in what are classically called shakes the head typically moves as a pendulum from the starting position to the side and then back *beyond* the original position and then back again.

Lateral movements that are not counted as shakes not only differ in their configuration but also in their amplitude and speed. A relatively fast movement of medium size, we call a sweep. Sometimes there is a pause at the end of the sweep or a small repositioning of the head before the retraction takes place. A sweep often marks stress in argumentative discourse. The start and end positions of the sweep or retraction movement may also be adorned with tiny shakes. Besides the movement of the head the gaze position changes alongside (often the eyes blink at transition points). The head and eyes are directed towards the addresse at the start of the sweep and are turned away (mostly downwards) at the end of the sweep. Retraction of the head also restores the gaze. This makes them different from the lateral movements by BLAAUW described in Section 3, because they are faster.

Most sweeps combine a movement left or right with a downward movement. They function as *underliners*.

A snapshot from the annotation window makes the difference in size between sweeps and the (tiny) shakes immediately clear.

Moves (downwards or lateral). In the definition by Kendon above, the head shake is distinguished from head turns of other kinds. There are lateral movements that are not part of shakes and which we will not classify as sweeps either but which

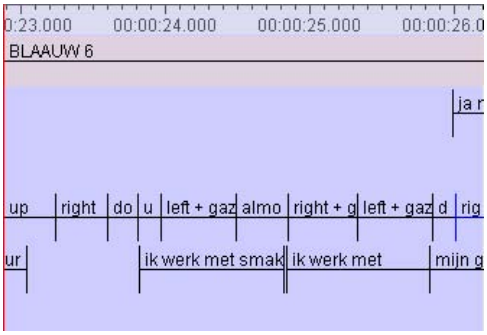

Fig. 5. Annotation illustrating sweeps

we call a move. An example of what we call a move is the turn away while listening depicted in the following picture by BLAAUW.

Here the picture shows a fairly slow movement where BLAAUW was looking at the speaker at first and then turned his head (and gaze) away. During this movement the speaker is finishing his question. Immediately after this the listener retracts to the initial position with a fast movement and after a slight motionless phase starts speaking.

Most of the movements categorized as 'moves' involve a turning away from looking at a person towards looking in front and downwards. Depending on the position of the other person this can be a movement sideways and down or simply a movement down. They can also involve a movement directing the head from one person to another. They can be faster or slower depending on the agitation of the person.

Deictic move. A special kind of move is where a speaker moves his head and gaze from one person (mostly) the addressee of the utterance to another person (or an object) and back typically when making reference to the person (or object) in the utterance. This can be accompanied by a pointing gesture with the hand.

Fig. 6. BLAAUW displaying a lateral move

Waggles. The waggle is similar to a shake. The main direction of movement is horizontal. However, the waggle has a more arch-like form. As with the shake the head moves from the start position to one side and then to the other side, but not just back to the starting position but beyond this. This is illustrated by the following screenshot. Here the speaker is *listing* all kinds of examples. Gradually, the waggle's amplitude grows.

Fig. 7. C displaying a waggle

The waggle is typically found when a speaker is making a list or when alternatives are being considered, which can occur without speech.

Head repositionings. When speakers return from a deictic move excursion to looking back at the addressee, we label this as a repositioning. The label is also applied at movements that position the head differently just prior to the onset of speech.

No motion. Episodes where a person in view does not show any movement occur mostly when they are not speaking. But there are occassions where speakers hold their head stationary as well for a shorter period of time.

These classes do not, of course, exhaust all possible movements, nor do they constitute the definitive classification of basic head movements. Some of the movements, or the way they are executed may be typical for the context and setting (for instance the table arrangement) We do believe though that similar constructs can be found in other contexts as well.

These basic movements combine in several ways, some of which have been mentioned before. There are various ways in which the basic movements combine to form more complex movements both by concatenation and by layering. In the case of nods, shakes, sweeps and waggles, the movement episode typically consist of repetitions of a two movements back and forth (left/right or up/down). As we mentioned above, in the case of sweeps and waggles, the start and end phase of each elementary movement may be adorned by either a stationary moment or some tiny nods or shakes. A typical case of layering occurs when somebody is speaking, producing tiny shakes along side and meanwhile turns his complete

head and gaze from one addressee to another. Again these cases do not exhaust all the noteworthy combinations.

6 Conclusion

The data described and analysed above is only one of the set of contexts that we are looking at. These preliminary investigations of which only a part is presented above, shows us in some detail how we can segment and label head movements of both speakers and listeners. The interpretation and the context shows a lot of overlap with previous findings in the literature with respect to the functions. One of the functions that has not been paid much attention to but which is a dimension that we are particularly investigating is what the head movements tell us about the mental state of the speaker or listener and the interpersonal attitudes. In particular, we find the level of excitement and the positive versus negative attitude (argumentative, defensive, etc.) towards the other important variables to work with in future work.

For the purpose of our investigations we use this data analysis in several ways. On the one hand we use the data to design and implement algorithms that can track, segment and label movements to be used as input for our sensitive artificial agent system. On the other hand we use this to establish a kind of lexicon of movements and contexts that we use in the generation of ECA behaviour because these simple movements can be highly expressive, as we hope to have shown.

Acknowledgements. Thanks are due to Justine Cassell for valuable comments and the Humaine Network of Excellence for partly supporting this work.

References

1. Heylen, D.: Head gestures, gaze and the principles of conversational structure. International Journal of Humanoid Robotics 3, 241–267 (2006)
2. Kendon, A.: Some uses of head shake. Gesture 2, 147–182 (2003)
3. Munhall, K., Jones, J.A., Callan, D.E., Kuratate, T., Vatikiotis-Bateson, E.: Visual prosody and speech intelligibility. Psychological Science 15, 133–137 (2004)
4. Hadar, U., Steiner, T., Rose, C.F.: Head movement during listening turns in conversation. Journal of Nonverbal Behavior 9, 214–228 (1985)
5. Heylen, D., Bevacqua, E., Tellier, M., Pelachaud, C.: Searching for prototypical facial feedback signals. In: Pelachaud, C., Martin, J.C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (eds.) Intelligent Virtual Agents, pp. 147–153. Springer, Berlin, Heidelberg (2007)
6. Gumpertz, J.: Discourse Strategies. Cambridge University Press, Cambridge (1982)
7. Heylen, D.: Multimodal backchannel generation for conversational agents. In: van der Sluis, I., Theune, M., Reiter, E., Krahmer, E. (eds.) Workshop on Multimodal Output Generation, Aberdeen, Scotland (2007)
8. Fukayama, A., Ohno, T., Mukawa, N., Sawaki, M., Hagita, N.: Messages embedded in gaze of interface agents - impression management with agent's gaze. In: Proceedings of CHI 2002, pp. 41–48. ACM, New York (2002)

9. Mignault, A., Chauduri, A.: The many faces of a neutral face: Heat tilt and perception of dominance and emotion. Journal of Nonverbal Behavior 27, 111–132 (2003)
10. Argyle, M., Cook, M.: Gaze and Mutual gaze. Cambridge University Press, Cambridge (1976)
11. Cassell, J., Thórisson, K.R.: The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. Applied Artificial Intelligence 13, 519–538 (1999)
12. Kendon, A.: Some functions of gaze direction in social interaction. Acta Psychologica 26, 22–63 (1967)
13. Heylen, D., van Es, I., van Dijk, B., Nijholt, A.: Experimenting with the gaze of a conversational agent. In: van Kuppevelt, J., Dybkjaer, L., Bernsen, N.O. (eds.) Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems, Kluwer Academic Publishers, Dordrecht (2005)
14. Reuderink, B.: The influence of gaze and head tilt on the impression of listening agents (2006)
15. Bevacqua, E., Heylen, D., Pelachaud, C., Tellier, M.: Facial feedback signals for ecas. In: Proceedings of AISB 2007: Artificial and Ambient Intelligence, Newcastle University, Newcastle upon Tyne, UK (2007)
16. Birdwhistell, R.L.: Kinesics and Context. Essays on Body Motion Communication. University of Pennsylvania Press, Philadelphia (1970)
17. McClave, E.Z.: Linguistic functions of head movements in the context of speech. Journal of Pragmatics 32, 855–878 (2000)
18. Chovil, N.: Social determinants of facial displays. Journal of Nonverbal Behavior 15, 141–154 (1991)
19. Hadar, U., Steiner, T., Grant, E., Rose, C.F.: Kinematics of head movements accompanying speech during conversation. Human Movement Science 2, 35–46 (1983)
20. Duncan, S.: Language, paralanguage, and body motion in the structure of conversations. In: McCormack, W., Wurm, S. (eds.) Language and Man. Anthropological Issues, pp. 239–268. Mouton, The Hague (1976)
21. Duncan, S.: On the structure of speaker-auditor interaction during speaking turns. Language in Society 2, 161–180 (1974)
22. Kendon, A.: Some functions of the face in the kissing round. In: Conducting Interaction. Editions de la Maison des Sciences de l'Homme, pp. 117–152. Cambridge University Press (1990)
23. Kendon, A.: Movement coordination in social interaction: some examples described. Acta Psychologica 32, 100–125 (1970)
24. Yngve, V.: On getting a word in edgewise. In: Papers from the sixth regional meeting of the Chicago Linguistic Society, Chicago: Chicago Linguistic Society, pp. 567–577 (1970)
25. Krumhuber, E., Manstead, A.S., Kappas, A.: Temporal aspects of facial displays in person and expression perception: the effects of smile dynamics, head-tilt, and gender. Journal of Nonverbal Behavior 31, 39–56 (2007)
26. Costa, M., Menzani, M., Bitti, P.E.R.: Head canting in paintings: An historical study. Journal of Nonverbal Behavior 25, 62–73 (2001)
27. Pendlton, K.L., Snyder, S.S.: Young children's perception of nonverbally expressed "preference": the effects of nonverbal cue, viewer age, and sex of actor. Journal of Nonverbal Behavior 6, 220–237 (1982)
28. O'Leary, M.J., Gallois, C.: The last ten turns: behavior and sequencing in friends' and stranger's conversational findings. Journal of nonverbabl behavior 9, 8–27 (1985)

29. Costa, M., Dinsback, W., Manstead, A.S., Bitti, P.E.R.: Social presence, embarrassment, and nonverbal behavior. Journal of nonverbal communication 25, 225–240 (2001)
30. Harrigan, J.A., Oxman, T.E., Rosenthal, R.: Rapport expressed through nonverbal behavior. Journal of Nonverbal Behavior 9, 95–110 (1985)
31. Bente, G., Feist, A., Elder, S.: Person perception effects of computer-simulated male and female head movement. Journal of Nonverbal Behavior 20, 213–228 (1996)
32. Cerrato, L., Skhiri, M.: Analysis and measurement of head movements signalling feedback in face-to-face human dialogues. In: Paggio, P., Jokinen, K., Jönsson, A. (eds.) First Nordic Symposium on Multimodal Communication, Copenhagen, pp. 43–52 (2003)