Project ref. no.	IST-1999-10647
Project title	ISLE Natural Interactivity and Multimodality Working Group

Deliverable status	Public
Contractual date of delivery	30 September 2001
Actual date of delivery	February 2002
Deliverable number	D9.1
Deliverable title	Survey of Multimodal Annotation Schemes and Best Practice
Type	Report
Status & version	Final
Number of pages	122
WP contributing to the deliverable	WP9
WP / Task responsible	Laila Dybkjær, NISLab
Editors	Malene Wegener Knudsen, Jean-Claude Martin and Laila Dybkjær
Authors	Malene Wegener Knudsen, Jean-Claude Martin, Laila Dybkjær, María Jesús Machuca Ayuso, Niels Ole Bernsen, Jean Carletta, Ulrich Heid, Sotaro Kita, Joaquim Llisterri, Catherine Pelachaud, Isabella Poggi, Norbert Reithinger, Gijs van Elswijk, Peter Wittenburg
EC Project Officer	Brian Macklin
Keywords	Natural Interactivity, Multimodality, Annotation Schemes, Gesture, Facial Expression
Abstract (for dissemination)	This ISLE Deliverable 9.1 from the ISLE Natural Interactivity and Multimodality (NIMM) Working Group presents a survey of NIMM annotation schemes and best practice in the field. The report reviews 7 facial annotation schemes and 14 gesture annotation schemes, some of them including speech. In addition, a chapter on best practice in the field is included. The descriptions of the 7 facial and 14 gesture annotation schemes have a common structure. The purpose of this structure is to facilitate comparison across annotation schemes by providing similar information about each scheme to the extent possible.

ISLE Natural Interactivity and Multimodality Working Group Deliverable D9.1

Survey of Multimodal Coding Schemes and Best Practice

February 2002

Authors

Malene Wegener Knudsen[1], Jean-Claude Martin[7], Laila Dybkjær[1], María Jesús Machuca Ayuso[2], Niels Ole Bernsen[1], Jean Carletta[5], Ulrich Heid[6], Sotaro Kita[8], Joaquim Llisterri[2], Catherine Pelachaud[4], Isabella Poggi[4], Norbert Reithinger[3], Gijs van Elswijk[8], Peter Wittenburg[8]

1: NISLab, University of Southern Denmark. 2: DFE, Barcelona, Spain. 3: DFKI, Saarbrücken, Germany. 4: DIS, University of Rome, Italy. 5: HCRC, Edinburgh, UK. 6: IMS, Stuttgart University, Germany. 7: LIMSI-CNRS, Orsay, France. 8: MPI, Nijmegen, The Netherlands

Contents

1 Introduction

This ISLE (International Standards for Language Engineering) Natural Interactivity and Multimodality (NIMM) Working Group report D9.1 provides a survey of NIMM coding schemes which include facial expression and/or gesture, possibly combined with speech. The report forms part of a series of European ISLE NIMM WG reports on data resources, coding schemes and coding tools for natural interactivity and multimodality. Report D11.1 on NIMM coding tools and report D8.1 on NIMM data resources are available at http://isle.nis.sdu.dk. This series of ISLE reports continues the work on coding schemes (deliverable D1.1) and coding tools (deliverable D3.1) surveys done for spoken dialogue in the MATE project (Multilevel Annotation Tools Engineering, 1998-2000). MATE reports are available at http://mate.nis.sdu.dk. The present survey comprises 21 different coding schemes of which 7 can be used to code facial expression possibly combined with speech, and 14 primarily can be used to code gesture possibly accompanied by speech. The report also includes a chapter on best practice and practices in coding schemes.

The present report is the result of work in two sub-groups, one looking at coding schemes for facial expression with or without accompanying speech, and one looking at coding schemes for gesture with or without accompanying speech. The work on facial coding schemes was led by NISLab in close collaboration with U-ROME and with contributions from HCRC, while the work on gesture resources was led by LIMSI-CNRS with contributions from IMS, MPI, DFKI and DFE. Not only did the distribution of expertise in the ISLE NIMM WG make it natural to proceed in this way but we also found that, at the present time, rather few coding schemes focus on both facial expression and gesture. Since it can be expected that an increasing number of data resources which focus more or less equally on speech, facial expression, and gesture will become available in the future, it is also expected that coding schemes that can be used to code all these modalities will become available in the future. When we refer to coding schemes that can be used to code facial expression data, this includes coding schemes which focus on specific facial parts such as the lips or the eyes. Lip movement and eye/gaze behaviour coding is thus included among the facial coding schemes. Likewise, it should be noted that the coding schemes for gesture presented include gesture made in the setting of human-system communication as well as gesture made in human-human communication.

To our knowledge, the present report is the most comprehensive survey to date on facial and gesture coding schemes. We hope that the report will be of interest to colleagues from academia and industry who have a need for, or take an interest in, working with NIMM coding schemes.

In the following, we first briefly define some central concepts (Section 1.1). We then describe our approach in terms of how the surveyed coding schemes were selected and described (Section 1.2). Section 1.3 provides an overview of the reviewed coding schemes and whether contact was established to the creator(s) of each coding scheme. Finally, Section 1.4 draws some general conclusions from this report as regards practices and best practice.

The following chapters describe (i) the facial coding schemes as divided into facial coding schemes (Chapter 2) and lesser known/used facial coding schemes for which we have found little information (Chapter 3), and (ii) the gesture coding schemes as divided into gesture coding schemes (Chapter 4) and lesser known gesture coding schemes for which we found little information (Chapter 5). Finally, Chapter 6 provides conclusions on practices and best practices in the field.

1.1 Definitions

Despite the fact that this report is termed a survey of NIMM (Natural Interactivity and Multimodality) data resources, it may be useful to point out that its focus is on natural interactivity rather than on multimodality. The following definitions are necessary for understanding this claim.

A *modality* is a particular way of presenting information in some medium. A *medium* is a physical substrate or vehicle for information presentation, such as light/graphics which is being perceived visually, or sound/acoustics which is being perceived auditively. Obviously, there are many different ways of representing information in a particular medium. This is why we need to distinguish between

different modalities presented in the same medium. For instance, spoken language is a modality expressed in the acoustic medium, whereas written language and facial expression are modalities expressed in the medium of light/graphics. In the form of lip movements or textual transcription, spoken language may also be expressed in the medium of light/graphics. Thus, the same modality can be represented (more or less adequately) in different media. During human-human-system interaction (see below), a modality may be used as an *input* modality (from a human to the system or to other humans) or as an *output* modality (from the system to humans or, rarely today, to another system), or both. *Multimodal* representations are representations which can be decomposed into two or more *unimodal* modalities. For more details on these basic concepts in Modality Theory, see Bernsen, N. O.: Multimodality in language and speech systems - from theory to design support tool. To appear in Granström, B. (Ed.): *Multimodality in Language and Speech Systems*. Dordrecht: Kluwer Academic Publishers 2002. *Interaction* refers to communication or information exchange between humans, possibly mediated by a computer system, or between humans and computer systems. The term *natural* qualifies the interaction and refers to the ways in which humans normally exchange information with one another.

Given those definitions, it is clear that natural interactive communication is multimodal, using several media and a large number of different modalities of information representation and exchange. However, multimodal exchange of information is not necessarily a case of natural interaction. Multimodal exchange of information might, for instance, and in fact often does, include media which are not perceivable by humans, such as magnetic fields, radar, or ultrasound. It follows that the modalities used in those media are not ones used by humans in their natural interactive communication with each other. In other words, the term "multimodal interaction" is a generic term which subsumes natural interaction as well as information representation and exchange which cannot be described as natural in this sense. It is clear from the surveyed data resources below that the vast majority of resources directly address, or are at least relevant to, natural interaction and its understanding for scientific or application-oriented purposes.

1.2 Approach

The approach adopted for producing the present report was to (i) first identify a common set of criteria for selecting the coding schemes to be described and decide upon issues concerning quality of content as well as of presentation; then (ii) establish a common template for describing each coding scheme; (iii) identify relevant coding schemes world-wide based on the web, literature, networking contacts among researchers in the field, etc., and, finally (iv), interact with the coding scheme creators to the extent possible in order to gather information on their resources and ask them to verify the data resource descriptions produced.

1.2.1 Selection criteria

To keep the survey focused, the following criteria were adopted for selecting the coding schemes to be included below:

Documentation: It is interesting if the coding scheme is well documented in the sense that it has a coding book which describes the purpose and the domain for which the scheme has been developed. Exception to this point is made if the coding scheme is rare in its domain or still under development. Moreover, the coding schemes should be substantiated by examples for better understanding and come with a contact address where one can gain further information.

Usability: The coding scheme should have been used by a decent number of researchers, due to the fact that coding schemes that have only been used by their developers tend to be too subjective and difficult to use. However, if a coding scheme has been used only by its developers but for a large data resource which has been included in ISLE Deliverable D8.1, it is still included in this survey. Moreover, to demonstrate its usability, the coding scheme must have been used to annotate a certain number of dialogues/interactions, and it must be in recent use or have potential for future use.

Mark-up language: The coding scheme must come with a list of phenomena which have been annotated by using the coding scheme (tag set) in order to make possible comparison among the different coding schemes. Moreover, the markup language should be described.

Evaluation desirability: It is interesting if users outside the group of developers have evaluated the coding scheme on, e.g., matters of intercoder agreement. It is interesting if the coding scheme has been used to code a certain number of the data resources included in the ISLE Deliverable 8.1 survey of NIMM data resources. Moreover, it is interesting if the coding scheme has tool support, and if the tool is possibly included in the ISLE Deliverable 11.1 survey on NIMM coding tools.

Exceptions: Exception may be made to the above if a coding scheme is so rare or innovative for its domain that its very existence might be of interest to researchers in the field. However, it is still desirable that the coding scheme is generalisable to a certain degree, at least, and not only suitable for one very particular and limited task.

1.2.2 Quality of content and presentation

NIMM coding schemes tend to be voluminous, and they are sometimes carefully protected against intruders in the sense that it costs time and money to become an approved-by-the-developers user of them. To realistically compromise among the above selection criteria, we have adopted the following guidelines for contents inspection, validation and presentation:

Hands-on: It is highly desirable that the describer of a certain coding scheme has actually tried to use the coding scheme.

Validation: All descriptions should be validated by someone other than the describer, if at all possible with the coding scheme creator in the loop, either as describer or as validator.

Examples: Whenever permissible, a short example of coding of a resource should be presented in the report. If, for whatever reason, it has not been possible to access and inspect a coding example first-hand, this should be stated clearly in the description.

1.2.3 Common description template

In order to help providers of coding scheme information to document their coding schemes, facilitate the reading of this report, and allow some measure of easy comparison among the coding schemes presented, coding scheme descriptions have a common structure which, to the extent possible, provides the same information about all coding schemes. The common structure includes 8 main entries as shown in Figure 1.2.1. Each main entry subsumes a number of more specific information items.

The common description template went through several revisions as work on the survey proceeded, for instance in order to take into account types of coding schemes which had not been anticipated from the outset. Other adjustments became necessary during the validation process where the close contact to the coding scheme creators often demonstrated that the creators took a critical approach to their own coding schemes, providing valuable information on what they would do different were they to create their coding schemes once more.

Reference (specify coding scheme by project name, main authors or laboratory)

Description header

Main actor (name and email)

Verifying actor (name and email)

Date of last modification of the description

References

Web site(s)

Short description

Illustrative example of coding

References to additional information on the coding scheme (journal or conference paper, whitepaper…)

Coverage

Which types of raw data are referenced?

Which modalities is the coding scheme meant to code?

Which annotation level(s) does the coding scheme cover, e.g. facial expression and prosody?

Which coding tasks has the coding scheme been used for?

Detailed description of coding scheme

Which header file information is included (meta-data)?

Coding purpose of the coding scheme?

List and description of phenomena which can be annotated by the scheme

Description of markup language/markup declaration

Examples

Description of coding procedure, if any

Creation notes, i.e. who wrote the coding scheme, when, and in which context?

Usage

Origin of the coding scheme and reasons for creating it

How many people have used the coding scheme and for what purposes?

How many dialogues/interactions have been annotated using the coding scheme?

Has the coding scheme been evaluated?

Is the coding scheme language dependent or language independent?

Is there tools support for using the coding scheme or API for editing/parsing coded descriptions? In which language?

7. Accessibility

How does one get access to the coding scheme?

Is the coding scheme available for free or how much does it cost?

8. Conclusion

How well described is the coding scheme?

How general and useful is the coding scheme?

Figure 1.2.1. Common coding scheme description template.

1.2.4 Interaction with coding schemes creators

Close interaction with the creators of the surveyed coding schemes has been sought throughout the writing of this report, first, of course, to seek their permission to publicly describe their coding scheme, and secondly to invite their collaboration in producing as useful and accurate information about the coding scheme as possible. Creators of the coding schemes reviewed in this report were invited to comment on the description of their coding scheme and to validate the final description of their coding scheme, resulting in feedback on, and validations of, most descriptions, cf. Figure 1.3.1. Many coding schemes creators pointed out the potential value of the present survey, arguing that had the survey been available when they needed coding schemes for their own research, this might have made their work easier because they might have been able to use an already available coding scheme instead of going through the laborious process of creating their own, or they might have been in a better position to learn from other researchers' experience with coding scheme creation.

1.3 Surveyed coding schemes

For each coding scheme described in the following chapters, an indication is included of which ISLE partner was the main actor in making the description, i.e. the partner that had the main responsibility for describing that particular coding scheme. In most cases, each coding scheme was verified by another ISLE partner or – in a few cases – by the coding scheme creators which is then also indicated. Only for a few coding schemes on which little information was available, no verifying actor was involved. For each described coding scheme we tried to establish contact to the creator(s) to also make them verify our descriptions and possibly provide additional information. In a number of cases we

received valuable feedback while in other cases we never succeeded in getting a response. Figure 1.3.1 lists the surveyed coding schemes in the order in which they are described in this report and indicates for each scheme whether we received feedback from the creator(s).

* after a coding scheme name indicates that the creator(s) of the coding scheme provided feedback on our description.

+ means that the coding scheme was created at the main actor's site and that feedback on our description thus was provided by a person located at the main actor's site.

- means that we did not succeed in establishing contact to the creator(s) of the coding scheme.

Sections 3.2, 4.11, 5.2 and 5.3 are not marked in Figure 1.3.1 since these provide more general descriptions and do not concern any particular coding scheme.

2	**Facial Coding Schemes**
2.1	The Alphabet of eyes: formational parameters of gaze+
2.2	Facial Action Coding System – FACS-
2.3	The Maximally Discriminative Facial Movement Coding System (MAX)*
2.4	MPEG-4 SNHC (Moving Pictures Expert Group, Synthetic/Natural Hybrid Coding)*
2.5	ToonFace*
3	**Lesser Known/Used Facial Coding Schemes**
3.1	BABYFACS – Facial Action Coding System for Baby Faces-
3.2	General description of coding schemes for hand annotation of mouth and lip movements and speech
4	**Gesture Coding Schemes**
4.1	DIME: National Autonomous Univ. of Mexico, Multimodal extension of DAMSL-
4.2	HamNoSys - Hamburg Notation System for Sign Languages*
4.3	HIAT -- Halbinterpretative Arbeitstranskriptionen*
4.4	LIMSI Coding Scheme for Multimodal Dialogues between Car Driver and Copilot+
4.5	MPI GesturePhone+
4.6	MPI Movement Phase Coding Scheme+
4.7	MPML - A Multimodal Presentation Markup Language with Character Agent Control Functions-
4.8	SmartKom Coding scheme+
4.9	SWML (SignWriting Markup Language)*
4.10	TUSNELDA Corpus Annotation standard*
4.11	General description of coding schemes for prosody, gestures and speech
5	**Lesser Known/Used Gesture Coding Schemes**
5.1	LIMSI TYCOON scheme for analysing cooperation between modality+
5.2	W3C Working Draft on Multimodal Requirements for Voice Markup Languages
5.3	The New England Regional Leadership Non-Verbal Coding scheme

Figure 1.3.1. Verification of coding scheme by its creator(s).

1.4 The need for NIMM coding schemes

Based on the survey work and our close interaction with coding scheme creators, we have observed a felt and growing user need for surveys which present coding scheme descriptions that are easily comparable when considering the creation process, accessibility issues and usage. Furthermore, a picture of best practice for NIMM coding schemes and surveys of these has emerged through the making of this survey and is described in Chapter 6 below.

Despite a positive attitude towards re-use of coding schemes, there is still a long way to go before this will be the rule rather than the exception as it seems to be today – not least as regards gesture markup. In the field of facial expression markup, steps towards standardisation seem far more advanced, so that several coding schemes are actually being used by many colleagues in the community.

2 Facial Coding Schemes

This introduction covers chapters 2 and 3, both of which have a primary focus on facial coding schemes. Chapter 2 presents facial coding schemes while Chapter 3 describes lesser know facial coding schemes for which we have found little information.

The descriptions cover existing coding schemes which have been applied to the coding of facial expression, including e.g. eyes, eye brows and lips. The schemes cover faces of adults as well as faces of babies and children, and they cover human faces as well as cartoon faces.

2.1 The Alphabet of eyes: formational parameters of gaze

2.1.1 Description header

Main actor

UROME: Catherine Pelachaud (cath@dis.uniroma1.it) and Isabella Poggi (poggi@uniroma3.it)

Verifying actor

NISLab: Malene Wegener Knudsen (mwk@nis.sdu.dk), Laila Dybkjær (laila@nis.sdu.dk) and Niels Ole Bernsen (nob@nis.sdu.dk)

Date of last modification of the description

June 18th, 2001.

2.1.2 Reference

Web site

The coding scheme has no official web site yet.

Short description

This coding scheme is concerned with gaze behaviour. The formational parameters that may be useful to take into account to analyse gaze behaviour are the parts of the eye region (eyebrows, wrinkles, eyelids, eyes) and some movements or features of them (humidity, reddening, pupil dilation, eye direction, eye movements).

An illustrative example of coding

A coding of Adriana Faranda, “Women at the cross-road”. “Women at the cross-road” was a talk show in Italian TV (one of the programs of Mediaset, Berlusconi’s net), conducted by Enza Sampò. She interviewed women who had had a dramatic experience. Adriana Faranda was a terrorist during the hard times of the seventies in Italy. She was in jail for some 15 years, and in this talk show she talked about her experiences not from a political but from a human point of view.

In Italian Adriana Faranda is saying: “della mia vita particolar-mente significa-tiva”. A word-to-word translation into English: “of my life particularly relevant”.

Below the speech and the eye movements are described and finally the meaning of the eye movements are deducted.

	della mia vita of my life	*particolar-mente significa-tiva* particularly relevant	
iris position	left	centre	centre
iris direction	left down	Interlocutor	Interlocutor
head direction	Interlocutor	Interlocutor	Interlocutor
eyebrow right	normal	internal up, external up	normal
left	normal	normal	normal
eyelids upper	slightly lowered	raised	normal
lower		lowered	corrugated
external corner right	normal	(tense)	
left	normal	tense	upward (laugh)
humidity	no	yes	yes
MEANING	**I am remembering**	**Important** **+** **terror**	**I ask you to pay attention**

Figure 2.1.1. A coding of Adriana Faranda, "Women at the cross-road".

References to additional information on the coding scheme

De Carolis, B., Pelachaud, C. and Poggi, I.: Verbal and nonverbal discourse planning. In Proceedings of the Workshop: Achieving Human-Like Behavior in Interactive Animated Agents, in conjunction with The Fourth International Conference on Autonomous Agents. Barcelona (Spain), June 3, 2000.

Pelachaud, C. and Poggi, I.: Talking faces that communicate by eyes. In S. Santi, B. Guaitella, C. Cavé and G.Konopczynski (Eds.): Oralite et gestualite, communication multimodale, interaction. Paris: L'Harmattan, , pp. 211-216, 1998.

Pelachaud, C., Poggi, I and De Rosis, F.: Study and Generation of Coordinated Linguistic and Gaze Communicative Acts. In E. Lamma and P. Mello (cur.): AI*IA '99. Atti del VI Congresso dell'Associazione Italiana per l'Intelligenza Artificiale. Bologna: Pitagora Ed., pp. 248-257, 1999.

Pezzato, N. and Poggi, I.: Le funzioni comunicative delo sguardo. In A. Tronconi (cur.): Atti del 6° Convegno Nazionale Informatica, Didattica e Disabilità. Andria (Bari), 46 November, pp. 27-31, 1999.

Poggi, I. and Pelachaud, C.: Emotional meaning and expression in performative faces. In A. Paiva (Ed.): Affective Interactions. Towards a new generation of computer interfaces. Berlin: Springer, pp. 182-195, 2000.

Poggi, I. and Pelachaud, C.: Emotional meaning and expression in performative faces. In A. Paiva and C. Martinho (Eds.): Affect in Interactions. In Proceedings of the Workshop of the 3rd i3 annual Conference, Siena, October 21-22, pp. 122-126, 1999.

Poggi, I., Pelachaud, C., and De Rosis, F.: Eye communication in a conversational 3D synthetic agent. In AI Communications 13, pp. 169-181, 2000.
Poggi, I., Pezzato, N. and Pelachaud, C.: Gaze and its meaning in animated faces. In P. McKevitt (Ed.): CSNLP-8 (Cognitive Science and Natural Language Processing). Proceedings of the Workshop on Language, vision and music. Galway (Ireland), 9-11 August, 1999.

2.1.3 Coverage

Which types of raw data are referenced?

The data referenced is video files.

Which modalities is the coding scheme meant to code?

The coding scheme is meant to code facial expressions and gaze behaviour.

Which annotation level(s) does the coding scheme cover?

The coding scheme covers gaze, namely: eyebrow movements, eyelid openness, wrinkles, eye direction, eye reddening and humidity, and pupil dilation.
It covers gaze and the relationship between gaze communication and speech.

Which coding tasks has the coding scheme been used for?

It has been used for analysing gaze in different types of videotaped data: talk shows, teacher's facial behaviour in class and other.

2.1.4 Detailed description of coding scheme

Which header file information is included (meta-data)?

None

Coding purpose of the coding scheme?

The coding purpose of the coding scheme is to describe gazes and clustering different tokens of the same gaze type.

List and description of phenomena, which can be annotated by the scheme

Any kind of gaze (glance, stare, blink, loving eyes, defying gaze....).
In a natural language we have a folk taxonomy of types of gaze. We do know what is a *blink*, what is a *glance*, what is a *defying gaze*; and we know that looking at someone with a glance is different from staring at him. In other words, by these names of gaze we have information on both how a gaze is performed (the signal) and what it means (the meaning). Our coding scheme allows one to state precisely what all the features are that characterize each different type of gaze and, possibly, its specific meaning. A blink for example implies a speedy closing of the eyelids; in a defying gaze usually both the eyes and the head are directed towards the interlocutor, while in a glance eye direction is oblique with respect to head direction. Moreover, a defying gaze means we want to defy someone, while a glace means we want to communicate to him in a furtive way. All of this information is provided by our scheme.

Description of markup language/markup declaration

No specific markup language has been developed for this coding scheme.

Examples

A case of emotional gaze (fear). The corresponding sample is recorded at the time only on VHS videotape and is therefore not online available.

Eyebrows	Right	Internal	Up
		Central	Up
		External	Up
	Left	Int.	Up
		Cent.	Up
		Ext.	Up
Eyelids	Right	Upper	Up/Tense
		Lower	Down / Tense
	Left	Upper	Up/Tense
		Lower	Down/Tense
Humidity			Default
Reddening			Default
Pupil dilation			No
Focusing			Yes
Iris Position			Central
Iris Direction			Right
Face Direction			Right
Head Inclination			Default
Trunk Direction			Forward
Interlocutor Direction			Right
Duration			Short

Figure 2.1.2. Table showing the different possible movements of the face, which the coding scheme takes into account.

Description of coding procedure, if any

Firstly the verbal signal concomitant to the gaze under analysis is written down, in order to have a reference point. Then the gaze is looked at (usually many times) in normal speed and slow motion, trying to catch the particular feature to be analysed: eye direction, pupil dilation, eyebrow movements, and so on. Finally the right values are written into the coding scheme.

Creation notes, i.e. who wrote the coding scheme, when,and in which context?

Created by I. Poggi (poggi@uniroma3.it), C. Pelachaud (cath@dis.uniroma1.it) and N. Pezzato. Contact person:
Isabella Poggi
Dipartimento di Scienze dell'Educazione
Università Roma Tre

Via del Castro Pretorio 20
00185 Roma - Italy
Tel. 0039-06-49229314
The coding scheme was created during a seminar on “Theory of Communication” and was the part of the work for an unpublished undergraduate thesis by Nicoletta Pezzato at Università Roma Tre, Faculty of Education, 1998/99. Nicoletta Pezzato graduated in 1999 with a thesis in “Theory of Communication”, with Dr. Isabella Poggi as her advisor.

2.1.5 Usage

Origin of the coding scheme and reasons for creating it

The coding scheme is the result of a research on gaze. It was created to analyse any single item of gaze in videotaped data.

How many people have used the coding scheme and for what purposes?

So far 4 people have used the coding scheme for research purposes.

How many dialogues/interactions have been annotated using the coding scheme?

Ten interactions have been annotated using the coding scheme.

Has the coding scheme been evaluated?

The coding scheme has been evaluated by the researchers’ own students by applying written and verbal instructions and using a VHS projector to show the interactions. The result of the evaluation and intercoder agreement was generally good.

Is the coding scheme language dependent or language independent?

The coding scheme is language independent.

Is there tools support for using the coding scheme or API for editing/parsing coded descriptions? In which language ?

At the moment there is no tool support for using the coding scheme but one is being prepared by I. Poggi and E. Magno Caldognetto. This tool will allow the annotation of multimodal conversation on sides of both, the signals (description and classification of gaze parameters, facial expressions and gestures) and of their literal and indirect meaning.

2.1.6 Accessibility

How does one get access to the coding scheme?

By contacting Isabella Poggi using the above address.

Is the coding scheme available for free or how much does it cost?

The coding scheme is available for free.

2.1.7 Conclusion

How well described is the coding scheme?

Quite generally described in published papers, more specifically described in three unpublished theses (written in Italian).

How general and useful is the coding scheme?

It is useful to analyse any possible gaze, to classify it as one of the possible gaze types, and then find the meaning attached to that specific gaze type. In this respect the coding scheme is quite general, since one may apply it to whatever data to the extent to which one needs to know what particular type of gaze a subject has performed and, in some cases, what the meaning might be, intentionally or not, conveyed by that gaze. One can apply it, say, in trials, to assess dismay or terror in the accused; in classroom interaction, to see how interested or bored pupils are; in oral examination at the university, to tell if a student has too many inferences to draw before catching the right answer and so on.

2.2 Facial Action Coding System - FACS

2.2.1 Description header

Main actor

UROME: Catherine Pelachaud (cath@dis.uniroma1.it) and Isabella Poggi (poggi@uniroma3.it)

Verifying actor

NISLab: Malene Wegener Knudsen (mwk@nis.sdu.dk), Laila Dybkjær (laila@nis.sdu.dk) and Niels Ole Bernsen (nob@nis.sdu.dk)

Date of last modification of the description

June 18th, 2001.

2.2.2 Reference

Web site

FACS has no website of its own.

Short description

In order to encode facial expression FACS was developed by Paul Ekman and Wallace Friesen. Paul Ekman, Wallace Friesen, and S.S. Tomkins had already developed another system called Facial Affect Scoring Technique (FAST) that is not used anymore. FACS is a further development of FAST.

An illustrative example of coding

See figure 1.2.2. and 1.2.3.
FACS involves four operations:

- determining which AUs are responsible for the observed movements,
- scoring the intensity of the actions on a three-point scale: low (X), medium (Y), and high (Z),
- deciding whether an action is asymmetrical or unilateral, and
- determining the position of the head and the position of the eyes during a facial movement.

Most of the AUs combine additively. In other cases, rules of dominance, substitution or alternation take over. The dominance rule dictates when an AU disappears for the benefit of another AU. The substitution rule allows for the elimination of certain AUs when others produce the same effects. Finally, the alternation rule takes over when AUs cannot combine. A facial expression is the results of different AUs. Describing a facial expression consists in recognizing which AU is responsible for which facial action.
For more examples see:
From P. Ekman, W. Friesen, "Unmasking the face: A guide to recognizing emotions from facial clues", Prentice-Hall, INC. Englewood Cliffs, New-Jersey, 1975:

- surprise eyebrow: AU 1 + AU2. "The brows are raised, so that they are curved and high.", p. 45
- fear eyebrow: AU1 + AU2 + AU4. "The brows are raised and drawn together", p. 63.
- sadness eyebrow: AU1 + AU4. "The inner corners of the eyebrows are drawn up", p. 126.

- anger eyebrow: AU4. “The brows are lowered and drawn together”, p. 95

References to additional information on the coding scheme

P. Ekman and W. Friesen: Facial Action Coding System. Consulting Psychologists Press, Inc. 1978.

P. Ekman and W. Friesen: Unmasking the Face: A guide to recognize emotions from facial clues. Prentice-Hall, Inc. 1975.

Ekman, P. and Rosenberg, E. (Editors): What the face reveals: Basic and Applied Studies of Spontaneous Expression using the Facial Action Coding System (FACS). Oxford University Press, Oxford. 1997.

Paul Ekman, Thomas S. Huang, Terrence J. Sejnowski, Joseph C. Hager: Report To NSF of the Planning Workshop on Facial Expression Understanding. July 30 to August 1, 1992. Can be found at: http://mambo.ucsc.edu/psl/nsf.txt

2.2.3 Coverage

Which types of raw data are referenced?

Any visual data, image or video.

Which modalities is the coding scheme meant to code?

The coding scheme is meant to code facial expressions.

Which annotation level(s) does the coding scheme cover?

The coding scheme covers the annotation level of facial expression.

Which coding tasks has the coding scheme been used for?

FACS is widely used to encode facial expression of emotion. It is also used as a parameterised scheme of 3D facial models. Recognition algorithms have been developed based on this notation system.

2.2.4 Detailed description of coding scheme

Which header file information is included (meta-data)?

No information available.

Coding purpose of the coding scheme?

The FACS system allows expert coders to manually measure facial expressions by breaking them down into component movements of individual facial muscles (Action Unit).

P. Ekman explains FACS function as follows: “It is by emphasizing patterns of movement, the changing nature of facial appearance, that distinctive actions are described – the movements of the skin, the temporary changes in shape and location of the features, and the gathering, pouching, bulging, and wrinkling of the skin”.

List and description of phenomena, which can be annotated by the scheme

Any visible change on the face (muscle deformation, apparition of wrinkles / bulges, folds), and secondary movements can be annotated by the scheme.

AUs	Name	AU	Name
AU1	Inner Brow Raiser	AU31	Jaw Clencher
AU2	Outer Brow Raiser	AU32	Lip Bite
AU4	Brow Lowerer	AU33	Cheek Blow
AU5	Upper Lid Raiser	AU34	Cheek Puff
AU6	Cheek Raiser, Lid Compressor	AU35	Cheek Suck
AU7	Lid Tightener	AU36	Tongue Bulge
AU8	Lips Toward Each Other	AU37	Lip Wipe
AU9	Nose Wrinkler	AU38	Nostril Dilator
AU10	Upper Lip Raiser	AU39	Nostril Compressor
AU11	Nasolabial Furrow Deepener	AU41	Lip Droop
AU12	Lip Corner Puller	AU42	Slit
AU13	Sharp Lip Puller	AU43	Eyes Closed
AU14	Dimpler	AU44	Squint
AU15	Lip Corner Depressor	AU45	Blink
AU16	Lower Lip Depressor	AU46	Wink
AU17	Chin Raiser	AU51	Head Turn Left
AU18	Lip Pucker	AU52	Head Turn Right
AU19	Tongue Show	AU53	Head Up
AU20	Lip Stretcher	AU54	Head Down
AU21	Neck Tightener	AU55	Head Tilt Left
AU22	Lip Funneler	AU56	Head Tilt Right
AU23	Lip Tightener	AU57	Head Forward
AU24	Lip Presser	AU58	Head Back
AU25	Lips Part	AU61	Eyes Turn Left
AU26	Jaw Drop	AU62	Eyes Turn Right
AU27	Mouth Stretch	AU63	Eyes Up
AU28	Lip Suck	AU64	Eyes Down
AU29	Jaw Thrust	AU65	Walleye
AU30	Jaw Sideways	AU66	Cross-eye

Figure 2.2.1. List of action units (AUs).

Description of markup language/markup declaration

An AU is a minimal visible action. It corresponds to the action of a muscle or a group of related muscles. Each AU describes the direct effect of muscle contraction as well as secondary effects due to movement propagation and the presence of wrinkles and bulges.

Examples

The following is not a sample from a resource, but a drawn illustration of examples of AUs.

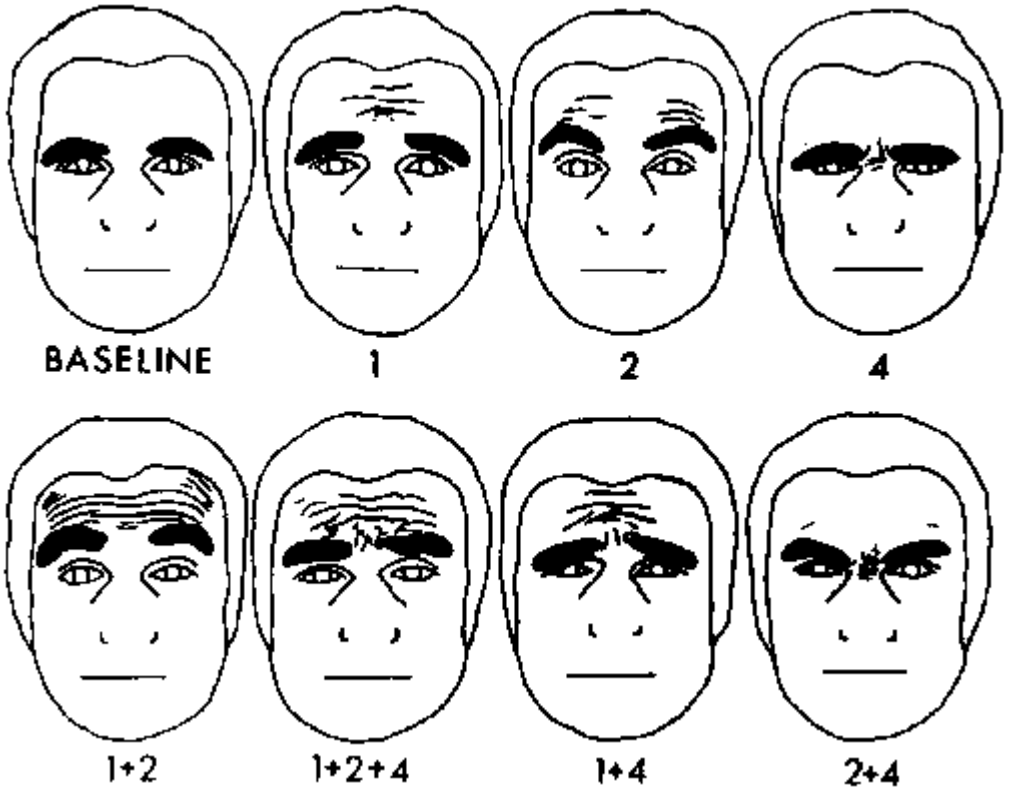


Figure 2.2.2. AU1 corresponds to the raise of the inner brow while AU2 to the raise of the outer brow. AU4 corresponds to a loweting of the brow. Their combination is the action of both AUs simultaneously: see example AU1+2 raises the inner and the outer part of the brow.

To perform AU1+2, one should simply raise one's complete eyebrows. It is an easy action to perform which is not the case of every AU. Indeed some AUs can be difficult to execute separately and/or consciously; but once they appear in combination or as part of a conversational signal, emotion or any type of spontaneous facial expression, they are easily produced.

Doing the expression of AU1+2 one can notice the following facial changes:

- apparition of wrinkles on the forehead. For some people no wrinkle appears but the skin of the forehead bulges nevertheless. The wrinkles are much more apparent than in neutral position (that is without expression). They are deeper.
- the entire eyebrow is raised.
- the eye cover (it is the part between the eye and the brow) becomes more apparent.
- in case people have heavy eye cover, this action raises it making apparent their eyelid that usually disappear under their eye cover.

In the description of each AU, minimum requirements for scoring the AU are defined. For AU1+2, the minimum requirements are not simply the sum of both minimum requirements for AU1 and AU2. They correspond to the facial changes explained above but in low intensity. That is that the entire eyebrow is raised slightly with slight apparition of wrinkles on the forehead and/or slight apparition of the eye cover.

Figure 2.2.3. AUs 10+15. AU-10 raises the upper lip and curves the nasolabial fold (action of levator labii superioris). AU-15 pulls the corners of the lips downwards obliquely (action of triangularis).

Description of coding procedure, if any

FACS involves four operations:

- determining which AUs are responsible for the observed movements.
- scoring the intensity of the actions on a three-point scale: low (X), medium (Y), and high (Z).
- deciding whether an action is asymmetrical or unilateral.
- determining the position of the head and the position of the eyes during a facial movement.

Paul Ekman precises: “A FACS coder "dissects" an observed expression, decomposing it into the specific AUs, which produced the movement. The coder repeatedly views records of behaviour in slowed and stopped motion to determine which AU or combination of AUs best account for the observed changes. The scores for a facial expression consist of the list of AUs, which produced it. The precise duration of each action is also determined, and the intensity of each muscular action and any bilateral asymmetry is rated. In the most elaborate use of FACS, the coder determines the onset (first evidence) of each AU, when the action reaches an apex (asymptote), the end of the apex period when it begins to decline, and when it disappears from the face completely (offset). These time measurements are usually much more costly to obtain than the decision about which AU(s) produced the movement, and in most research only onset and offset have been measured.” Quotation is taken from:

Paul Ekman, Thomas S. Huang, Terrence J. Sejnowski, Joseph C. Hager: Report To NSF of the Planning Workshop on Facial Expression Understanding. July 30 to August 1, 1992. Can be found at: http://mambo.ucsc.edu/psl/nsf.txt

In total 46 AUs are defined.

Creation notes, i.e. who wrote the coding scheme, when,and in which context?

FACS was developed by Paul Ekman and Wallace Friesen at the Langley Porter Neuropsychiatric Institute in San Francisco in 1975.

Paul Ekman, Ph.D.
Professor of Psychology
Department of Psychiatry
University of California
San Francisco
email: ekmansf@itsa.ucsf.edu

2.2.5 Usage

Origin of the coding scheme and reasons for creating it

FACS was developed by Paul Ekman and Wallace Friesen in 1975 in order to encode facial expression. Paul Ekman, Wallace Friesen, and S.S. Tomkins had already developed another system called Facial Affect Scoring Technique (FAST) that is not used any more. FACS is a further development of that technique.

How many people have used the coding scheme and for what purposes?

Many – more than 300 researchers are expert FACS coders (in 1992). Coders spend approximately 100 hours learning FACS. Self instructional materials teach the anatomy of facial activity, i.e., how muscles singly and in combination change the appearance of the face. Workshops are organised on the topic of FACS: http://emotions.psychologie.uni-sb.de/facs/program.htm

How many dialogues/interactions have been annotated using the coding scheme?

Many. The exact number is difficult to quantify as a large numbers of FACS coders are using FACS for their research.

Has the coding scheme been evaluated?

Prior to using FACS, all learners are required to score a videotaped test (provided by P. Ekman), to ensure that they are measuring facial behaviour in agreement with prior learners. To date, more than 300 people have achieved high inter- coder agreement on this test.
Jeffrey Cohn, University of Pittsburgh, says: "FACS is still a subjective method, but it's rigorously based on description of facial motion. Therefore, it provides a ground truth that can be used in expression recognition."

Is the coding scheme language dependent (which language(s)) or language independent?

The coding scheme is language independent.

Is there tools support for using the coding scheme or API for editing/parsing coded descriptions?

In order to recognize the subtle changes of facial expressions, several researchers propose to recognize minimal facial signals and combine the signals to recognize the complete facial expressions. That is rather than trying to recognize the entire facial expression they are working on recognizing singular facial actions. The facial expression is then deduced by combining the several facial actions. This recognition method is based on The Facial Action Coding System, FACS. FACS is a system to measure facial signals using minimal action units (AUs). This recognition method follows the same logic as FACS as it looks at singular facial actions.
Cohn developed a facial recognition system composed of three modules. These modules are used to extract information on facial actions. One module extracts information on particular facial features (e.g., brows, mouth), another gets data from larger facial regions (such as chin and cheek) and the final one looks at the appearance of the wrinkles and furrows. The combination of the information provided from the three modules give good and precise results. More precisely, the feature-point tracking module tracks feature points within a small feature window. The authors also employ a neural network to recognize the action units after the facial features are correctly extracted and suitably represented. Eleven basic lower face action units and combinations (Neutral, AU9, AU 10, AU 12, AU 15, AU 17, AU 20, AU 25, AU 26, AU 27, and AU23+24) and seven basic upper face action units (Neutral, AU1, AU2, AU4, AU5, AU6, AU7) are identified by a single neural network for lower face and upper face separately. The recognition rate results are comparable to the rate obtained by high-trained FACS

coders. The results of recognising single AUs (except for the combination of AU23 + AU24) is more than 95% of agreement.

Bartlett has examined several techniques to measure facial actions from the upper face: PCA, optical flow and feature measurement. The authors also combined all three methods (hybrid system) in a single neural network. The four techniques were compared on the same dataset of face image sequences. The dataset was built by asking an actor to perform specific facial actions (corresponding to Action Units as defined in FACS). Each sequence started with the neutral expression. All the faces in the datasets were transformed (scaled, rotated...) so that the facial features of all the faces were aligned. Moreover the images were cropped in order to contain only the upper part of the face. The best recognition result is obtained for the hybrid system: 92.2%. The results are compared with recognition rate from both naive and expert coders. To both types of coders were given a set of pair of images consisting in the neutral image along with the test image. The neutral image serves as a reference basis for the recognition task. Test images contained different faces performing several AUs at low, medium, and high intensity. A training session was given to the naive subjects. Naive coders arrived to 73.7% agreement while expert coders show 91.8% agreement in their recognition results. The recognition rate obtained with the hybrid method is similar to the one from the expert coders. This is an encouraging results. But as the authors pointed out the system should be tested on spontaneous expressions. These expressions are often blended expressions increasing enormously the complexity of the recognition task.

Bartlett, M.S., Hager, J.C., Ekman, P., and Sejnowski, T.J.: Measuring facial expressions by computer image analysis. Journal of Psychophysiology, vol. 36, pp. 253-263, 1999.

Bartlett, M.S., Viola, P. A., Sejnowski, T. J., Golomb, B.A., Larsen, J., Hager, J. C., and Ekman, P.: Classifying facial action. Advances in Neural Information Processing Systems 8, MIT Press, Cambridge, MA. p. 823-829, 1996.

Donato, G.L., Bartlett, M.S., Hager, J.C., Ekman, P., and Sejnowski, T.J.: Classifying Facial Actions. IEEE Transactions on Pattern Analysis and Machine Intelligence 21(10) p. 974-989. 1999.

Kanade, T., Cohn, J. F., and Tian, Yingli: Comprehensive Database for Facial Expression Analysis. The 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), France. March 2000. Can be downloaded from: http://www.cs.cmu.edu/afs/cs/project/face/www/world-ftp/FG3.pdf

Tian, Yingli, Kanade, T., and Cohn, J. F.: Recognizing Action Units for Facial Expression Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 2, February, 2001. Can be downloaded from: http://www.cs.cmu.edu/afs/cs/project/face/www/world-ftp/PAMI_YLfinal.ps.gz

Tian, Yingli, Kanade, T., and Cohn, J. F.: Recognizing Upper Face Action Units for Facial Expression Analysis: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR'00), Hilton Head Island, South Carolina. June 2000. Can be downloaded from: http://www.cs.cmu.edu/afs/cs/project/face/www/world-ftp/cvpr00camera.pdf

Tian, Yingli, Kanade, T., and Cohn, J. F.: Recognizing Lower Face Action Units for Facial Expression Analysis. The 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), France. March 2000. Can be downloaded from:
http://www.cs.cmu.edu/afs/cs/project/face/www/world-ftp/FG2camera.ps.gz

2.2.6 Accessibility

How does one get access to the coding scheme?

One can get access to the coding scheme by contacting:
Paul Ekman, Ph.D.
Professor of Psychology
Department of Psychiatry
University of California
San Francisco
email: ekmansf@itsa.ucsf.edu

Furthermore the coding scheme is thoroughly described in Ekman and Friesen, 1978 and 1975 and Ekman and Rosenberg, 1997.

Is the coding scheme available for free or how much does it cost?

The coding scheme described in the self-instructional system includes: a 300 page manual; 146 illustrative facial photographs; practice material; Investigator's Guide, Part 1; Investigator's Guide, Part 2; Score Sheets; and a VHS Video. The price: $320.00, plus shipping. Payment must be in advance. One must make a check payable to Paul Ekman in U.S. Dollars, and mail it to: Paul Ekman, UCSF/Psychiatry, 401Parnassus Avenue, San Francisco, CA 94143-0984.

2.2.7 Conclusion

How well described is the coding scheme?

FACS is extremely powerful to describe facial expressions of emotion. It is very widely used.

How general and useful is the coding scheme?

The coding scheme is able to encode any visible facial action. It is specially indicated to encode facial expressions of emotion. It is not particularly indicated to encode lip movement during speech.

2.3 The Maximally Discriminative Facial Movement Coding System (MAX)

2.3.1 Description header

Main actor

NISLab: Malene Wegener Knudsen (mwk@nis.sdu.dk), Laila Dybkjær (laila@nis.sdu.dk) and Niels Ole Bernsen (nob@nis.sdu.dk)

Verifying actor

UROME: Catherine Pelachaud (cath@dis.uniroma1.it) and Isabella Poggi (poggi@uniroma3.it)

Date of last modification of the description

June 18th, 2001.

2.3.2 Reference

Web site

The coding scheme has no web site.
An evaluation of the coding scheme can be found in: http://www.pitt.edu/~jeffcohn/fp.pdf

Short description

The coding scheme was created in 1983 by Carroll Izard at the University of Delaware. It was last revised in 1995. The principal reason for developing the coding system was to provide an efficient, reliable and valid system for measuring the emotion signals in the facial behaviours of infants and young children. With some modifications in the descriptions of changes in the appearance the coding scheme can be used to measure emotion signals at any age.

An illustrative example of coding

No example of coding with a corresponding sample of the resource has been available to the main actor, but drawn examples representing the different MAX codes can be found in the book referenced below. Two examples from the book can be seen here. The first example shows a face with no observable movement with a smooth forehead, brows in resting position, eyes open and mouth closed and relaxed (MAX code 0/0/0), cf. figure 1.3.1. The second example is a representation of MAX code no. 22/0/0: The brows are raised and drawn together. There is a thickening or bulging of the mid-region of the forehead and the nasal root is narrowed, cf. figure 1.3.2.

Figure 2.3.1. An example of MAX code 0/0/0. Neutral face.

Figure 2.3.2. An example of MAX code 22/0/0. The brows are raised and drawn together. There is a thickening or bulging of the mid-region of the forehead and the nasal root is narrowed The eyes, nose, cheeks, lips and mouth are in neutral position.

References to additional information on the coding scheme

Izard, C.E.: The Maximally Discriminative Facial Movement Coding System. Academic Computing Services and University Media Services, University of Delaware, Newark, Delaware. Revised Edition, January, 1983.

2.3.3 Coverage

Which types of raw data are referenced?

MAX has been used on photographs taken in medical clinics for healthy babies. The procedure for obtaining the photographs has been to take photographs during four different phases during visits by mothers and their infants to the clinic. The four situations are: playful interaction between the mother, nurse and infant before the medical procedure, the medical procedure itself, the mother's comforting of the infant after the medical procedure, and a brief period of playful interaction between the mother, nurse and infant at the end of the visit. The photographs have been transferred to videotapes, were they are shown in a slide-show manner.

Which modalities is the coding scheme meant to code?

The coding scheme is meant to cover face corpora in the form of video recording of the faces of babies of the age of 0-2 years. With a small adaptation of the coding scheme, the scheme can be used to cover other face data than infant faces.

Which annotation level(s) does the coding scheme cover?

The coding scheme covers the annotation level of facial expression.

Which coding tasks has the coding scheme been used for?

The coding scheme has been used for coding expressions of affect and the emotions of interest, joy, surprise, sadness, anger, fear and disgust which were all identified in infants aged 1 to 9 months.

2.3.4 Detailed description of coding scheme

Which header file information is included (meta-data)?

The meta-data header file includes information of which tape number one is coding from, which age the infant has, which coder is coding, the beginning and end time of coding and which date the coding is taking place.

Coding purpose of the coding scheme?

The principal reason for developing the coding system was to provide an efficient, reliable and valid system for measuring the emotion signals in the facial behaviours of infants and young children.

List and description of phenomena, which can be annotated by the scheme

The following phenomena can be annotated by the scheme:

1. Interest-Excitement (IE)
2. Enjoyment-Joy (EJ)
3. Surprise-Astonishment (SA)
4. Sadness-Dejection (SD)
5. Anger-Rage (AR)
6. Fear-Terror (FT)
7. Discomfort-Pain (DP) (Physical distress)

The description of the seven phenomena is the following:

1. IE consists of brows that are drawn together, no observable movements in the eyes, nose and cheeks and maybe slightly lowered and opened and relaxed lips and mouth.

2. EJ consists of no observable movement in the brows, forehead and nasal root, narrowed or squinted eyes, nose and cheeks and corners of the mouth pulled back and slightly up. The mouth may be opened or closed.
3. SA consists of raised brows in normal or arched shape, long transverse furrows or thickening across the forehead and a narrowed nasal root, wide open, round cheeks and opened, round or oval mouth.
4. SD consists of brows where the inner corners raised and changed in shape under inner corner, the forehead bulges in centre above the brow corners and the nasal root is narrowed. The eyes, nose and cheeks are narrowed or squinted and the corners of the mouth are drawn downward/outward. The chin may push up the centre of the lower lip and the mouth may be opened or closed.
5. AR consists of brows that are sharply lowered and drawn together, vertical furrows across the forehead or a bulge between the brows and a broadened and bulged nasal root. The eyes/nose/cheeks are narrowed or squinted and an angular, squarish mouth, probably open.
6. FT consists of brows that are raised and drawn together with a straight or normal shape. The forehead has short transverse furrows or thickening in the mid-region and the nasal root is narrowed. The eye fissure is wide opened, the upper lid is raised and the eyes show more white than normal. The mouth is opened with tense corners that are retracted straight back.
7. DP consists of sharply lowered and drawn together brows. Vertical furrows across the forehead or a bulge between the brows and a broadened and bulged nasal root. The eye fissure is scrunched and tightly closed and the mouth is angular, squarish and open.

Description of markup language/markup declaration

The markup language consists of 68 MAX number codes, each of which represents a certain expression of the face. The expressions are divided into placement and look of the brows, forehead, nasal root, eyes, nose, cheeks, lips and mouth. The expressions are grouped into regional groups consisting of 1: brows, forehead and nasal root, 2: eyes, nose and cheeks, and 3: lips and mouth. The description of the expression of each number code is based on the anatomically possible movements of the facial muscles and is a description of what the face looks like when the movements have taken place. Region 1 is coded first, then region 2 and last region 3.

Max codes	
Codes for brows (B), forehead (F) and nasal root (N) (from 20-29) *	
Code no.	
0	No observable movement. Smooth forehead, brows in resting position.
20	B: raised in normal or arched shape. F: long transverse furrows or thickening. N: narrowed.
21	B. one brow raised higher than the other – other may be slightly lowered.
22	B: raised, drawn together, straight or normal shape. F: short transverse furrows or thickening in mid-region. N: narrowed.
23	B: inner corners raised, change in shape under inner corner. F: bulge or furrows in centre above brow corners. N: narrowed.
24	B: drawn together, may be slightly lowered. F: vertical furrows or bulge between brows.
25	B: sharply lowered and drawn together. F: vertical furrows or bulge between brows. N: broadened, bulged.
Codes for eyes, nose and cheeks (from code 30-49) *	
0	No observable movement. Eyes normally open.
31	Eye fissure widened, upper lid raised – white in the eyes shows more than normal.
33	Narrowed or squinted eyes, by action of eye sphincters or brow depressors.
34	Squinting without cheek movement.
35	Visual scanning.
37	Eye fissure scrunched, tightly closed.

42	Nasal bridge furrowed or shows lumpy ridge running diagonally upward from nasolabial fold.
Codes for lips and mouth (from code 50-63) *	
0	No observable movement. Mouth closed and relaxed.
50	Opened, round or oval.
51	Opened, relaxed.
52	Corners, pulled back and slightly up (opened or closed).
53	Opened, tense, corners retracted straight back.
54	Angular, squarish (open).
55	Open, tense.
56	Corners drawn downward/outward (open or closed) chin may push up centre of lower lip.
57	Mouth corners slightly retracted; lip corners press against teeth.
58	Bilateral 57.
59A	Opened relaxed; tongue forward beyond gum line, may be moving.
59B	Opened, angular, upper lip pulled up; tongue forward beyond gum, may be moving.
61	Upper lip raised on one side.
63	Lower lip lowered (may be slightly forward).
Other codes	
NS	Non-scorable. Movement is suspected, but not clear enough to code.
OBS	Obscured. Area to be coded is out of view for at least one second.
NC	Non-codable.

Figure 2.3.3. Description of the MAX codes. * It has not been possible to find description of the codes 1-19, 25-30, 32, 36, 38-41, 43-39, 60, 62 and 64-68.

Examples

Examples are MAX code 20 where the brows are raised in normal or arched shape and long transverse furrow or a thickening shows across the forehead together with a narrowed nasal root. MAX code 34 represents squinting eyes and nose without cheek movements and MAX code 57 represents a mouth where the corners are slightly retracted and the lip corners are pressed against the teeth.

Another code it O – which is no observable movement, NS – Non-scorable movement, where movement is suspected but not clear enough to code, OBS – Obscured view, where the area to be coded is out of view for a period and NC – Non-codable, where the coder is not sure of which code to apply.

Since the corresponding samples from the resource only exist on VHS videotape, it is not possible to put in an illustration of the examples.

Description of coding procedure, if any

The coding procedure is based on a situation, where one has coders who are trained in the coding system available for one's coding. One can become a trained coder, through the use of the MAX training videotapes, which can be obtained by contacting Carroll Izard – izard@udel.edu. The MAX system has its own procedure especially for coders using a video machine with a slow motion control, but the procedure should be applicable to any equipment.

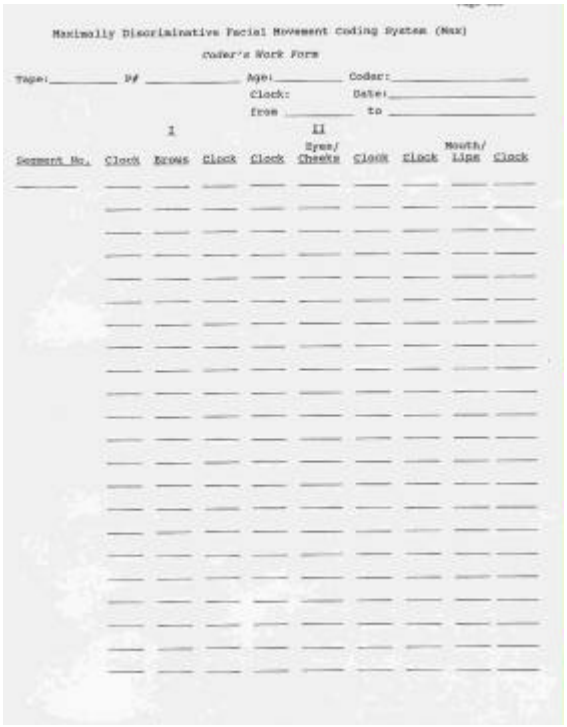
A. General rules:

1. Adjust the brightness and contrast of the monitor as necessary; recheck these adjustments any time the picture becomes unclear.
2. It is helpful to divide the material to be coded into definite time units or segments of 3-10 second in length.
3. Code only one facial region at a time.

4. Never allow one region's muscle movements or appearance change to affect judgement of movements in other facial regions.
5. Do not let a joystick controlled tape stay on one frame for more than 30 sec. at a time. It will damage the tape.
6. Viewing in real time is essential. It is often useful and sometimes essential in confirming codes detected in the customary stop-frame/slow-motion mode. Because infants' faces often vary greatly in structure and amount of fatty tissue, it is also strongly advised that coders initially view all segments in real time to obtain a judgement of baseline facial tone. It is good practice to review coded movements in real time.

B. Movement coding:

1. Start the playback of the videotape and divide the tape into segments of 3-10 seconds.
2. Having divided the videotape into segments, start the playback again and note which tape and segment number you are coding at the moment in the space saying Segment No. In the Coder's Work Form, cf. figure 1.3.4. Place the onset time of the segment under the brow column.


Maximally Discriminative Facial Movement Coding System (Max)

Coder's Work Form

Tape:_____ Subj _____ Age:_____ Coder:_____

Clock: Date:_____

from _____ to _____

	I			II			
Segment No.	Clock	Brows	Clock	Clock	Eyes/ Cheeks	Clock	Clock	Mouth/ Lips	Clock
_____									

Figure 2.3.4. Coder's Work Form.

3. Begin by running through the segment in real time concentrating on the brow region only (repeat the real-time run for each area, before coding it)
4. Next, slowly proceed through the segment with the joystick (slow-motion/stop frame) control. Watch brow area for any muscle activity or appearance change described in the MAX codes. When a movement is observed find the code that best describes it.
5. Find the onset and offset times of the appearance change and along with the appearance change code note the times on the coding form.
6. Code the entire segment in this way, concentrating only on the brows. Code the brow region for a set of three segments.
7. Repeat steps 1-6 for the eye/cheek region and then again for the mouth/chin region for the set of 3 segments. Before proceeding to the next set of 3 segments, compute your reliability for the segments you have just coded.

Reliability or percent agreement = $\frac{\text{agreements}}{\text{agreements+disagreements}}$

C. Time coding:

1. Time (or frame number) should be displayed in one corner of the image of any video recordings to be coded.
2. To find the onset time of a movement or appearance change, proceed until the movement peaks or is clearly visible. Then move backward until the movement stops. Conform or adjust the onset time by moving the tape forward and backward to the estimated onset point.
3. Find the offset time in the same manner.
4. Code all onset and offset times to the nearest 0.1 sec.

D. Zero(0), Non-coded Movement (NM), Non-scorable (NS) and Obscure (Obs)

1. 0 may be used for a region when it shows no movement. When all 3 regions are 0, it indicate that the baby is awake =0/0/0 or asleep = 0/0/0
2. Movements not described as affect signals in MAX are coded NC (non-coded movement). The region may show movement, but if this is not part of one of the appearance changes in the MAX codes, it is designated NC.
3. NS: The guide for scoring NS is visibility of the target movement. Ideally the face should be zoomed in so that it is approximately life-size or larger in clear focus and adequately lighted to prevent shadows. However, these conditions need not be met completely if the target movement is clearly visible. A good check as to whether a segment should be designated NS is to ask oneself how difficult it is to make the judgement. If one has to strain or think ones judgement is partly guesswork then the segment should be marked NS. Another check is to have two trained coders attempt to code the questionable segment. If either coder marks it NS, it should be so designated.
4. Obscure (Obs): Score the whole face or any region when for more than 1 second the face or target area is blocked from view or turned such that less than two thirds of the face measured horizontally is visible. The two-thirds rule can be disregarded if one is sure that one can code the region under consideration.
5. If one half of the face measured vertically is Obs and the other half predict an emotion expression, it should be written as Obs/code# or code#/Obs.

E. Further suggestions for coding:

1. Be alert to make the effect of head tilting. For example head tilting can often make stationary brows appear to be moved up or eyes to be widened.
2. If unsure whether eyes are narrowed, blinked shut or tightly closed, proceed slowly through the sequence looking for glints or other indication that the eyelids are parted and for bulges and furrows on and around closed eyelids. If the eyes are closed only for a fraction of a second without tissue displacement on and around the eyelids, it is probably a blink, which can be coded 0 or ignored.
3. Be alert to any apparent bulges or dimples that may be part of the baby's facial structure.
4. When trying to see cheeks raising watch for the deepening (or formation) of the furrow under the eye.

Creation notes, i.e. who wrote the coding scheme, when,and in which context?

The coding scheme was created in 1983 by Carroll Izard at the University of Delaware. It was last revised in 1995.

Contact person:

Professor Carroll E. Izard

The department of Psychology

105 McKinly Lab
220 Wolf Hall
University of Delaware
Newark, DE 19716
E-mail: izard@udel.edu

2.3.5 Usage

Origin of the coding scheme and reasons for creating it

The principal objective in developing MAX was to provide an efficient, reliable, and valid system for measuring the emotion signals in the facial behaviours of infants and young children. With some modifications in the descriptions of the appearance changes the coding scheme can be used to measure emotion signals at any age.

How many people have used the coding scheme and for what purposes?

No information available.

How many dialogues/interactions have been annotated using the coding scheme?

No information available.

Has the coding scheme been evaluated?

The coding scheme has been evaluated and compared with FACS by another facial expression researcher, Professor Jeffrey Cohn at Pittsburgh University. The evaluation can be found in a paper that can be downloaded from: http://www.pitt.edu/~jeffcohn/fp.pdf

Is the coding scheme language dependent (which language(s)) or language independent?

The coding scheme is language independent.

Is there tools support for using the coding scheme or API for editing/parsing coded descriptions?

There is no tools support for the coding scheme.

2.3.6 Accessibility

How does one get access to the coding scheme?

The description of the coding scheme is based on (Izard, 1983).
Contact person:
Professor Carroll E Izard
The department of Psychology
105 McKinly Lab
220 Wolf Hall
University of Delaware
Newark, DE 19716
E-mail: izard@udel.edu
Homepage: http://www.psych.udel.edu/people/detail.php?firstname=Carroll&lastname=Izard

Is the coding scheme available for free or how much does it cost?

The coding scheme is available for free.

2.3.7 Conclusion

How well described is the coding scheme?

Due to being less comprehensive and intended for a more narrow area than FACS, MAX has been used and described less than FACS.

How general and useful is the coding scheme?

Compared with FACS, MAX is less comprehensive and was intended to include only facial displays (referred to as movements in MAX) related to emotion. MAX does not allow one to encode facial expressions that are not emotions while FACS does. MAX does not discriminate among some anatomically distinct displays (e.g. inner and outer brow raises) and considers as autonomous some movements that are not anatomically distinct. Unlike FACS, MAX makes explicit claims that specific combinations of displays are expressions of emotion, and the goal of MAX coding is to identify these MAX-specified emotion expressions.

2.4 MPEG-4 SNHC (Moving Pictures Expert Group, Synthetic/Natural Hybrid Coding)

2.4.1 Description header

Main actor

NISLab: Malene Wegener Knudsen (mwk@nis.sdu.dk), Laila Dybkjær (laila@nis.sdu.dk) and Niels Ole Bernsen (nob@nis.sdu.dk)

Verifying actor

UROME: Catherine Pelachaud (cath@dis.uniroma1.it) and Isabella Poggi (poggi@uniroma3.it)

Date of last modification of the description

June 21st, 2001.

2.4.2 Reference

Web site

MPEG-4 SNHC homepage: http://www.es.com/mpeg4-snhc/index.html

A description of MPEG-4: http://ligwww.epfl.ch/mpeg4

The ACTS MPEG Forum: http://www.ccir.ed.ac.uk/mpeg4/

A description of parameter based facial animation: http://www.research.att.com/~osterman/AnimatedHead/index.html

A description of Face and 2-D Mesh Animation in MPEG-4: http://www.cselt.it/leonardo/icjfiles/mpeg-4_si/8-SNHC_visual_paper/8-SNHC_visual_paper.htm

A description of MPEG-4 compliant facial animation and Hybrid Video Coding: http://www-dsp.com.dist.unige.it/~pok/RESEARCH/index.htm

The home page of the Moving Picture Experts Group: http://www.cselt.it/mpeg

Short description

MPEG-4 is an object-based multimedia compression standard, which allows to encode different audio-visual objects (AVO). MPEG-4 SNHC group proposes an architecture for the efficient representation and coding of synthetically and naturally generated audio-visual information.

An illustrative example of coding

The picture of the 3D face below has been generated from the FAPS value listed below. They come from the wow.fap example that can be found in the web pages of MPEG-4. The image is the frame 71 of the file wow.fap:

0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0

71 97 -52 -175 -33 25 -16 -48 -92 -106 22 0 -55 56 -54 0 0 0 166 103 145 155 68 102 -24 0 -36 5 0

Figure 2.4.1. Frame 71 of the wow.fap.

References to additional information on the coding scheme

Tutorial issue on the MPEG-4 standard: Elsevier (http://www.cselt.it/leonardo/icjfiles/mpeg-4_si/)

P. Doenges ,F. Lavagetto, J. Ostermann, I.S. Pandzic and E. Petajan: MPEG-4: Audio/Video and Synthetic Graphics/Audio for Mixed Media. Image Communications Journal, vol. 5(4), May 1997.

J. Ostermann: Animation of synthetic faces in MPEG-4. Computer Animation'98, Philadelphia, USA, pp. 49-51, June 1998.

E. Petajan: Facial Animation Coding, Unofficial Derivative of MPEG-4. Work-in-Progress, Human Animation Working Group, VRML Consortium, 1997.

2.4.3 Coverage

Which types of raw data are referenced?

The raw data represents the displacements of the feature points defined by MPEG-4 of a facial model.

Which modalities is the coding scheme meant to code?

Facial expression of synthetic agents.

Which annotation level(s) does the coding scheme cover, e.g. facial expression plus prosody?

Facial expression

Which coding tasks has the coding scheme been used for?

The coding scheme has been used to drive synthetic facial models.

2.4.4 Detailed description of coding scheme

Which header file information is included (meta-data)?

A file FAP (Facial Animation Parameters) is constituted of a header and a sequence of FAP masks and FAP values, one set per frames. The header record is of the form: 'version number' 'sequence base name' 'frame rate' 'sequence length':

- version number: 2.0
- sequence base name: note:face feature filename= <basename>.fff,
- video filename= <basename>.vid, separate image filenames= <basename>.nnnnn.img,audio filename= <basename>.aud,
- FAP filename = <basename>.fap, FDP filename = <basename>.fdp,
- BAP filename = <basename>.bap, BDP filename = <basename>.bdp
- frame rate: fps (note that this may be equal to the video field rate since interlace has no meaning here)
- sequence length: Nframe

Coding purpose of the coding scheme?

The coding purpose of the coding scheme is to define a set of parameters to define and control facial models.

The shape, texture and expressions of the face are controlled by Facial Definition Parameters (FDPs) and/or Facial Animation Parameters (FAPs). Upon construction, the face object contains a generic face with a neutral expression. This face can already be rendered. It can also immediately receive the animation parameters from the bitstream, which will produce animation of the face: expressions, speech etc. Meanwhile, definition parameters can be sent to change the appearance of the face from something generic to a particular face with its own shape and (optionally) texture. If so desired, a complete face model can be downloaded via the FDP set.

Deforming a neutral face model according to some specified FAP values at each time instant generates a facial animation sequence. The FAP value for a particular FAP indicates the magnitude of the corresponding action, e.g., a big versus a small smile or deformation of a mouth corner. For an MPEG-4 terminal to interpret the FAP values using its face model, it has to have predefined model specific animation rules to produce the facial action corresponding to each FAP. The terminal can either use its own animation rules or download a face model and the associated face animation tables (FAT) to have a customized animation behaviour. Since the FAPs are required to animate faces of different sizes and proportions, the FAP values are defined in face animation parameter units (FAPUs). The FAPU are computed from spatial distances between major facial features on the model in its neutral state.

A FAPU and the feature points used to derive the FAPU are defined next with respect to the face in its neutral state.

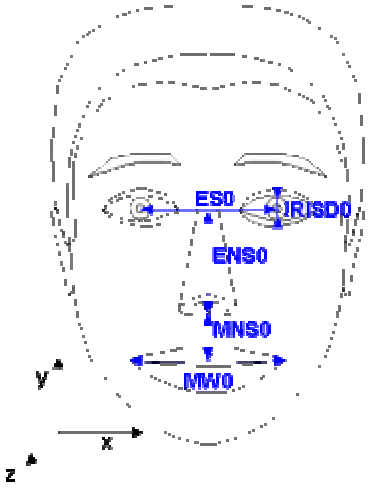


Figure 2.4.2. A face model in its neutral state and the feature points used to define FAP units (FAPU). Fractions of distances between the marked key features are used to define FAPUs (Face Animation Parameter Units).

In order to define face animation parameters for arbitrary face models, MPEG-4 defines FAPUs that serve to scale facial animation parameters for any face model. FAPUs are defined as fractions of distances between key facial features (see Figure 2.4.2.). These features, such as eye separation, are defined on a face model that is in the neutral state. The FAPU allows interpretation of the FAPs on any facial model in a consistent way, producing reasonable results in terms of expression and speech pronunciation. The measurement units are shown in figure 2.4.3.

IRISD0	Iris diameter (by definition it is equal to the distance between upper and lower eyelid) in neutral face	IRISD = IRISD0 / 1024
ES0	Eye separation	ES = ES0 / 1024
ENS0	Eye - nose separation	ENS = ENS0 / 1024
MNS0	Mouth - nose separation	MNS = MNS0 / 1024
MW0	Mouth width	MW=MW0 / 1024
AU	Angle unit	10E-5 rad

Figure 2.4.3. Facial animation parameter units and their definitions.

Feature Points:

MPEG-4 specifies 84 feature points on the neutral face (see figure 2.4.4). The main purpose of these feature points is to provide spatial references for defining FAPs. Some feature points such as the ones along the hairline are not affected by FAPs. However, they are required for defining the shape of a proprietary face model using feature points. Feature points are arranged in groups like cheeks, eyes, and mouth. The location of these feature points has to be known for any MPEG-4 compliant face model. The feature points on the model should be located according to figure 2.4.4.

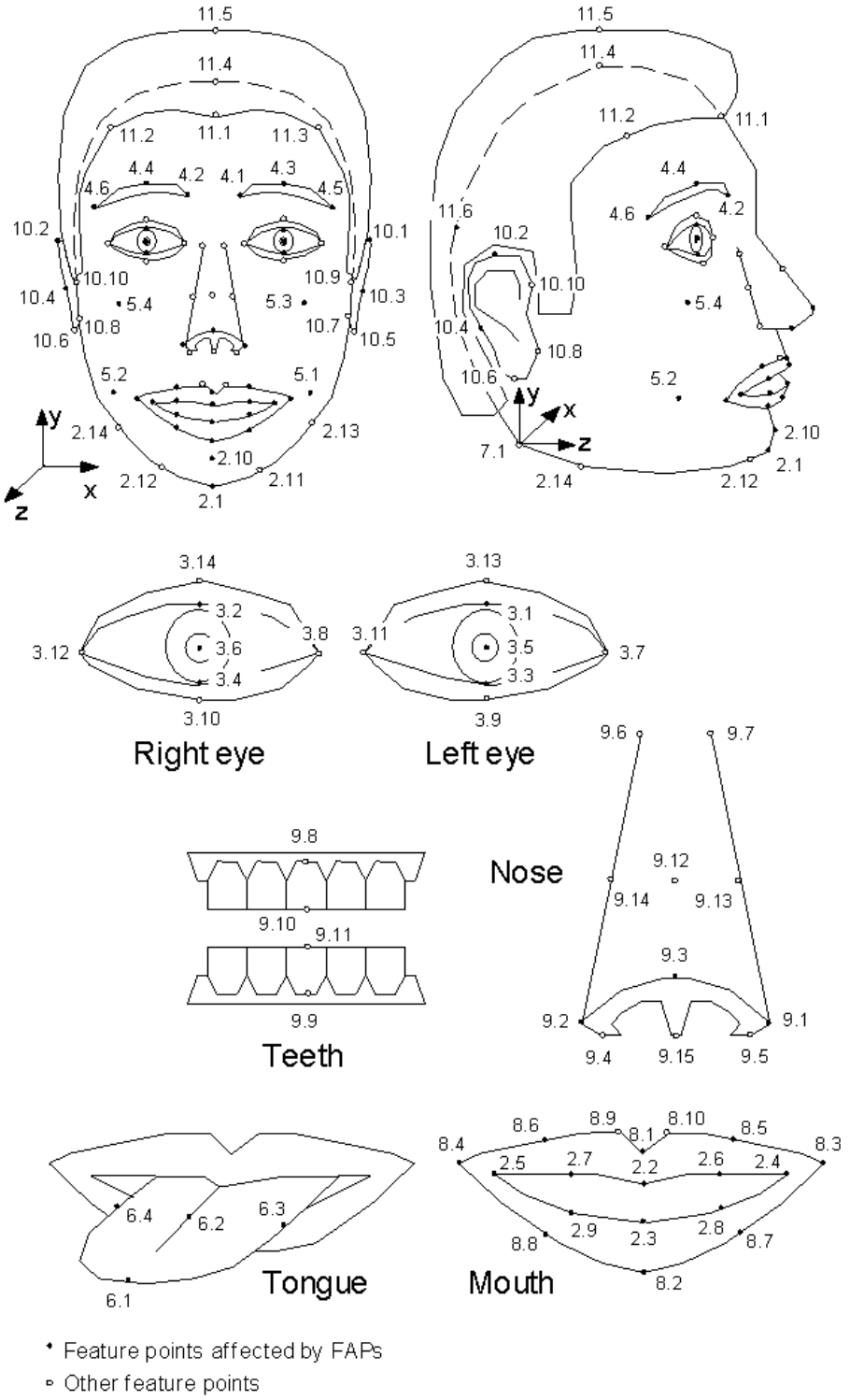


Figure 2.4.4. Feature points may be used to define the shape of a proprietary face model. The facial animation parameters are defined by motion of some of these feature points.

Face Animation Parameters:

The FAPs are based on the study of minimal perceptible actions and are closely related to muscle actions. FAPs represent a complete set of basic facial actions including head motion, tongue, eye, and mouth control. They allow representation of natural facial expressions. There are 68 parameters that are categorized into 10 groups related to parts of the face, cf. figure 2.4.5. For each FAP, the standard defines the appropriate FAPU, FAP group, direction of positive motion and whether the motion of the feature point is unidirectional (see e.g. FAP 3, open jaw) or bi-directional (see e.g. FAP 48, head

pitch). FAPs can also be used to define facial action units. Exaggerated amplitudes permit the definition of actions that are normally not possible for humans, but are desirable for cartoon-like characters.

The FAP set contains two high-level parameters, visemes and expressions (FAP group 1). A viseme (FAP 1) is a visual correlate to a phoneme. Only 14 static visemes that are clearly distinguished are included in the standard set. In MPEG-4, transitions from one viseme to the next are defined by blending only two visemes with a weighting factor.

The expression parameter FAP 2 defines the 6 primary facial expressions (anger, joy, fear, sadness, disgust and surprise). In contrast to visemes, facial expressions are animated by a value defining the excitation of the expression. Two facial expressions can be animated simultaneously with amplitude in the range of [0-63] defined for each expression. The facial expression parameter values are defined by textual descriptions. The expression parameter allows for an efficient means of animating faces. They are high-level animation parameters. A face model designer creates them for each face model.

The remaining FAPs (66) are clustered in different groups (such as outer lip, cheeks, eyebrow). With the exception of some FAPs which control the head rotations, the eyeball rotations etc, each low-level FAP indicates the translation of the corresponding feature point, with respect to its position in the neutral face, along one of the coordinate axes.

List and description of phenomena, which can be annotated by the scheme

Figure 2.4.5 contains the list and descriptions of the 68 FAPS.

#	**FAP name**	**FAP description**	Units	Uni- or bidirectiona	Position motion	Group	FDP subgroup number
1	viseme	Set of values determining the mixture of two visemes for this frame (e.g. pbm, fv, th)	na	na	na	1	na
2	expression	A set of values determining the mixture of two facial expression	na	na	na	1	na
3	open_jaw	Vertical jaw displacement (does not affect mouth opening)	MNS	U	down	2	1
4	lower_t_midlip	Vertical top middle inner lip displacement	MNS	B	down	2	2
5	raise_b_midlip	Vertical bottom middle inner lip displacement	MNS	B	up	2	3
6	stretch_l_cornerlip	Horizontal displacement of left inner lip corner	MW	B	left	2	4
7	stretch_r_cornerlip	Horizontal displacement of right inner lip corner	MW	B	right	2	5
8	lower_t_lip_lm	Vertical displacement of midpoint between left corner and middle of top inner lip	MNS	B	down	2	6
9	lower_t_lip_rm	Vertical displacement of midpoint between right corner and middle of top inner lip	MNS	B	down	2	7
10	raise_b_lip_lm	Vertical displacement of midpoint between left corner and middle of bottom inner lip	MNS	B	up	2	8
11	raise_b_lip_rm	Vertical displacement of midpoint between right corner and middle of bottom inner lip	MNS	B	up	2	9
12	raise_l_cornerlip	Vertical displacement of left inner lip corner	MNS	B	up	2	4
13	raise_r_cornerlip	Vertical displacement of right inner lip corner	MNS	B	up	2	5
14	thrust_jaw	Depth displacement of jaw	MNS	U	forward	2	1
15	shift_jaw	Side to side displacement of jaw	MW	B	right	2	1
16	push_b_lip	Depth displacement of bottom middle lip	MNS	B	forward	2	3
17	push_t_lip	Depth displacement of top middle lip	MNS	B	forward	2	2
18	depress_chin	Upward and compressing movement of the chin (like in sadness)	MNS	B	up	2	10
19	close_t_l_eyelid	Vertical displacement of top left eyelid	IRISD	B	down	3	1
20	close_t_r_eyelid	Vertical displacement of top right eyelid	IRISD	B	down	3	2
21	close_b_l_eyelid	Vertical displacement of bottom left eyelid	IRISD	B	up	3	3

#	FAP name	FAP description	Units	Uni- or bidirectiona	Position motion	Group	FDP subgroup number
22	close_b_r_eyelid	Vertical displacement of bottom right eyelid	IRISD	B	up	3	4
23	yaw_l_eyeball	Horizontal orientation of left eyeball	AU	B	left	3	5
24	yaw_r_eyeball	Horizontal orientation of right eyeball	AU	B	left	3	6
25	pitch_l_eyeball	Vertical orientation of left eyeball	AU	B	down	3	5
26	pitch_r_eyeball	Vertical orientation of right eyeball	AU	B	down	3	6
27	thrust_l_eyeball	Depth displacement of left eyeball	ES	B	forward	3	5
28	thrust_r_eyeball	Depth displacement of right eyeball	ES	B	forward	3	6
29	dilate_l_pupil	Dilation of left pupil	IRISD	B	growing	3	5
30	dilate_r_pupil	Dilation of right pupil	IRISD	B	growing	3	6
31	raise_l_i_eyebrow	Vertical displacement of left inner eyebrow	ENS	B	up	4	1
32	raise_r_i_eyebrow	Vertical displacement of right inner eyebrow	ENS	B	up	4	2
33	raise_l_m_eyebrow	Vertical displacement of left middle eyebrow	ENS	B	up	4	3
34	raise_r_m_eyebrow	Vertical displacement of right middle eyebrow	ENS	B	up	4	4
35	raise_l_o_eyebrow	Vertical displacement of left outer eyebrow	ENS	B	up	4	5
36	raise_r_o_eyebrow	Vertical displacement of right outer eyebrow	ENS	B	up	4	6
37	squeeze_l_eyebrow	Horizontal displacement of left eyebrow	ES	B	right	4	1
38	squeeze_r_eyebrow	Horizontal displacement of right eyebrow	ES	B	left	4	2
39	puff_l_cheek	Horizontal displacement of left cheek	ES	B	left	5	1
40	puff_r_cheek	Horizontal displacement of right cheek	ES	B	right	5	2
41	lift_l_cheek	Vertical displacement of left cheek	ENS	U	up	5	3
42	lift_r_cheek	Vertical displacement of right cheek	ENS	U	up	5	4
43	shift_tongue_tip	Horizontal displacement of tongue tip	MW	B	right	6	1
44	raise_tongue_tip	Vertical displacement of tongue tip	MNS	B	up	6	1
45	thrust_tongue_tip	Depth displacement of tongue tip	MW	B	forward	6	1
46	raise_tongue	Vertical displacement of tongue	MNS	B	up	6	2
47	tongue_roll	Rolling of the tongue into U shape	AU	U	concave upward	6	3, 4
48	head_pitch	Head pitch angle from top of spine	AU	B	down	7	1
49	head_yaw	Head yaw angle from top of spine	AU	B	left	7	1
50	head_roll	Head roll angle from top of spine	AU	B	right	7	1
51	lower_t_midlip _o	Vertical top middle outer lip displacement	MNS	B	down	8	1
52	raise_b_midlip_o	Vertical bottom middle outer lip displacement	MNS	B	up	8	2
53	stretch_l_cornerlip_o	Horizontal displacement of left outer lip corner	MW	B	left	8	3
54	stretch_r_cornerlip_o	Horizontal displacement of right outer lip corner	MW	B	right	8	4
55	lower_t_lip_lm _o	Vertical displacement of midpoint between left corner and middle of top outer lip	MNS	B	down	8	5
56	lower_t_lip_rm _o	Vertical displacement of midpoint between right corner and middle of top outer lip	MNS	B	down	8	6
57	raise_b_lip_lm_o	Vertical displacement of midpoint between left corner and middle of bottom outer lip	MNS	B	up	8	7
58	raise_b_lip_rm_o	Vertical displacement of midpoint between right corner and middle of bottom outer lip	MNS	B	up	8	8
59	raise_l_cornerlip_o	Vertical displacement of left outer lip corner	MNS	B	up	8	3
60	raise_r_cornerlip _o	Vertical displacement of right outer lip corner	MNS	B	up	8	4
61	stretch_l_nose	Horizontal displacement of left side of nose	ENS	B	left	9	1
62	stretch_r_nose	Horizontal displacement of right side of nose	ENS	B	right	9	2
63	raise_nose	Vertical displacement of nose tip	ENS	B	up	9	3
64	bend_nose	Horizontal displacement of nose tip	ENS	B	right	9	3

#	FAP name	FAP description	Units	Uni- or bidirectiona	Position motion	Group	FDP subgroup number
65	raise_l_ear	Vertical displacement of left ear	ENS	B	up	10	1
66	raise_r_ear	Vertical displacement of right ear	ENS	B	up	10	2
67	pull_l_ear	Horizontal displacement of left ear	ENS	B	left	10	3
68	pull_r_ear	Horizontal displacement of right ear	ENS	B	right	10	4

Figure 2.4.5. The list and descriptions of the 68 facial animation parameters (FAPs).

Description of markup language/markup declaration

After the header the animation FAP masks and FAP values are defined:

FAP mask: one 0 or 1 for each of the 68 FAPs (masked or unmasked) indicating whether it is used. The FAP mask precedes each FAP data line on a separate line.

Each frame is described as one line of integer *FAP values* starting with the frame number. The FAP mask in the header indicates which FAP groups are included. The 68 parameter values are defined in MPEG doc N1454. A zero value for any FAP indicates a neutral face.

Example:

<frame number> <viseme> <expression> <open_jaw>...<pull_r_ear>

Examples

This example has no corresponding sample from a resource. This example corresponds to the action of a different FAP at each frame. At frame 1, the facial model will exhibit the action of FAP 14 with the intensity 40; at frame 2, the FAP 15 with intensity -100; and so on.

#Header description

2.0 allfaps 30 5

#fap 14 thrust_jaw

0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

1 20

#fap 15 shift_jaw

0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0

2 -100

#fap 18 depress_chin

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0

3 25

#fap 50 head_roll

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0

4 17450

#fap 62 stretch_r_nose

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 1 0 0 0 0 0 0

5 100

Description of coding procedure, if any

Does not apply.

Creation notes, i.e. who wrote the coding scheme, when,and in which context?

The Moving Picture Experts Group (MPEG) is a working group of ISO/IEC in charge of the development of international standards for compression, decompression, processing, and coded representation of moving pictures, audio and their combination.

Contact person:
Dr. Leonardo Chiariglione
Multimedia Technology and Services
CSELT – via G. Reiss Romoli, 274
10148 Torino (Italy)
tel: +39 911 228 6120 / 6116 / 5111
fax: +39 011 228 6299 / 6190 /5520
email: leonardo.chiariglione@cselt.it
http://www.cselt.it/leonardo

2.4.5 Usage

Origin of the coding scheme and reasons for creating it

MPEG-4 is an object-based multimedia compression standard, which allows for encoding of different audio-visual objects (AVO). The MPEG-4 SNHC group proposes an architecture for the efficient representation and coding of synthetically and naturally generated audio-visual information. MPEG-4 foresees that talking heads will serve an important role in future customer service applications. For example, a customized agent model can be defined for games or web-based customer service applications. To this effect, MPEG-4 enables integration of face animation with multimedia communications and presentations and allows face animation over low bit rate communication channels, for point-to-point as well as multi-point connections with low-delay. MPEG-4 also has derived a standard for facial animation coding.

How many people have used the coding scheme and for what purposes?

Many groups.

How many dialogues/interactions have been annotated using the coding scheme?

Many.

Has the coding scheme been evaluated?

Model calibration methods have been elaborated by MPEG-4. Once a model has been calibrated it should be able to render any FAPs file in a seemingly manner as it would in any other calibrated models. That is, the same animation file should give the same result (i.e. the same facial expression on each frame) on all calibrated models.

Is the coding scheme language dependent or language independent?

The coding scheme is language independent.

Is there tools support for using the coding scheme or API for editing/parsing coded descriptions?

No information available.

2.4.6 Accessibility

How does one get access to the coding scheme?

The coding scheme is described in the ISO/IEC MPEG-4 Part 2 (Visual) document that can be found on the web pages dedicated to MPEG-4.

Is the coding scheme available for free or how much does it cost?

The coding scheme is publicly available for free.

2.4.7 Conclusion

How well described is the coding scheme?

Very well described.

How general and useful is the coding scheme?

The coding scheme is general enough to encode facial expression of emotions or lip movements.

2.5 ToonFace

2.5.1 Description header

Main actor

UROME: Catherine Pelachaud (cath@dis.uniroma1.it) and Isabella Poggi (poggi@uniroma3.it)

Verifying actor

NISLab: Malene Wegener Knudsen (mwk@nis.sdu.dk), Laila Dybkjær (laila@nis.sdu.dk) and Niels Ole Bernsen (nob@nis.sdu.dk)

Date of last modification of the description

June 21st, 2001.

2.5.2 Reference

Web site

A description of the ToonFace animation framework: http://xenia.media.mit.edu/~kris/gandalf.html

Short description

Toonface allows one to create and animate in short time interactive cartoon faces. It codes facial expression with limited detail.

An illustrative example of coding

Not available.

References to additional information on the coding scheme

Thórisson, K. R. ToonFace: A System for Creating and Animating Cartoon Faces. Learning & Common Sense Section Technical Report 1-96. 1996. Can be downloaded from : http://xenia.media.mit.edu/~kris/ftp/toonface.pdf

An extension of ToonFace, CharToon has been developed at CWI under the supervision of P.J.W ten Hagen as part of European project Facial Analysis and Synthesis of Expressions (FASE), which started in 1997 and ended in 2000.

Drs H. Noot and Ms. Dr. Zs.M. Ruttkay: CharToon 2.0 manual. INS-R0004, ISSN 1386-3681. 2000 Can be downloaded from: http://www.cwi.nl/ftp/CWIreports/INS/INS-R0004.ps.Z

For more details see:

The Facial Analysis and Synthesis of Expression homepage: http://www.cwi.nl/projects/FASE/

The Facial Analysis and Synthesis of Expression Chartoon homepage: http://www.cwi.nl/projects/FASE/CharToon/

2.5.3 Coverage

Which types of raw data are referenced?

2D facial model.

Which modalities is the coding scheme meant to code?

Facial expression.

Which annotation level(s) does the coding scheme cover?

Facial expression and creation of model.

Which coding tasks has the coding scheme been used for?

To create and animate cartoon faces.

2.5.4 Detailed description of coding scheme

Which header file information is included (meta-data)?

No header file.

Coding purpose of the coding scheme?

ToonFace is a coding scheme to code facial expression with limited detail (especially with much less detail than FACS).

List and description of phenomena, which can be annotated by the scheme

A face is divided into seven main features: Two eye brows, two eyes, two pupils and a mouth. The eyebrows have three control points each, the eyes and mouth four and pupils one each.

Description of markup language/markup declaration

The following is a complete list of all one-dimensional motors that can be manipulated in a face [control point number in brackets]:
Brl = brow/right/lateral [3]; Brc = brow/right/central [2]; Brm = brow/right/medial [1]
Bll = brow/left/lateral [6]; Blc = brow/left/central [5]; Blm = brow/left/medial [4]
Eru = eye/right/upper [7]; Erl = eye/right/lower [9]
Elu = eye/left/upper [8]; Ell = eye/left/lower [10]
Plh = pupil/right/horizontal [15]; Plv = pupil/left/vertical [15]
Prh = pupil/right/horiz [16-h]; Prv = pupil/right/vertical [16-v]
Mlh = mouth/left/horizontal [14-h]; Mlv = mouth/left/vertical [14-v]
Mrh = mouth/right/horizontal [13-h]; Mrv = mouth/right/vertical [13-v]
Mb = mouth/bottom [12]
Hh = head/horizontal [17-h]; Hv = head/vertical [17-v]

Examples

Figure 2.5.1. Examples of faces created with ToonFace.

Description of coding procedure, if any

The graphics program allows the creation of faces and/or facial expressions interactively.

Creation notes, i.e. who wrote the coding scheme, when,and in which context?

ToonFace was created by Kristinn R. Thórisson at the M.I.T.Media Laboratory:
http://www.media.mit.edu/

2.5.5 Usage

Origin of the coding scheme and reasons for creating it

The origin of the coding scheme is the easy creation of a 2D synthetic agent in the Ymir's application.

How many people have used the coding scheme and for what purposes?

The ToonFace animation mechanism has been duplicated in Java by researchers at the Fuji-Xerox Palo Alto Research Laboratories. The CWI - Stichting Mathematical Institute in Amsterdam is using ToonFace principles to build a suit of new animation software.

How many dialogues/interactions have been annotated using the coding scheme?

Many.

Has the coding scheme been evaluated?

It is difficult to talk about evaluation here. Animators have reported how easy and fast it is to create cartoon faces and animation.

Is the coding scheme language dependent (which language(s)) or language independent?

ToonFace is language independent.

Is there tools support for using the coding scheme or API for editing/parsing coded descriptions? In which language ?

ToonFace comes with a program that allows the creation of synthetic faces.

2.5.6 Accessibility

How does one get access to the coding scheme?

The coding scheme is described in papers.

Kristinn R. Thórisson, ToonFace: A System for Creating and Animating Interactive Cartoon Faces, Learning and Common Sense Section Technical Report 96-01, April 1996. Can be downloaded from : http://xenia.media.mit.edu/~kris/ftp/toonface.pdf

An extension of ToonFace, CharToon has been developed at CWI under the supervision of P.J.W ten Hagen as part of an European project Facial Analysis and Synthesis of Expressions (FASE), which started in 1997 and ended in 2000.

Drs H. Noot and Ms. Dr. Zs.M. Ruttkay: CharToon 2.0 manual. INS-R0004, ISSN 1386-3681. 2000 Can be downloaded from: http://www.cwi.nl/ftp/CWIreports/INS/INS-R0004.ps.Z

For more details see:

The Facial Analysis and Synthesis of Expression homepage: http://www.cwi.nl/projects/FASE/

The Facial Analysis and Synthesis of Expression Chartoon homepage: http://www.cwi.nl/projects/FASE/CharToon/

Is the coding scheme available for free or how much does it cost?

The coding scheme is available for free.

2.5.7 Conclusion

How well described is the coding scheme?

The scheme is well described and is an interesting approach when one desires to create cartoon like facial animation.

How general and useful is the coding scheme?

The coding scheme is adapted for 2D faces not really for 3D faces.

3 Lesser Known/Used Facial Coding Schemes

3.1 BABYFACS – Facial Action Coding System for Baby Faces

3.1.1 Description header

Main actor

HCRC: Jean Carletta (jeanc@cogsci.ed.ac.uk)

Verifying actor

NISLab: Malene Wegener Knudsen (mwk@nis.sdu.dk), Laila Dybkjær (laila@nis.sdu.dk) and Niels Ole Bernsen (nob@nis.sdu.dk)

Date of last modification of the description

June 21st, 2001.

3.1.2 Reference

References to additional information on the coding scheme

Paper by Harriet Oster et. al.: The effect of craniofacial anomalies on infant facial expression and maternal adjustment. Can be downloaded from: http://www.sbg.ac.at/psy/events/facs/apo_oste.htm

3.1.3 Short description

BABYFACS is a coding scheme for facial expressions, which is based on FACS (see earlier entry in this report) but tailored to infants. This tailoring is useful, among other reasons, because infants have a less expressive range for negative affect than adults. BABYFACS was devised by Harriet Oster of NYU and has been used for a number of research projects in this area.

It has not been possible to find any further information, since it has not been possible to establish contact to Harriet Oster or find any other references than the above.

3.2 General description of coding schemes for hand annotation of mouth and lip movements and speech

3.2.1 Description header

Main actor

HCRC: Jean Carletta (jeanc@cogsci.ed.ac.uk)

Verifying actor

NISLab: Malene Wegener Knudsen (mwk@nis.sdu.dk), Laila Dybkjær (laila@nis.sdu.dk) and Niels Ole Bernsen (nob@nis.sdu.dk)

Date of last modification of the description

June 21st, 2001.

3.2.2 Description

There are no coding systems for hand-annotating mouth or lip movements during speech, probably because this would be a very laborious process. During the 1970s there was some work on measuring mouth movements using direct instrumentation in the form of electrodes, but for obvious reasons this had to be limited to individuals with damaged nerve endings. Currently, rather than hand-annotating such movements, it would be more efficient to analyse them automatically from video. It is possible to extract features from video signals such as the corners of the mouth and the nose quite reliably. For instance, one very active field of research is that of using features extracted from a video signal to improve speech recognition on the corresponding audio (see http://www.clsp.jhu.edu/ws2000/groups/av_speech/).

4 Gesture Coding Schemes

This introduction covers chapters 4 and 5, both of which have their primary focus on gesture coding schemes. Chapter 4 covers gesture coding schemes while Chapter 5 describes lesser known gesture coding schemes for which we have found little information.

The descriptions cover existing coding schemes which have been applied to the coding of pure gesture (hand, arm, other), pure hand manipulation of objects, pure body movement, and any of the preceding as accompanied by speech.

4.1 DIME: National Autonomous Univ. of Mexico, Multimodal extension of DAMSL

4.1.1 Description header

Main actor

LIMSI: Jean-Claude MARTIN (martin@limsi.fr)

Verifying actor

UNAM: Luis Alberto PINEDA CORTÉS (luis@leibniz.iimas.unam.mx): contacted on 9 august 2001

Date of last modification of the description

9 August 2001

4.1.2 Reference

Web site

http://cic2.iimas.unam.mx/multimod/dime/
Publications, samples of video and transcription, slides explaining the coding scheme.

Short description

This coding scheme is an extension of the DAMSL dialogue act markup in several layers scheme (Allen and Core 1997). The extension concerns

1) a more structured conversational unit than the utterance : the contribution,
2) deictic gestures and information conveyed through external representation such as paper or graphical screen.

One illustrative example of coding

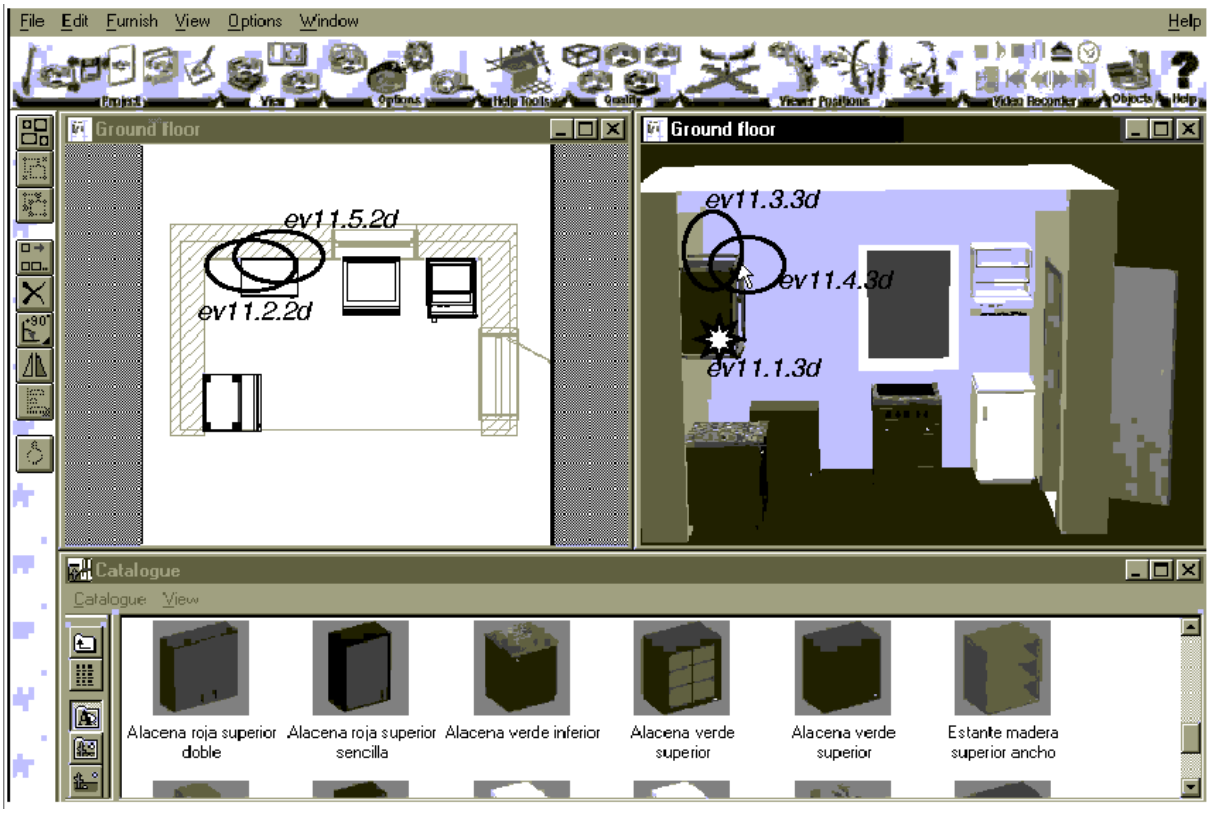


Figure 4.1.1. A sample video frame (subjects spoken utterances are also recorded).

	Information level/ Forward function/ Backward function	Interventions	Translation	Refer-ents
u: utt11	Task / Influence-on-listener = Action-directive / ∅	Puedes poner este <sil> **este estante (ev11.1.3d)** lo puedes poner eh también en **esta pared (ev11.2.2d)** pero o se[a] más o menos a **esta altura (ev11.3.3d)** en **la pared de este lado (ev11.4.3d)** e-en **la pared del fondo (ev11.5.2d)**	Can you put this <sil> **this shelf (ev11.1.3d)** can you put it erm also on **this wall (ev11.2.2d)** but I mean more or less at **this high (ev11.3.3d)** on **the wall of this side (ev11.4.3d)** on **the back-wall. (ev11.5.2d)**	furn.2 wall.3 region.1 wall.3 wall.3
s: utt12	Task / ∅ / understanding = Ack(utt11), agree-ment=Accept?	Ok	Ok	
s: utt13	Task /	¿quieres que ponga	Do you want me that to put	
	Info-request=yes, influence-on-speaker = Offer / ∅	**este estante (ev13.1.3d)** en **esta esquina (ev13.2.2d)**?	**this shelf (ev13.1.3d)** in **this corner (ev13.2.2d)**?	furn.2 region.2
u: utt14	Task / ∅ / Answer(utt13), Agreement=Accept	Sí	Yes	
s: utt15	Task / influence-on-speaker=Commit / understanding = Ack(utt14), agreement=Accept	Ok	Ok	
s: utt16	Task / ∅ / information-relations = Action-accomplishment(utt11)	*<conjunto de acciones para colocar el estante>*	<sequence of actions to put the shelf>	move(furn.2, region.2)
s: utt17	Task / Info-request=yes / ∅	¿ Así está bien ?	Is that all right ?	
u: utt18	Task / Answer(utt17) / agreement=Accept	Sí, así esta bien	Yes, that's fine	

Figure 4.1.2. Annotation of the sample in Figure 4.1.1 using the DAMSL coding scheme extended by the authors to include multimodal interaction and contributions.

Introduction	Referents
Initial Context	wall.1 = left wall wall.2 = right wall wall.3 = back wall window.1 = back window ... furn.1 = stove furn.2 = shelf furn.3 = refrigerator ...
ev11.3.3d	region.1 = medium height wall.1
ev13.2.2d	region.2 = corner wall.1 & wall.3

Figure 4.1.3. Annotation of the graphical context (objects and mouse events).

References to additional information on the coding scheme

Luis Villaseñor, Antonio Massé, Luis A. Pineda (2000) A Multimodal Dialogue Contribution Coding Scheme http://www.mpi.nl/world/ISLE/documents/papers/villasenor_paper.pdf

Towards a Multimodal Dialogue Coding Scheme by Luis Villaseñor, Antonio Massé and Luis Pineda. Presented at CICLing-2000 Conference on Intelligent text processing and Computational Linguistics, February 13 to 19, 2000. México City, México.

(talk and paper available at http://cic2.iimas.unam.mx/multimod/dime/index.html)

Allen, J. & Core, M. (1997) Draft of DAMSL: Dialog Act Markup in Several Layers. Pp 32.

http://www.cs.rochester.edu/research/cisd/resources/damsl/

4.1.3 Coverage

Which types of raw data are referenced?

Utterances are numbered and in the DRAT tool, utterances are linked to audio and video files (see below).

Which modalities is the coding scheme meant to code?

Speech, graphical actions of the user via the mouse.

Which annotation level(s) does the coding scheme cover?

Annotation levels of DAMSL:

A dialogue is divided into units called turns

A turn is composed by one or several utterance units

The notion of utterance is based on an analysis of the intentions of the speaker

Each utterance is annotated to indicate its contribution to the dialogue in 4 orthogonal aspects:Communicative Status. Records whether the utterance is intelligible and whether it was successfully completed.Information Level. A characterization of the semantic content of the utterance.The Forward Looking Function. How the current utterance constrains

the future believes and actions of the participants, and affects the discourse.The Backward Looking Function. How the current utterance relates to the previous discourse.

The DAMSL scheme has been extended by the authors in two dimensions:

1) labels for graphical actions have been added (mouse events + references to graphical objects)
2) a level of contribution sequences

Thus, as can be seen in the example above, the multimodal annotation includes the following information: contribution task level, contribution description level, utterance, DAMSL label, interventions (speech transcription + mouse events), referents.

Which coding tasks has the coding scheme been used for?

Code multimodal behavior observed in simulated Wizard of Oz sessions

4.1.4 Detailed description of coding scheme

Which header file information is included (meta-data)?

Example taken from the web:
Dialogue: Mayra-tarea2
Number of utterances files: 264
Length of dialogue: 1182.329727
Estimated number of turns: 170

Coding purpose of the coding scheme?

Code multimodal behavior observed in simulated sessions in order to specify a multimodal information system.

List and description of phenomena, which can be annotated by the scheme

Sequences of contributions, multimodal deictic expressions.

Description of markup language/markup declaration

There is no markup language. Annotations are done in text tables (cf. example above).

Examples

Cf. above (on the web site only a Spanish-only example in a slide is available as well as transcriptions of speech only).

Description of coding procedure, if any

The extension to DAMSL consists of a table, which has entries for utterance identifiers, annotation labels, interventions and their referents. Deictic expressions within interventions are highlighted and the corresponding pointing gestures are labeled as demonstrative events. These are identified as "evX.Y.Dd", where X is the utterance in which the demonstration is made, Y is the number of demonstration in the utterance, and D indicates whether the gesture has been made on the 2-D or 3-D window of the multimodal interface. The table has also a column in which the referent for every spatial demonstration is explicitly stated. Intuitively, the referent of a highlighted term is an individual that can be identified in the region pointed out in the graphical domain, taking into account the conceptual constraints imposed by the linguistic term.

Creation notes, i.e. who wrote the coding scheme (contact details), when, and in which context?

Contact: Luis Alberto PINEDA CORTÉS (luis@leibniz.iimas.unam.mx)
Partners of the DIME project (Developing Domain Independent Dialogue Models): A Prototype in Spanish in the Domain of Geometric Design" is a joint project between the Department of Computer Science at IIMAS, UNAM Mexico and the Department of Computer Science at the University of Rochester, USA.

4.1.5 Usage

Origin of the coding scheme and reasons for creating it

The existing schemes such as DAMSL are not enough to annotate interactive multimodal behaviour.

How many people have used the coding scheme and for what purposes?

Only the authors

How many dialogues/interactions have been annotated using the coding scheme?

30 dialogues in the kitchen design domain (15 subjects with 2 tasks).

Has the coding scheme been evaluated?

No

Is the coding scheme language dependent (which language(s)) or language independent?

Language independent (labels are in English but can be translated)

Is there tools support for using the coding scheme or API for editing/parsing coded descriptions? In which language?

Figures 4.1.4 and 4.1.5 show the DRAT tool graphical interface. DRAT is used to label the dialogue's speech acts and to mark discourse referents. By using this tool it is possible to hear the audio, read the orthographic transcription and see the segment of video of each utterance in such a way that it is possible to analyse the relation between deictic and verbal events, and to establish what is the contribution of the utterance in the dialogue.
The tool consists of three windows:
to annotate the speech acts of an utterance
to label the referential expressions of the utterance
to play the video segment associated with the utterance

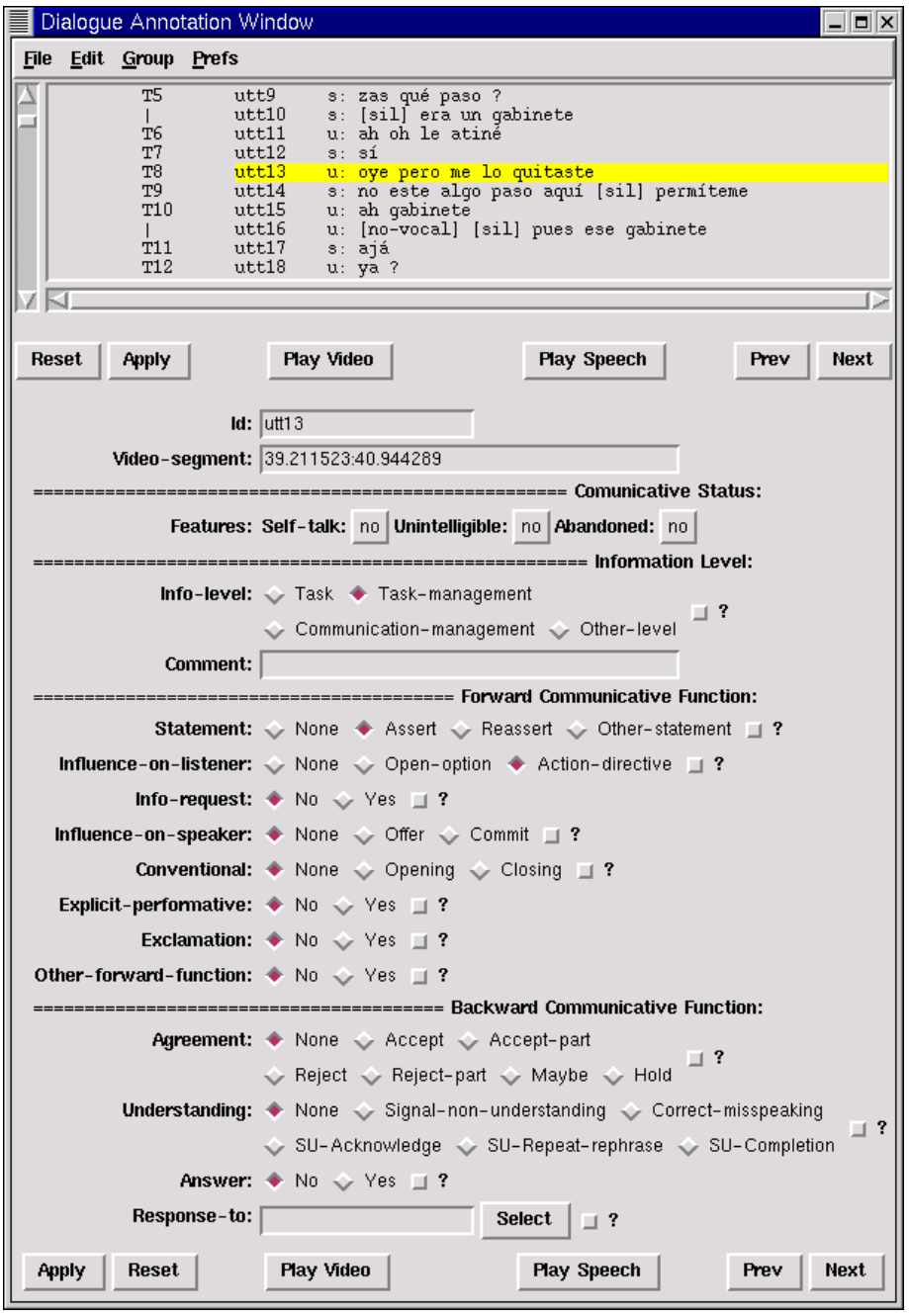


Figure 4.1.4. DRAT tool graphical interface used to label the speech acts in recorded dialogues.

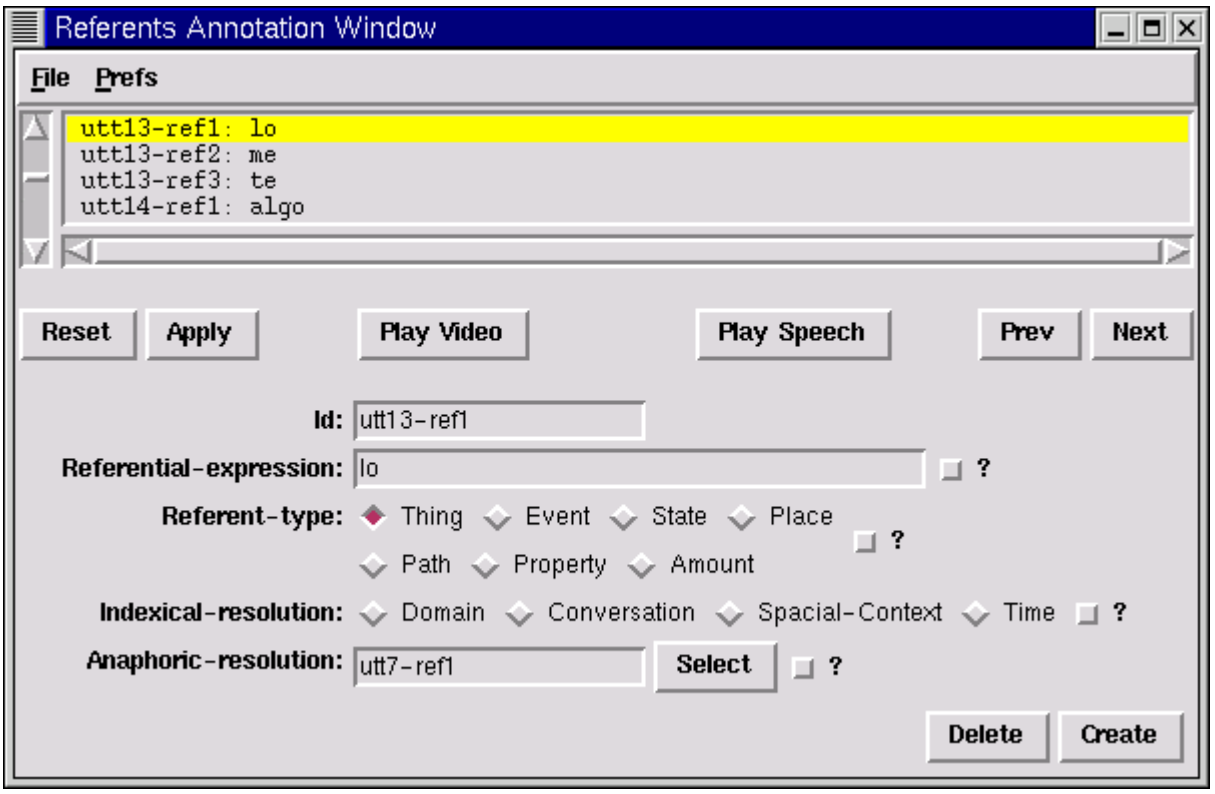


Figure 4.1.5. DRAT tool graphical interface used to label the discourse referents in recorded dialogues.

4.1.6 Accessibility

How does one get access to the coding scheme?

Luis Alberto PINEDA CORTÉS (luis@leibniz.iimas.unam.mx)

Is the coding scheme available for free or how much does it cost?

No information available

4.1.7 Conclusion

How well described is the coding scheme?

Well although still informal yet (no markup language definition).

How general and useful is the coding scheme?

Useful for mark-up of combination of spoken dialogs and 2D gestures on graphics.

4.2 HamNoSys - Hamburg Notation System for Sign Languages

4.2.1 Description header

Main actor

IMS: Ulrich Heid (heid@IMS.Uni-Stuttgart.DE)
We would like to thank Thomas Hanke, Hamburg, for his help in reviewing a draft of this summary sheet and for his detailed comments.

Verifying actor

LIMSI: Jean-Claude MARTIN (martin@limsi.fr)

Date of last modification of the description

11 October 2001

4.2.2 Reference

Web site

http://www.sign-lang.uni-hamburg.de/Projects/HamNoSys.html

Short description

HamNoSys was created to capture different sign languages: it thus does not make reference to specific national finger spelling systems. As HamNoSys is exclusively meant as a transcription system for sign languages, its details are not directly relevant for ISLE. However, a few general aspects of the design of the transcription system are of potential interest for work in ISLE. This is the reason for mentioning HamNoSys briefly in this report.

One illustrative example of coding

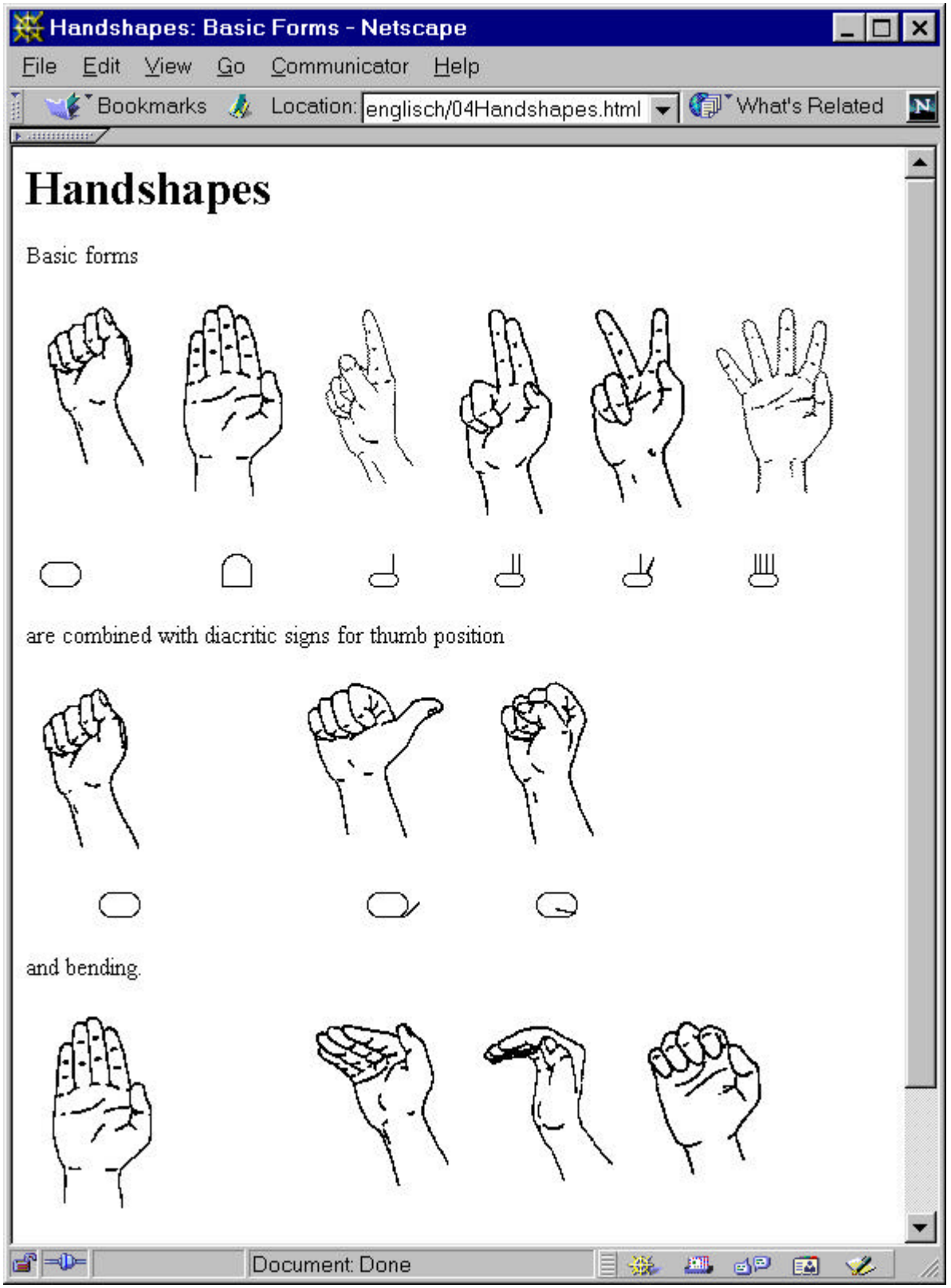


Figure 4.2.1. Basic hand shapes and their symbolic representation in HamNoSys.

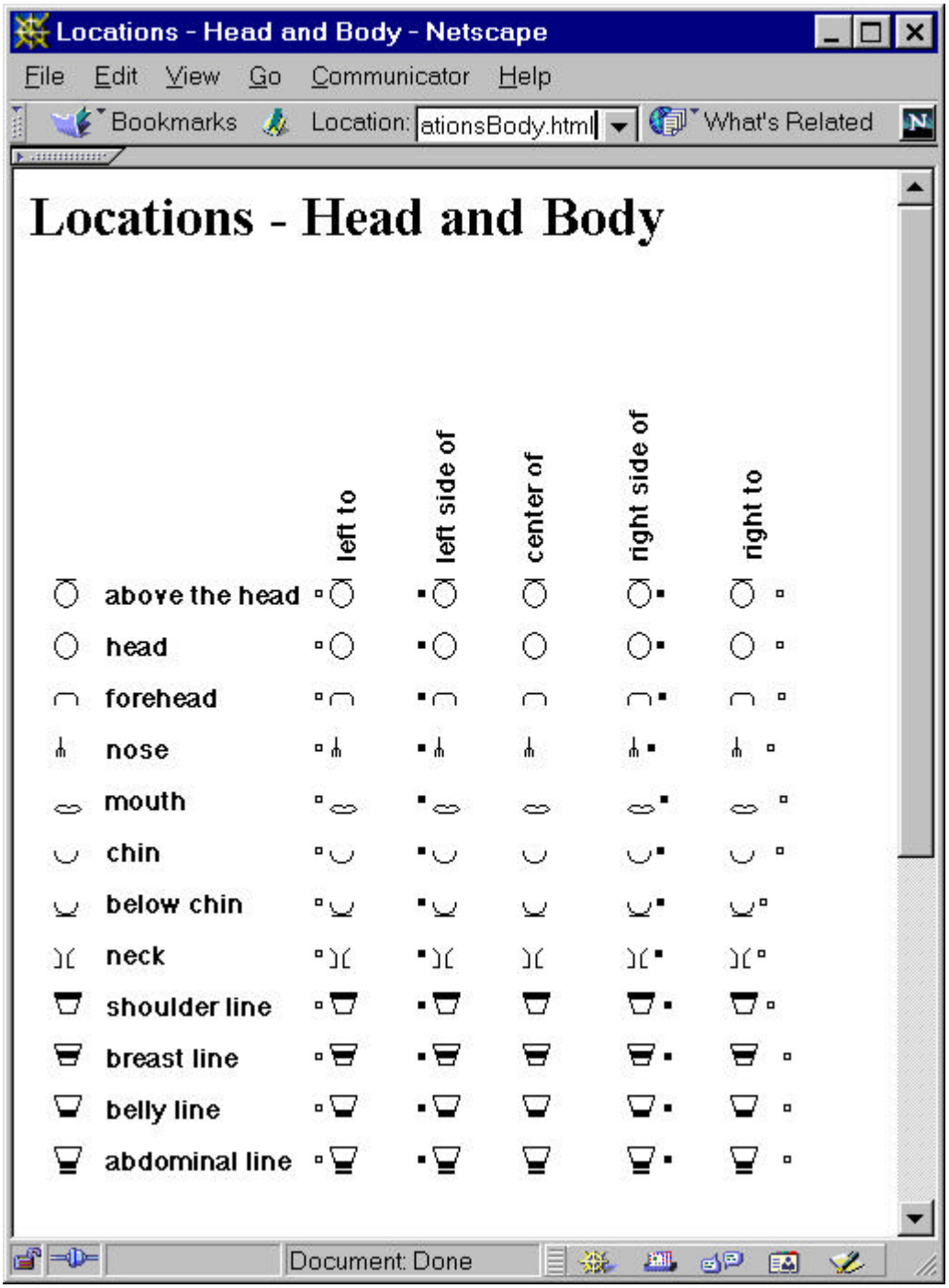


Figure 4.2.2. Table of the coding of the position of the hand with respect to head and body.

References to additional information on the coding scheme

The web site includes a detailed description of the coding scheme in German and English (http://www.sign-lang.uni-hamburg.de/bibweb/lidat.acgi?ID=8439).

For those who know version 2.0 of HamNoSys, there is a summary of new features of vs. 3 on the web site

http://www.sign-lang.uni-hamburg.de/Projekte/HamNoSys/HamNoSysErklaerungen/englisch/Contents.html.

A new description, based on version 4 can be found at the following URL:

http://www.sign-lang.uni-hamburg.de/Projekte/HamNoSys/HNS4.0/englisch/HNS4.pdf

There exist lexicons based on HamNoSys, asa well as more descriptive material (see bibliography on URL: http://www.sign-lang.uni-hamburg.de/tlex/Intro/Frame5.htm).

The lexicons can be downloaded from the following URLs, at no cost:

http://www.sign-lang.uni-hamburg.de/plex/
http://www.sign-lang.uni-hamburg.de/hlex/
http://www.sign-lang.uni-hamburg.de/tlex/
A lexicon of computing terms ('clex') should become available later in 2001.

4.2.3 Coverage

Which types of raw data are referenced?

The information accessible to us does not say much about the direct annotation of video. On the HamNoSys website there is a small QuickTime video with sign language statements on it. The syncWRITER tool provides the possibility to synchronize and annotate different "tracks", e.g. different levels of annotation, including a video track and a comment track. For details, see the ISLE deliverable D-11.1.

Which modalities is the coding scheme meant to code?

HamNoSys is designed as a transcription scheme for (different) sign languages. This implies that it focused originally in very much detail on the hands of the speaker it mainly keeps track of hand movements; additional means of expression used in sign languages (gesture at large, gaze, head movement, head position, etc.) are the object of extensions. Non-manual gesture is covered in projects related with the development of HamNoSys; cf. Transcription and glossing of sign language texts: Examples from AUSLAN (Australian Sign Language by Trevor Johnston, see URL

http://www.sign-lang.uni-hamburg.de/bibweb/LiDat.acgi?ID=29746

HamNoSys can specify, instead ot the hands, another body part as means of articulation, but the inventory of descriptors available was originally conceived for hand movements.

The ViSiCAST project (http://www.visicast.co.uk) has developed its own coding guidelines for non-manual gesture, but these are unfortunately so far internal to the project (thanks to Thomas Hanke, Hamburg, for pointing this out) and not part of HamNoSys.

Which annotation level(s) does the coding scheme cover?

HamNoSys covers all parts of hand shapes and hand movements:

Is the gesture one-handed or two-handed? Which is the main hand?

Initial configuration; subsequent actions;

Hand form

Hand position

Location of the hand wrt. the body.

Which coding tasks has the coding scheme been used for?

The HamNoSys scheme has been used for the transcription of different sign languages and for the creation of lexicons of sign languages. Note that HamNoSys will be used in the ongoing ViSiCAST project (http://www.visicast.co.uk), the purpose of which is to develop tools for access to public services, web sites, commercial use web devices, interactive television etc. for the deaf. This will be done by converting speech and text into sign language; HamNoSys will for that purpose be made "XML-compatible" and used as a representation medium for the sign language to be generated.

4.2.4 Detailed description of coding scheme

Which header file information is included (meta-data)?

No information available

Coding purpose of the coding scheme?

Rendering of sign languages

List and description of phenomena, which can be annotated by the scheme

Single-handed vs. both handed;
Initial configurations:

- in terms of non-manual components;
- in terms of hand shapes:
  - basic forms
  - position of thumb
  - inclination
- in terms of hand position:
  - pointing direction;
  - orientation of the hand;
  - both described with respect to the person gesturing;
- in terms of the hand location:
  - with respect to the body part at the level of which the hand is positioned;
    - above the head
    - head, nose, mouth, ...
    - shoulder
    - etc.
  - with respect to the distance from the body of the person gesturing.
- actions (i.e. combinations of hand movements, at the same time or immediately after each other);
  - line shape, bow shape, zigzag shape, etc.
  - change of hand position, etc.

In total, HamNoSys comprises about 200 symbols.

Description of markup language/markup declaration

The markup language includes about 200 symbols, most of them iconic, to remind the user of the shape, position etc. of the hand in the real gesture. The icons have been designed to be symbols on a special piece of software for Macintosh PCs; in addition there exist symbols for diacritics. So far, HamNoSys is not re-representable in XML; making it XML-tractable is one of the objectives of the ViSiCAST project (see above, under *coding tasks* and URL http://www.visicast.co.uk). Details of the markup declaration can be found in the slides on http://www.sign-lang.uni-hamburg.de/Projekte/HamNoSys/HamNoSysErklaerungen/englisch/Contents.html. A project-internal version of HamNoSys in XML exists in the ViSiCAST project. It would be interesting to follow up this development when it will become publicly available. This XML version also comprises the coding of non-manual gesture mentioned above.

Examples

An example of a text in English, along with the HamNoSys annotation for a right-handed signer can be found on the following URL: http://www.signwriting.org/forums/linguistics/ling007.html

Description of coding procedure, if any

Since HamNoSys is meant to be a "meta-transcription scheme", i.e. a representational system for a set of sign languages, and since each sign language has its alphabet and lexicon(s), there is no specific coding procedure needed for HamNoSys.

Creation notes, i.e. who wrote the coding scheme (contact details), when, and in which context?

Cf. Prillwitz, Siegmund et al: HamNoSys. Version 2.0; Hamburg Notation System for Sign Languages. An introductionary guide. (International Studies on Sign Language and Communication of the Deaf; 5) Hamburg: Signum 1989 - 46 p.
URL http://www.sign-lang.uni-hamburg.de/Projects/HamNoSys.html

4.2.5 Usage

Origin of the coding scheme and reasons for creating it

HamNoSys was created some 15 years ago and is now at its versions 3 (operational) and 4 (pre-release). It was created to capture different sign languages: it thus does not make reference to specific national finger spelling systems.

How many people have used the coding scheme and for what purposes?

We do not have figures about the use of HamNoSys, but Miller says
(cf. URL http://www.sign-lang.uni-hamburg.de/intersign/Workshop2/Miller/Miller.html)
it seems to have been used massively over the last 15 years.

How many dialogues/interactions have been annotated using the coding scheme?

No figures available. Details about the lexicons may be found on the SIGNUM Lexicon CD-ROM: FachgebaerdenLexikon Computer, Hamburg, Signum, 1993.

Has the coding scheme been evaluated?

There are no publications about evaluation of the HamNoSys coding scheme, and no intercoder consistency tests seem so far to have been made.

Is the coding scheme language dependent (which language(s)) or language independent?

Information not available

Is there tools support for using the coding scheme or API for editing/parsing coded descriptions? In which language?

The syncWRITER transcription tool has been developed in close connection with HamNoSys. It has been described in detail in ISLE deliverable D-11.1.
URL: http://www.sign-lang.uni-hamburg.de/software/syncwriter/info.english.html
For lexicon building work, there is another tool as well: GlossLexer,
cf. URL: http://www.sign-lang.uni-hamburg.de/Intersign/Workshop1/HankeKonradSchwarz

4.2.6 Accessibility

How does one get access to the coding scheme?

Information not available

Is the coding scheme available for free or how much does it cost?

Information not available

4.2.7 Conclusion

How well described is the coding scheme?

The HamNoSys system is very well described. Much material is on the web site; more is in preparation. Even more can be purchased on CD-ROM.

How general and useful is the coding scheme?

HamNoSys deals with sign language, not primarily with gesture in oral language although is has also been used for such tasks several times.
Given its widespread use and the tool support available (cf. D-11.1, section 12), it is a major resource for sign languages and some of the notions used in the HamNoSys classifications of hand shapes, positions and movements may be directly relevant for ISLE.

4.3 HIAT -- Halbinterpretative Arbeitstranskriptionen

4.3.1 Description header

Main actor

IMS: Ulrich Heid (heid@IMS.Uni-Stuttgart.DE)

Verifying actor

We would like to thank Dr. Gabriele Graefen, Munich, as well as Dr. Susanne Scheiter and Wolfgang Schneider, Dortmund, for providing us with many very helpful comments on an earlier version of this section, as well as with most recent updates.

Date of last modification of the description

Friday, 28 December 2001

4.3.2 Reference

Web site

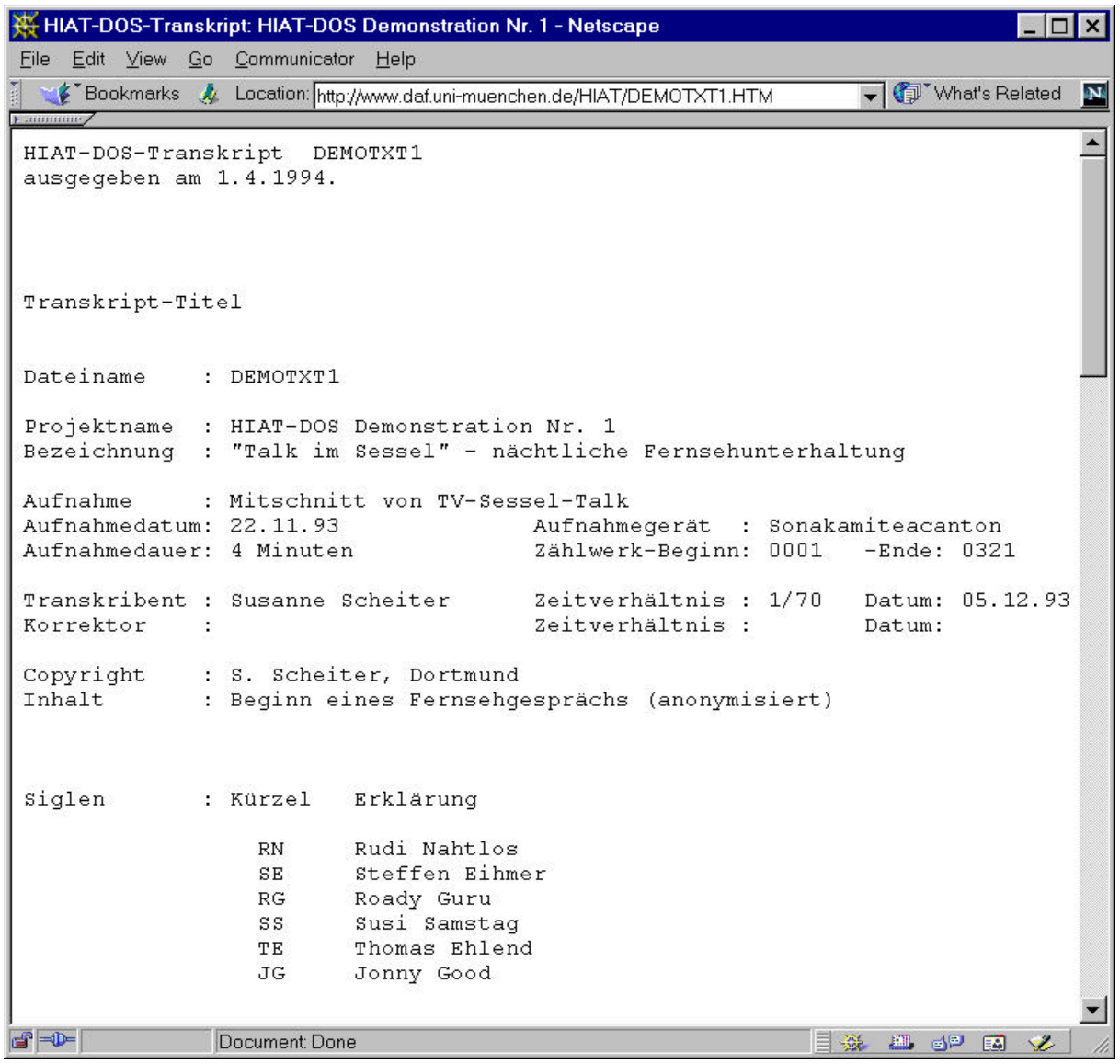
http://www.daf.uni-muenchen.de/HIAT/HIAT.HTM

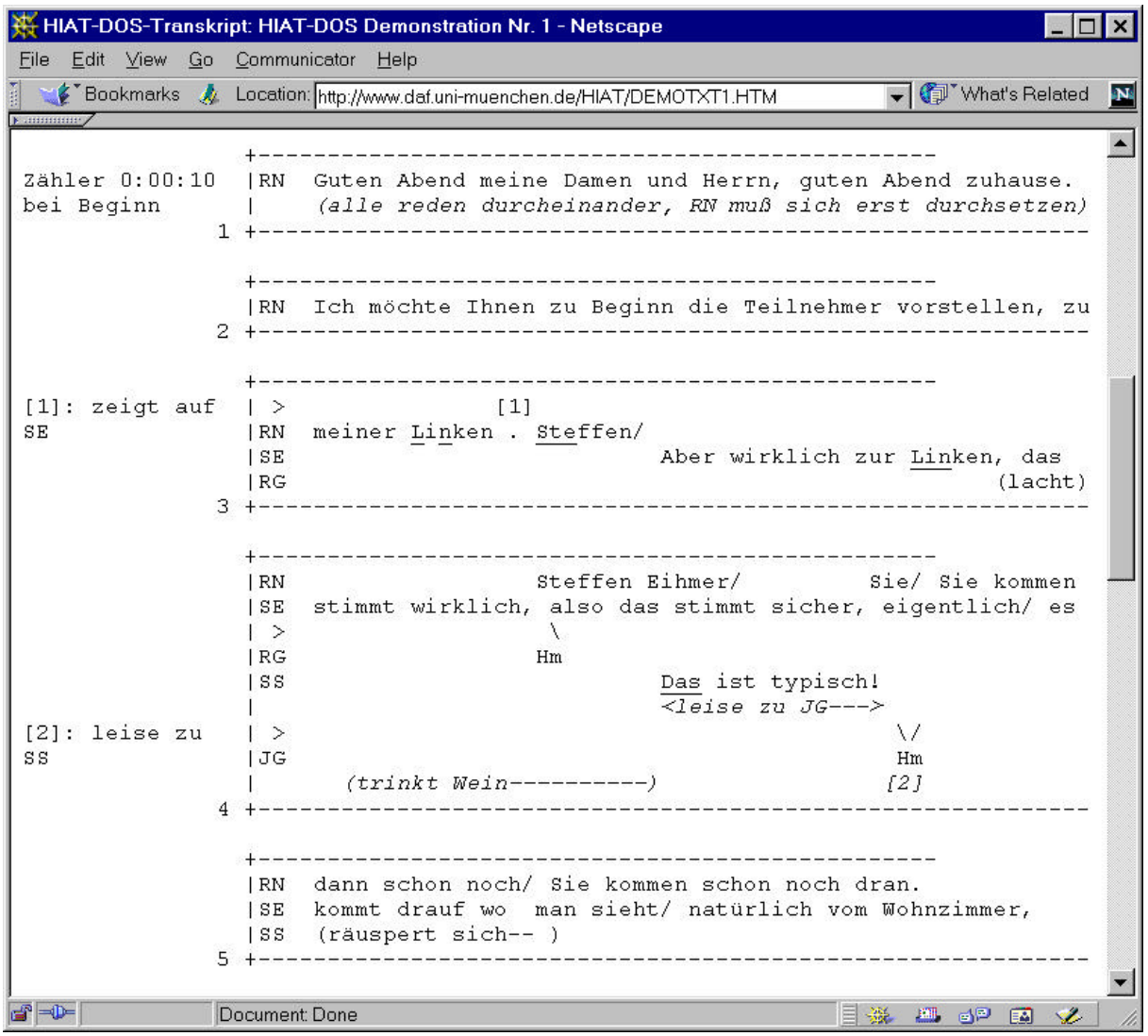
Short description

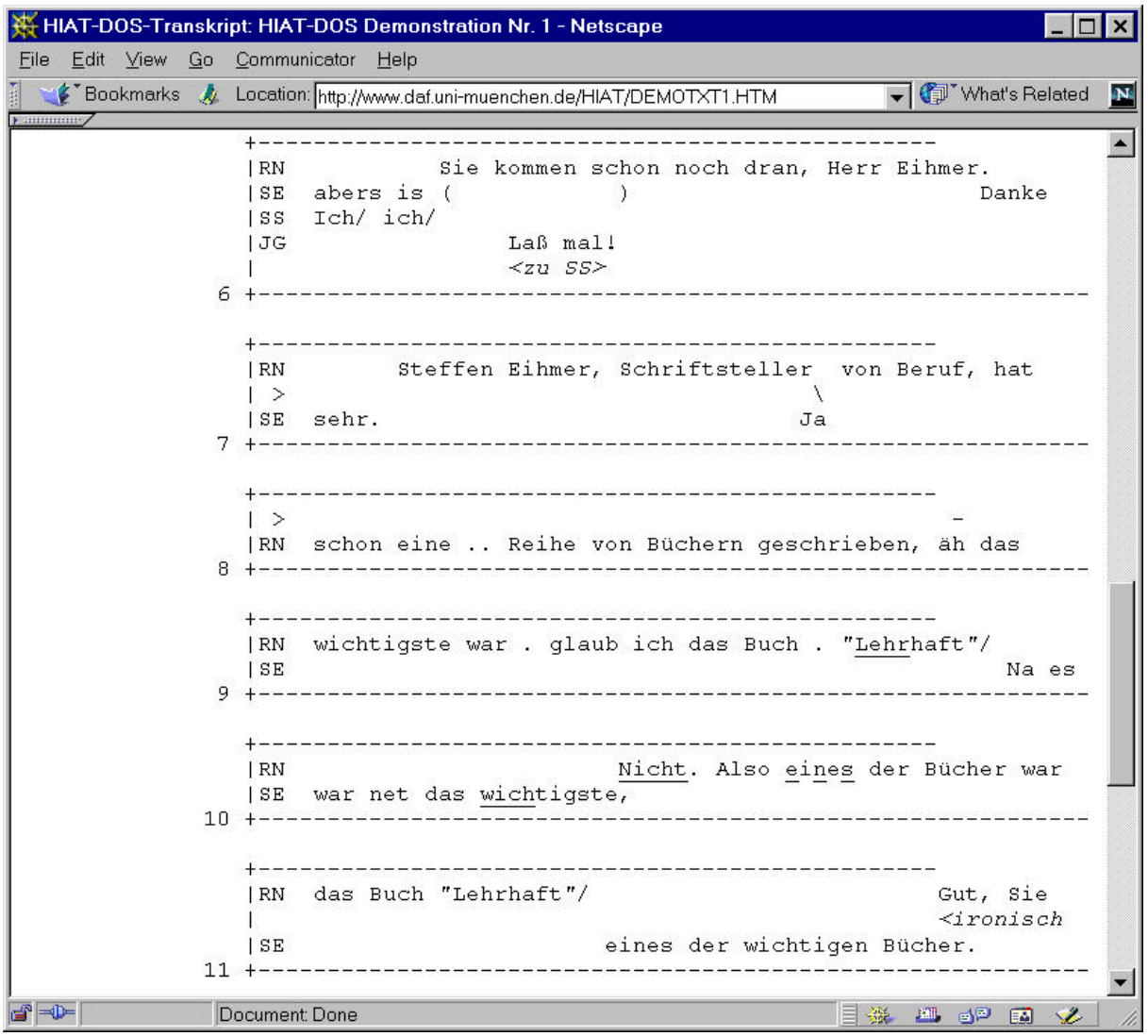
The purpose of the annotation scheme (and of the pertaining tool HIAT-DOS) is to describe and annotate parallel tracks of verbal and non-verbal (e.g. gesture) communication in a simple way.

One illustrative example of coding

Below is a sample transcript from a TV talk show, including meta data at the beginning (taken from URL: http://WWW.DaF.Uni-Muenchen.De/HIAT/DEMOTXT1.HTM)


HIAT-DOS-Transkript: HIAT-DOS Demonstration Nr. 1 - Netscape
File Edit View Go Communicator Help
Bookmarks Location: http://www.daf.uni-muenchen.de/HIAT/DEMOTXT1.HTM What's Related
HIAT-DOS-Transkript DEMOTXT1
ausgegeben am 1.4.1994.
Transkript-Titel
Dateiname : DEMOTXT1
Projektname : HIAT-DOS Demonstration Nr. 1
Bezeichnung : "Talk im Sessel" - nächtliche Fernsehunterhaltung
Aufnahme : Mitschnitt von TV-Sessel-Talk
Aufnahmedatum: 22.11.93 Aufnahmegerät : Sonakamiteacanton
Aufnahmedauer: 4 Minuten Zählwerk-Beginn: 0001 -Ende: 0321
Transkribent : Susanne Scheiter Zeitverhältnis : 1/70 Datum: 05.12.93
Korrektor : Zeitverhältnis : Datum:
Copyright : S. Scheiter, Dortmund
Inhalt : Beginn eines Fernsehgesprächs (anonymisiert)
Siglen : Kürzel Erklärung
RN Rudi Nahtlos
SE Steffen Eihmer
RG Roady Guru
SS Susi Samstag
TE Thomas Ehlend
JG Jonny Good
Document: Done


HIAT-DOS-Transkript: HIAT-DOS Demonstration Nr. 1 - Netscape
File Edit View Go Communicator Help
Bookmarks Location: http://www.daf.uni-muenchen.de/HIAT/DEMOTXT1.HTM
What's Related
Zähler 0:00:10 bei Beginn
RN Guten Abend meine Damen und Herrn, guten Abend zuhause.
(alle reden durcheinander, RN muß sich erst durchsetzen)
1
RN Ich möchte Ihnen zu Beginn die Teilnehmer vorstellen, zu
2
[1]: zeigt auf SE
RN meiner Linken . Steffen/
SE Aber wirklich zur Linken, das
RG (lacht)
3
RN Steffen Eihmer/ Sie/ Sie kommen
SE stimmt wirklich, also das stimmt sicher, eigentlich/ es
RG Hm
SS Das ist typisch!
<leise zu JG--->
[2]: leise zu SS
JG Hm
(trinkt Wein----------) [2]
4
RN dann schon noch/ Sie kommen schon noch dran.
SE kommt drauf wo man sieht/ natürlich vom Wohnzimmer,
SS (räuspert sich--)
5
Document: Done


HIAT-DOS-Transkript: HIAT-DOS Demonstration Nr. 1 - Netscape
File Edit View Go Communicator Help
Bookmarks Location: http://www.daf.uni-muenchen.de/HIAT/DEMOTXT1.HTM What's Related
|RN Sie kommen schon noch dran, Herr Eihmer.
|SE abers is ( ) Danke
|SS Ich/ ich/
|JG Laß mal!
| <zu SS>
6
|RN Steffen Eihmer, Schriftsteller von Beruf, hat
| > \
|SE sehr. Ja
7
| > -
|RN schon eine .. Reihe von Büchern geschrieben, äh das
8
|RN wichtigste war . glaub ich das Buch . "Lehrhaft"/
|SE Na es
9
|RN Nicht. Also eines der Bücher war
|SE war net das wichtigste,
10
|RN das Buch "Lehrhaft"/ Gut, Sie
| <ironisch
|SE eines der wichtigen Bücher.
11
Document: Done

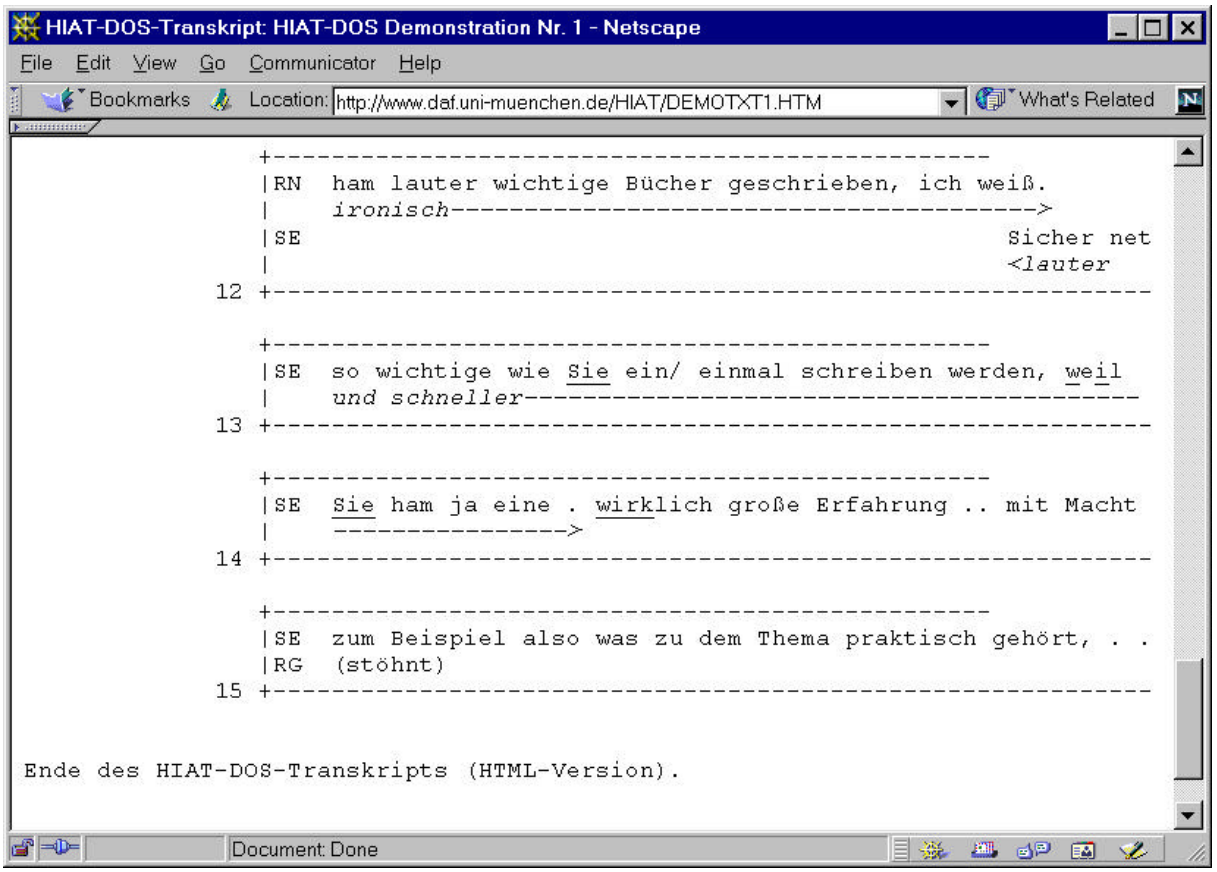


Figure 4.3.1. Sample transcript from a TV talk show, including meta-data at the beginning.

References to additional information on the coding scheme

Ehlich, Konrad; Rehbein, Jochen (1976) Halbinterpretative Arbeitstranskriptionen (HIAT). In: Linguistische Berichte, 45, 21-41

Ehlich, Konrad; Rehbein, Jochen (1979) Erweiterte halbinterpretative Arbeitstranskriptionen (HIAT2); Intonation. In: Linguistische Berichte, 59, 51-75

Ehlich, Konrad; Rehbein, Jochen (1981a) Zur Notierung nonverbaler Kommunikation für diskursanalytische Zwecke (Erweiterte halbinterpretative Arbeitstranskriptionen HIAT 2). In: Winkler, P. (Hg.) Methoden der Analyse von Face-to-Face-Situationen, Stuttgart: Metzler, 302-329

Ehlich, Konrad; Rehbein, Jochen (1981b) Die Wiedergabe intonatorischer, nonverbaler und aktionaler Phänomene im Verfahren HIAT. In: Lange-Seidl, A. (Hg.) Zeichenkonstitution, Bd. 2 (Akten des 2. Semiotischen Kolloquiums), Berlin: de Gruyter, 174-186

Ehlich, Konrad (1992a) Computergestütztes Transkribieren - das Verfahren HIAT-DOS. In: Richter, Günther (Hg.) Methodische Grundfragen der Erforschung gesprochener Sprache, Frankfurt a.M.: P. Lang, 47-59

Ehlich, Konrad (1992b) HIAT - a Transcription System for Discourse Data. In: Jane A. Edwards; Martin D. Lampert (eds.) Talking Data: Transcription and Coding in Discourse Research, Hillsdale, NJ.: L. Erlbaum Ass., 123-148

Ehlich, Konrad; Redder, Angelika (1994) Einleitung. In: Redder, Angelika; Ehlich, Konrad (Hg.) Gesprochene Sprache - Transkripte und Tondokumente, Tübingen: Niemeyer, 1-17

Lenk, Uta (1999) Notation Systems in Spoken Language Corpora. In Verschueren, Jef et al. (eds.) Handbook of Pragmatics 1999. Amsterdam: John Benjamins

Glas, Reinhold; Ehlich, Konrad (2000 Deutsche Transkripte 1950 bis 1955. Ein Repertorium. Hamburg: HAZEMS

Schneider, Wolfgang (2001) Der Transkriptionseditor HIAT-DOS. In: Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion, 2 (2001), 29-33

4.3.3 Coverage

Which types of raw data are referenced?

Direct referencing of non-transcribed material does not seem to be foreseen. The tool and the annotation (meta-scheme) are meant to make use of transcripts. These can possibly be linked (outside the HIAT tool), to video etc.

Which modalities is the coding scheme meant to code?

HIAT is mainly intended for conversation analytical research; its main application is thus in the transcription of situated dialogue, of multiparty conversation (talk show, classroom, etc.). Up to three freely definable tracks for non-verbal communication (examples mainly include gesture and body movement) per speaker are foreseen.

Which annotation level(s) does the coding scheme cover?

The HIAT system and the HIAT transcription convention set can both accommodate all kinds of linguistic, paralinguistic and non-verbal communication levels. Examples are given below. HIAT was produced with independence from a specific coding scheme for a given level and openness to all kinds of relevant coding schemes.

Which coding tasks has the coding scheme been used for?

Frequent types of usage include the following:

- different speakers; temporal relationships between speakers;
- transcripts, gesture, facial expression, paralinguistic events and prosody, and their temporal interrelationships;
- transcripts and comments of discourse analysts, aligned to each other.

There are more applications (outside the realm of this survey), such as interlinear translation, collations of manuscripts, semantic, co-referential etc. relationships between component parts of texts, discourses etc., and think-about protocols.

4.3.4 Detailed description of coding scheme

Which header file information is included (meta-data)?

Header information is given on the following types of information:

- name (and file name) of transcript;
- project name and short explanation of the type of data transcribed;
- data and technicalities of recording;
- transcriber and corrector, dates, etc.;
- short informal descriptions of the contents of the transcript.

An example of a header screen from the HIAT-DOS tool, taken from the website (see URL mentioned above) is reproduced in the figure (taken from http://www.daf.uni-muenchen.de/HIAT/HIAT.HTM#HIAT_Programm) below:

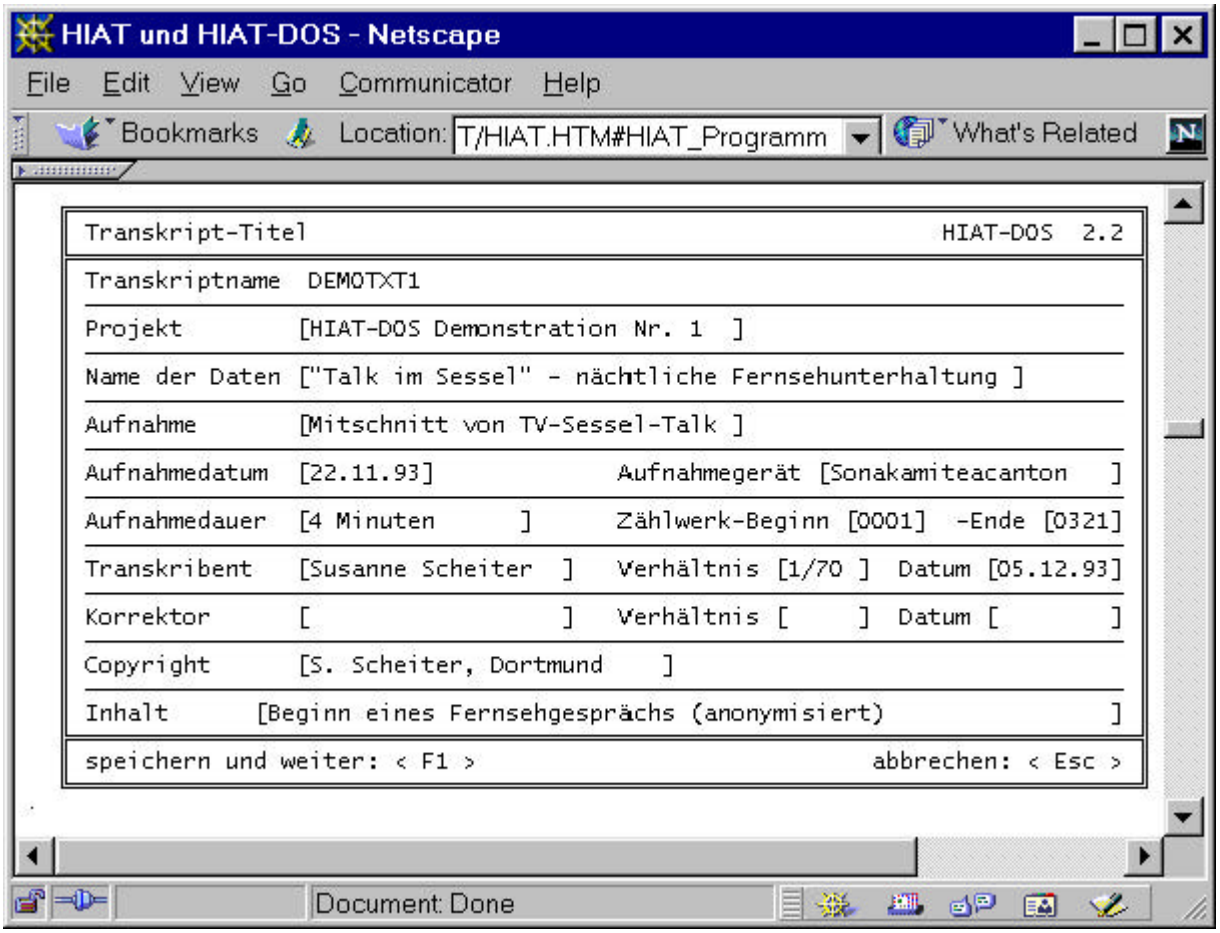


Figure 4.3.2. Example of a header screen from the HIAT-DOS tool.

Coding purpose of the coding scheme?

We see HIAT as a meta-scheme for different kinds of conversational analysis-aware annotations; it covers a range of phenomena from language, speech (intonation) and gesture/facial expression. It was designed to be sufficiently general to serve a broad range of applications.

List and description of phenomena, which can be annotated by the scheme

1 Prosodic features (intonation, stress, pauses (location and duration)

2 Paralinguistic features (volume, tempo, quality of articulation)

3 Speech management phenomena: self correction reformulations, hesitations, back channelling utterances

4 Transturn data: simultaneity of communication, represented by score notation

5 non-phonological acoustic phenomena (Laughter, sighing)

6 non-phonological non-acoustic phenomena: mainly gestures, described in up to three additional lines in each speaker's band, plus a comment field at the left of the bands

7 Speaker attribution: by prefixes that precede each turn

8 Mode of transcription: literal

Description of markup language/markup declaration

HIAT defines annotation-conventions but does not deal with storage formats of annotated data and computer-technical details.

HIAT-DOS (for DOS/Windows-PCs) is one of the two editors (SyncWriter for MACs is the other one), specialized to help annotate according to the HIAT-conventions. They differ eminently in the technical way used to carry out this task. Both editors store their data in undocumented proprietary formats, but they do not produce SGML- or XML-based tag-oriented text. The preferred transfer-route for text from HIAT-DOS to standard editors or analysis-tools is via an export towards rich-text-format (rtf-Unicode on demand).
A HIAT-DOS-from-RTF-to-HTML-converter is available (for free) to create HIAT-texts in HTML. This is for presentational purposes only, not for an automated tags-based-analysis.

Examples

No example available.

Description of coding procedure, if any

The coding scheme is described in **Ehlich/Rehbein** (1981a)

Creation notes, i.e. who wrote the coding scheme (contact details), when, and in which context?

Contact address:
Prof. Dr. Konrad Ehlich
Ludwig-Maximilians-Universität München
Institut für Deutsch als Fremdsprache
Kennwort HIAT

D-80539 München
Fax: 089/ 2180 - 3999

4.3.5 Usage

Origin of the coding scheme and reasons for creating it

HIAT (Halbinterpretative Arbeitstranskriptionen; "semi-interpretative working transcripts") is a system created in the 1970s (first references from 1976) in the framework of research in conversation analysis. The purpose of the annotation scheme (and of the pertaining tool HIAT-DOS) is to describe and annotate parallel tracks of verbal and non-verbal (e.g. gesture) communication in a simple way. The system is meant for relatively coarse-grained annotation; HIAT itself is a set of conventions for annotation (a sort of meta-scheme) much more then an annotation scheme for one specific modality.

How many people have used the coding scheme and for what purposes?

R.Glas (2000) gives an overview of "Deutsche Transkripte 1950 bis 1995". For this period he found 1274 transcripts. 221 (17%) of them were produced on the basis of HIAT. Since 1994, the computer program HIAT-DOS 2 is available. There are no actual data, but the number of users is growing from year to year.

How many dialogues/interactions have been annotated using the coding scheme?

See above for the figures from Glas/Ehlich 2000.

Has the coding scheme been evaluated?

So far, only an informal evaluation has taken place, in the form of contacts between users and developers, as well as via user support. Given the amount of material annotated already with HIAT, this gives a massive body of (so far unpublished) experience.

Is the coding scheme language dependent (which language(s)) or language independent?

The coding scheme is not in itself language dependent.

Is there tools support for using the coding scheme or API for editing/parsing coded descriptions? In which language?

The tool HIAT-DOS is designed to provide support for the annotation in the framework of HIAT. A brief mention of the tool in the D-11.1 report would have been useful. The HIAT-DOS-tool may be one of the earliest tools allowing for the annotation of text, gesture and other modalities.

4.3.6 Accessibility

How does one get access to the coding scheme?

HIAT is documented in publications, see URL:
http://www.daf.uni-muenchen.de/HIAT/HIAT.HTM#HIAT_Literatur

Is the coding scheme available for free or how much does it cost?

The tool is available for 20 EUR (students), 45 EUR (scholars) or 200 EUR (institutions).

4.3.7 Conclusion

How well described is the coding scheme?

The description of the coding scheme on the web is rather general. We expect there to be more details available in the individual publications cited above.

How general and useful is the coding scheme?

The HIAT scheme is very general (not language-bound, generic enough to potentially cater for more than just multimodal dialogue (multiparty communication, more parallel tracks than just gesture and facial expression), but it does not specify individual gestures or facial expressions.

4.4 LIMSI Coding Scheme for Multimodal Dialogues between Car Driver and Co-pilot

4.4.1 Description header

Main actor

LIMSI: Jean-Claude MARTIN (martin@limsi.fr)

Verifying actor

LIMSI: Xavier BRIFFAULT (briffault@limsi.fr)

Date of last modification of the description

29 August 2001

4.4.2 Reference

Web site

http://www.limsi.fr/Individu/xavier/Articles/RapportInterneXBMDAideALaNavigation/NDXBMD.html
and more specifically :
http://www.limsi.fr/Individu/xavier/Articles/RapportInterneXBMDAideALaNavigation/NDXBMD.html#RTFToC11

Short description

The coding scheme has been created for the annotation of a resource which contains multimodal dialogues between drivers and co-pilots during real car driving tasks.

One illustrative example of coding

dialogue: cni-v-5/02/93-1-dt. Pilot: V. Co-pilot: J.

At the beginning of the route:

[LIMSI/RPN118]

<00000/00335>

%v(c): bon je t'explique un peu...tu veux que je te dise où on va en gros? je vais te dire quand même hein

%v(p): oui, je veux bien...

%v(c): on va aller à paris faire un tour...on va faire un tour dans paris alors...on va partir par la 118 et puis on va revenir par l'autoroute

%v(p): ah d'accord, par la porte d'orléans

%v(c): voilà, entre, heu...oui, ça sera marqué, je te dirai à ce moment là

%v(p): oui

%v(c): pour le retour...donc là tu prends la 118

%v(p): donc là je vais prendre la 118 au rond point de saclay...au rond point là-bas?

%v(c): alors attends, je vais te dire...je vais te dire où tu vas la prendre la 118...donc tu vas prendre la 118 au carrefour habituel près de corbeville...

(...)

Reaching the end of the route:

{après pont}

<10902/11200>

v(c): ouais, je crois

v(p): donc...

v(c): on va vérifier

v(p): et maintenant?

v(c): on va, non /là-bas je/ pense, tout droit

g(c): ixtddr

% rectification expé qui indique de prendre la rue suivante pour retrouver le boulevard charles de gaulle.

[BLD GABRIEL PERI/PORTE DE CHATILLON]

<11200/11540>

[PORTE DE CHATILLON/ACCES AUTOROUTE]

<11540/11950>

% c indique de suivre la rue qui longe le périph jusqu'à trouver l'entrée de la A6. Mais après avoir examiné la carte plus en détail, décide de poursuivre jusqu'à porte de gentilly

[ACCES AUTOROUTE]

<11950/12010>

[A6/A10]

<12010/12500>

[A10/SORTIE SACLAY]

<12500/12800>

[SORTIE SACLAY/RPN118]

<12800/13000>

[RPN118/LIMSI]

References to additional information on the coding scheme

Chalme, S., Briffault, X., Denis, and M., Gaunet, F.: Experiments For Designing Multimodal Dialogue Interfaces In Navigational Aid Systems : Real Versus Simulated Driving Situations. Dsc'99 (Driving Simulation Conference), Paris, Juillet 1999.

Denis M., Briffault, X.: Analyse Des Dialogues De Navigation À Bord D'un Véhicule Automobile. Le Travail Humain, Tome 63, N° 1/2000.

Briffault, X., and Denis, M.: Analyses D'un Corpus De Dialogues De Navigation À Bord D'un Vehicule Automobile. Technical Report, Limsi N°95-28, 1995.

4.4.3 Coverage

Which types of raw data are referenced?

Video files are referenced via their name and a time code in the transcription.

Which modalities is the coding scheme meant to code?

Speech, hand gesture, head gesture, gaze.

Which annotation level(s) does the coding scheme cover?

Facial expressions and prosody are not covered.

Which coding tasks has the coding scheme been used for?

The coding scheme has been created for the annotation of a resource which contains multimodal dialogues between drivers and co-pilots during real car driving tasks.

4.4.4 Detailed description of coding scheme

Which header file information is included (meta-data)?

The experimental protocol (cni-v). The date of the recording (i.e. 5/02/93). The number of recording on that date (i.e. 1). The name of the experimenter who made the recording, the name of the pilot and the name of the co-pilot.

Coding purpose of the coding scheme?

The objective was to investigate the components of the dialogue between the driver and the co-driver, the communication modalities used and the strategies of direction giving in order to provide guidelines for the specifications of an interactive navigational aid system for cars.

List and description of phenomena, which can be annotated by the scheme

Monomodal behaviour (verbal, hand gesture, gaze, head gesture) and temporal relationships between modalities.

Description of markup language/markup declaration

v stands for verbal
g stands for gesture
c stands for human copilot
p stands for human pilot

/ and \ stands for begin and end of gesture. Such markup are integrated in the verbal descriptions and are followed by the coding of gesture on a line beginning by g

% stands for a comment written by the encoder

[and] are used for defining successive segments of the itinerary ({ and } are used for coding subparts of such segments)

< timecode-begin / timecode-end > is used to refer to a segment of the video

coding of the gesture shape :

coding of the body part:

te=tête (head)

ma=main (hand)

mo=menton (chin)

ms=mains (both hands)

fingers : ix=index (first finger), mj=majeur (middle finger), an=annulaire (ring finger), au=auriculaire (little finger), po=pouce (thumb)

gaze : oc= short glance on the map, ol= long glance on the map

coding of the shape of the body part:

td=tendu (tense)

sp=souple (loose)

cr=crochet (hook)

coding of the global movement:

mv=mouvement ample (wide mouvement),

mr=mouvements répétés (repeated movement)

( )=statique

coding of the direction of movement:

ar=arrière (backwards)

tr=transversal (side)

ci=circular

coding of the meaning of gesture:

ds=designation (ie. reference to a street, a building …)

ca= designation on the map

dr=direction

dc=description (i.e. Configuration of a place)

pc=position (i.e. In order to prepare for a change of direction)

These different codes are combined into a global coding of the observed gesture. For example, first finger tense in direction of ... is coded "ixtddr".

Examples

Cf. previous example in Section 4.4.2. Other examples are available on the web site.

Description of coding procedure, if any

Information not available.

Creation notes, i.e. who wrote the coding scheme (contact details), when, and in which context?

Information not available.

4.4.5 Usage

Origin of the coding scheme and reasons for creating it

The coding scheme has been created for a specific corpus.

How many people have used the coding scheme and for what purposes?

Information not available.

How many dialogues/interactions have been annotated using the coding scheme?

54 subject have been recorded in the whole resource (27 drivers and 27 co-pilots).
The global size of the annotation files is 400 000 characters.

Has the coding scheme been evaluated?

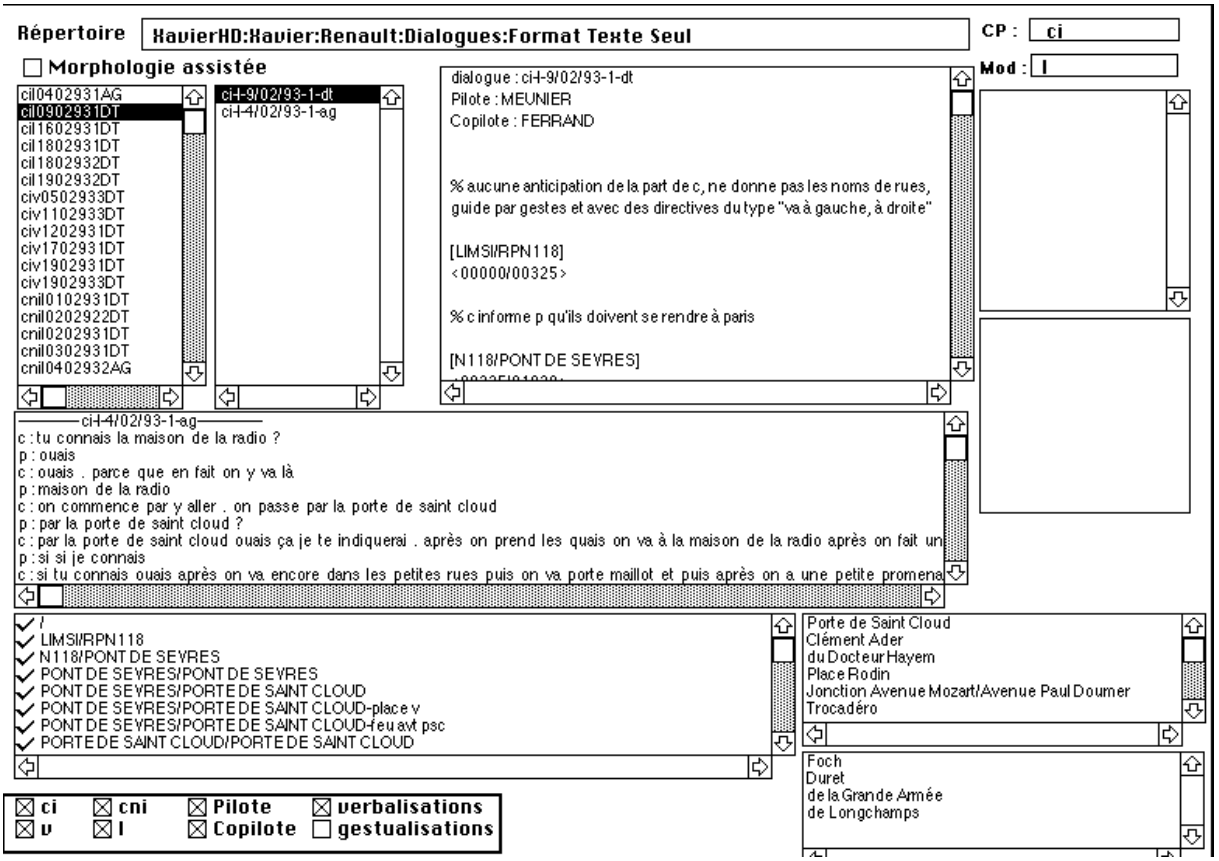
Information not available.

Is the coding scheme language dependent (which language(s)) or language independent?

Language dependent (the codes are abbreviations of French words).

Is there tools support for using the coding scheme or API for editing/parsing coded descriptions? In which language?

Software was developed (in Smalltalk) for computing statistics (syntactical categories, length of sentences) and comparing them between the pilot and co-pilot, and in different protocol settings.



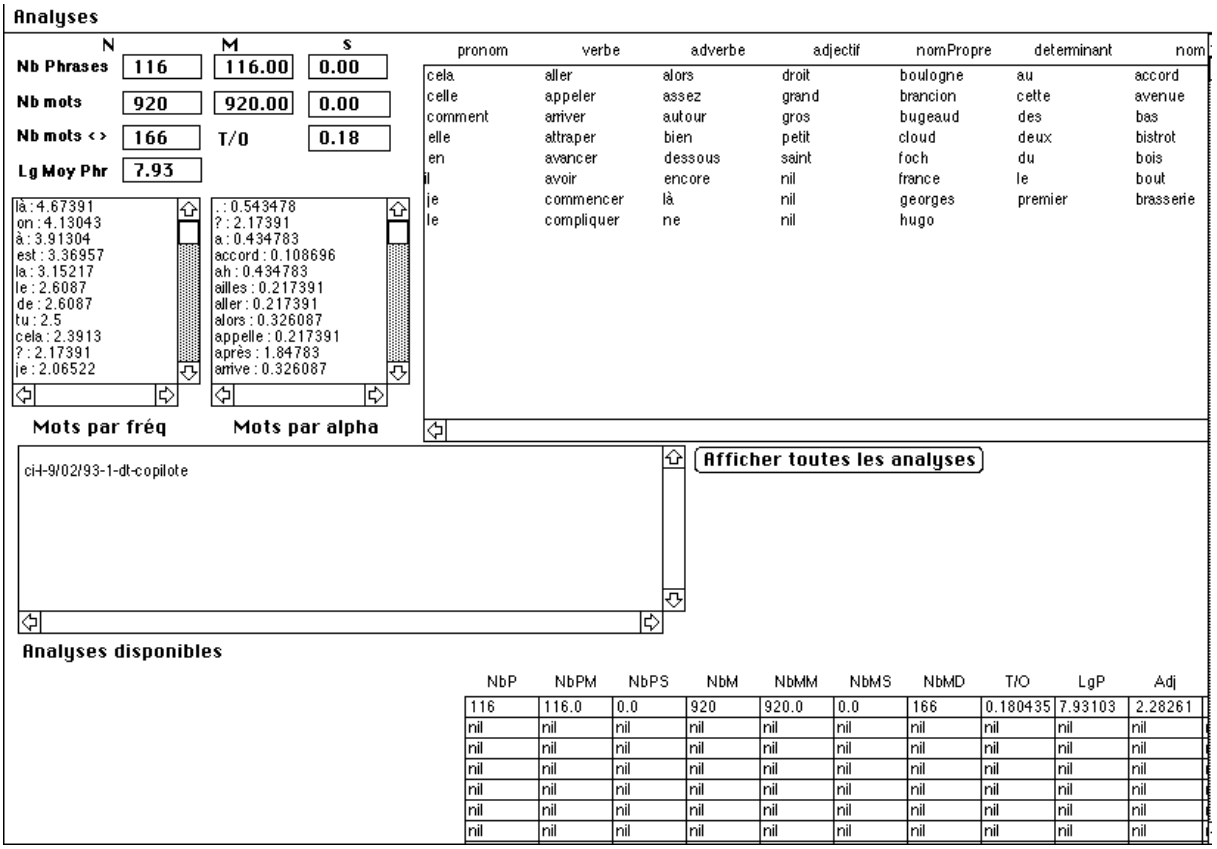


Figure 4.4.1. Screendump of the software developed to enable the filtering of specific behaviours in the corpus and the computation of statistics.

4.4.6 Accessibility

How does one get access to the coding scheme?

Contact briffault@limsi.fr

Is the coding scheme available for free or how much does it cost?

Information not available.

4.4.7 Conclusion

How well described is the coding scheme?

The coding scheme is well described in the document.

How general and useful is the coding scheme?

Although applied to only one (but huge) corpus, it seems that the coding scheme's principles can be applied to other application which has similar needs than the car application.

4.5 MPI GesturePhone

4.5.1 Description header

Main actor

MPI: Gijs van Elswijk (gijsve@mpi.nl), Peter Wittenburg (pewi@mpi.nl)

Verifying actor

MPI: Sotaro Kita (kita@mpi.nl)

Date of last modification of the description

10/05/2001

Date of last verification of the description

16/05/2001

4.5.2 Reference

Web site

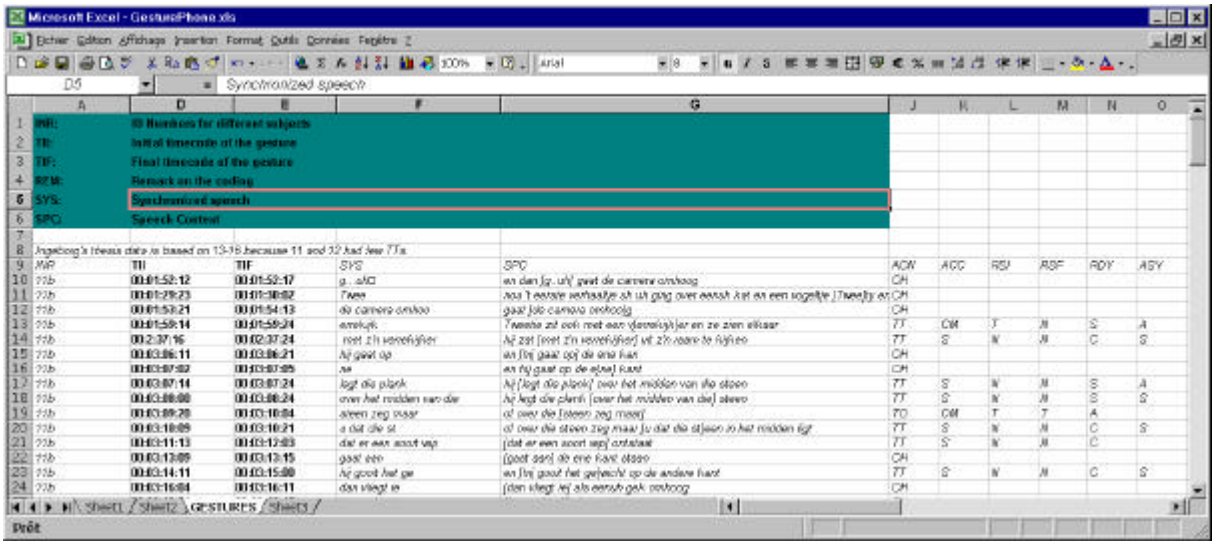
N/A

Short description

A coding scheme to transcribe signs and gestures.

One illustrative example of coding

See Excel worksheet "GesturePhone.xls"



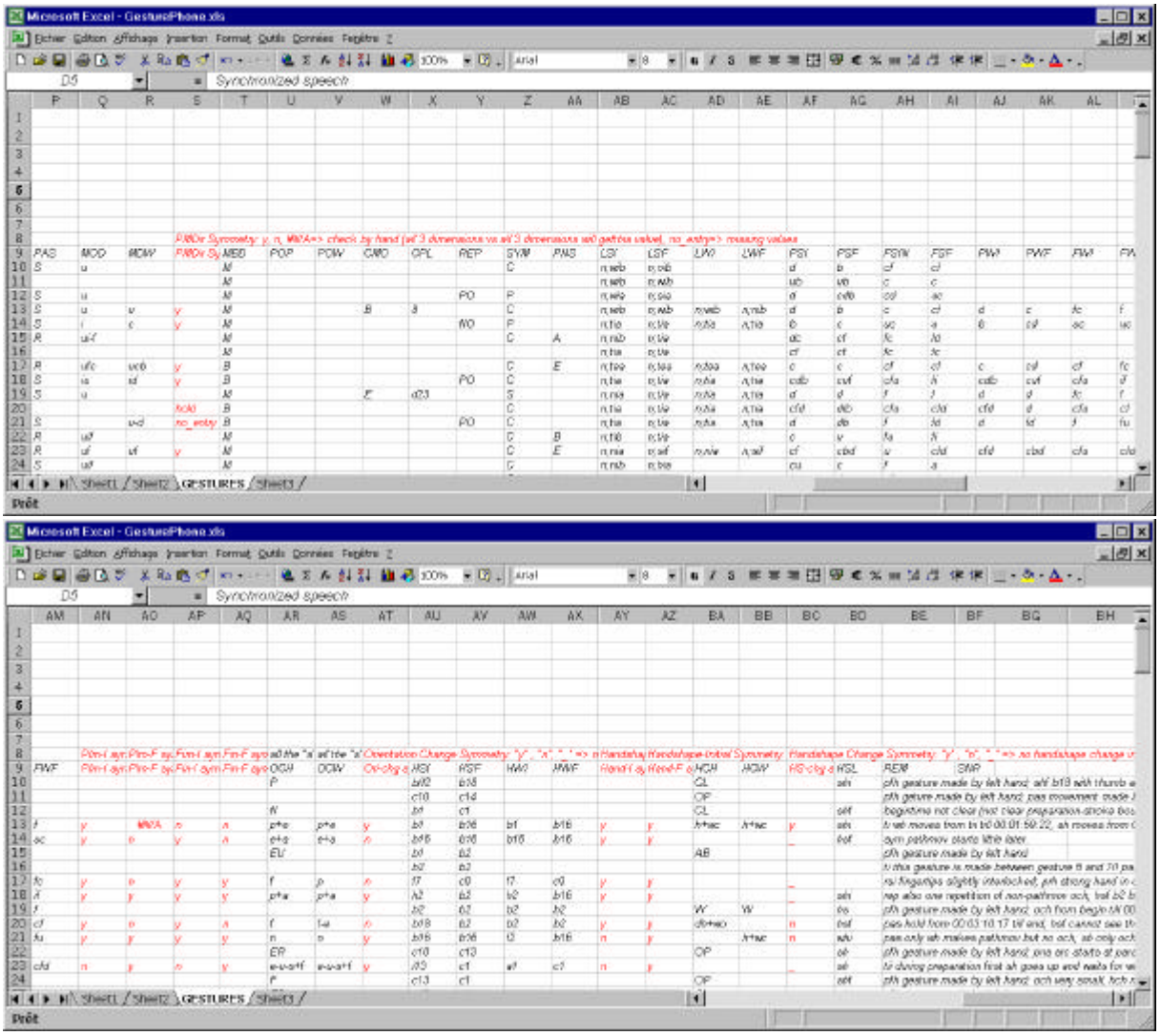


Figure 4.5.1. The figure gives an example of the used gesture annotations in a condensed form as an Excel spreadsheet. For each time segment gestures are coded as vectors with 48 formal codified annotations. The categories with help of which gestures are coded are described in the coding scheme.

References to additional information on the coding scheme

Van Gijn, I., Kita, S., van der Hulst, H. (submitted). How phonetic is the symmetry condition in sign language? HILP4 Proceedings.

4.5.3 Coverage

Which types of raw data are referenced?

Video files.

Which modalities is the coding scheme meant to code?

Hand and arm, sign and gesture.

Which annotation level(s) does the coding scheme cover?

Gesture and sign.

Which coding tasks has the coding scheme been used for?

Most of the gesture resources in the MPI gesture-database.

4.5.4 Detailed description of coding scheme

Which header file information is included (meta-data)?

IMDI standard.

Coding purpose of the coding scheme?

Research of the relation between gesture and speech.

List and description of phenomena, which can be annotated by the scheme

Path movement shape, path movement direction, hand orientation change, hand shape change, hand orientation, hand shape

Description of markup language/markup declaration

Tables in a relational database.

Examples

```
Path movement direction
A movement direction can involve more than one axis at the same time (e.g. up + front)
<up,down>, <front,back>,<ipsilateral, contralateral>
```

Description of coding procedure, if any

Well-trained student assistants who understand the principles of the coding system in detail apply the rules by first identifying the time boundaries of a gesture. Subsequently they have to select the codes from predefined controlled vocabularies for each category to limit the amount of errors.

Creation notes, i.e. who wrote the coding scheme (contact details), when, and in which context

van Gijn, I. (1997). Vormkarakteristieken van gestures en gebaren. MA thesis, Leiden University.
This coding scheme is based on SignPhon, see:
Blees, M., Crasborn, O., van der Hulst, H., & van der Kooij, E. (1996). SignPhon. A database tool for phonological analysis of sign languages. Manual, version 0.1. Leiden Sign Phonology Group.

4.5.5 Usage

Origin of the coding scheme and reasons for creating it

Study of Symmetry Condition in gesture in Dutch and Dutch Sign Language.

How many people have used the coding scheme and for what purposes?

Ingeborg van Gijn and Sotaro Kita.

How many dialogues/interactions have been annotated using the coding scheme?

4 for Dutch Sign Language.

6 for Dutch.

Has the coding scheme been evaluated?

The coding scheme has been used by three laboratories and has proven its great usefulness for the scientific analysis of gestures.

Is the coding scheme language dependent (which language(s)) or language independent?

Language independent.

Is there tools support for using the coding scheme or API for editing/parsing coded descriptions? In which language ?

The coding scheme (the categories and its controlled vocabularies) are supported in two tools developed by the MPI: TranscriptionEditor and MediaTagger. A support in the EUDICO tool set is intended.

4.5.6 Accessibility

How does one get access to the coding scheme?

Via Sotaro Kita.

Is the coding scheme available for free or how much does it cost?

Available for free.

4.5.7 Conclusion

How well described is the coding scheme?

Well described in an MA thesis (see section 4).

How general and useful is the coding scheme?

The presented coding scheme allows an exhaustive encoding of movement patterns which occur in gestures. In so far it will allow to answer many different scientific questions of relevance. Not all categories have to be filled in immediately, i.e. iterative encoding is possible. Applying the whole scheme is very labour intensive of course.

4.6 MPI Movement Phase Coding Scheme

4.6.1 Description header

Main actor

MPI: Gijs van Elswijk (gijsve@mpi.nl), Peter Wittenburg (pewi@mpi.nl)

Verifying actor

MPI: Sotaro Kita (kita@mpi.nl)

Date of last modification of the description

10/05/2001

Date of last verification of the description

16/05/2001

4.6.2 Reference

Web site

N/A

Short description

A syntagmatic rule system for movement phases that applies to both co-speech gestures and signs.

One illustrative example of coding

R-GP-K: This contains phase coding for the right hand.
R-ST-K: This tier is an attribute tier for the above, this contains the stroke types (coded only for stroke phases).
L-GP-K: This contains phase coding for the left hand
L-ST-K: This tier is an attribute tier for the above, this contains the stroke types (coded only for stroke phases).

```
00:11:47:18 17693    00:11:48:00 17700   R-GP-K preparation
00:11:48:01 17701    00:11:48:07 17707   R-GP-K stroke
00:11:48:08 17708    00:11:48:16 17716   R-GP-K retraction
00:11:50:12 17762    00:11:50:15 17765   R-GP-K preparation
00:11:50:16 17766    00:11:50:24 17774   R-GP-K stroke
00:11:51:00 17775    00:11:52:05 17805   R-GP-K hold
00:11:52:06 17806    00:11:52:11 17811   R-GP-K preparation
00:11:52:12 17812    00:11:52:20 17820   R-GP-K stroke
00:11:52:21 17821    00:11:53:16 17841   R-GP-K hold
00:11:53:17 17842    00:11:54:04 17854   R-GP-K preparation
00:11:54:05 17855    00:11:54:22 17872   R-GP-K stroke
00:11:54:23 17873    00:11:56:16 17916   R-GP-K hold
00:11:56:17 17917    00:11:57:03 17928   R-GP-K preparation
00:11:48:01 17701    00:11:48:07 17707   R-ST-K single segment
00:11:50:16 17766    00:11:50:24 17774   R-ST-K bounce back
00:11:52:12 17812    00:11:52:20 17820   R-ST-K single segment
00:11:54:05 17855    00:11:54:22 17872   R-ST-K single segment
00:12:49:21 19246    00:12:50:09 19259   L-GP-K preparation
```

```
00:12:50:10 19260    00:12:50:16 19266   L-GP-K stroke
00:12:50:17 19267    00:12:50:23 19273   L-GP-K stroke
00:12:50:24 19274    00:12:51:13 19288   L-GP-K retraction
00:13:03:06 19581    00:13:03:14 19589   L-GP-K preparation
00:13:03:15 19590    00:13:04:02 19602   L-GP-K stroke
00:13:04:03 19603    00:13:05:03 19628   L-GP-K retraction
00:13:05:12 19637    00:13:05:23 19648   L-GP-K stroke
00:13:05:24 19649    00:13:06:12 19662   L-GP-K stroke
00:13:06:13 19663    00:13:07:03 19678   L-GP-K retraction
00:13:11:03 19778    00:13:11:07 19782   L-GP-K stroke
00:13:11:08 19783    00:13:11:12 19787   L-GP-K retraction
00:13:13:17 19842    00:13:14:01 19851   L-GP-K preparation
00:13:14:02 19852    00:13:14:15 19865   L-GP-K stroke
00:13:14:16 19866    00:13:14:24 19874   L-GP-K retraction
00:12:50:10 19260    00:12:50:16 19266   L-ST-K single segment
00:12:50:17 19267    00:12:50:23 19273   L-ST-K single segment
00:13:03:15 19590    00:13:04:02 19602   L-ST-K multi-segment
00:13:05:12 19637    00:13:05:23 19648   L-ST-K single segment
00:13:05:24 19649    00:13:06:12 19662   L-ST-K bounce back
00:13:11:03 19778    00:13:11:07 19782   L-ST-K single segment
00:13:14:02 19852    00:13:14:15 19865   L-ST-K multi-segment
```

References to additional information on the coding scheme

Kita, S., van Gijn, I., & van der Hulst, H. (1998). Movement phases in signs and co-speech gestures, and their transcription by human coders. In Wachsmuth, I., & Froehlich, M. (Eds.). Gesture and Sign Language in Human Computer Interaction: Proceedings / International Gesture Workshop, Bielefeld, Germany, September 17-19, 1997. Springer-Verlag Berlin Heidelberg.

4.6.3 Coverage

Which types of raw data are referenced?

Video files.

Which modalities is the coding scheme meant to code?

Gestures and signs.

Which annotation level(s) does the coding scheme cover?

Gestures and signs.

Which coding tasks has the coding scheme been used for?

Most of the gesture resources in the MPI-gesture database.

4.6.4 Detailed description of coding scheme

Which header file information is included (meta-data)?

IMDI standard.

Coding purpose of the coding scheme?

Research of the relation between gesture and speech.

List and description of phenomena, which can be annotated by the scheme

Preparation of gesture/sign, expressive phase of gesture/sign, retraction of gesture/sign.

Description of markup language/markup declaration

Tables in a relational database.

Examples

```
Gesture Unit = Gesture Phrases*
Gesture Phrase = (Preparation) => Stroke => (Retraction)
Preparation = Preparation => (Pre-stroke hold)
Stroke = Stroke => (Post-stroke hold)
Retraction is optional when another Gesture Phase follows.

Notations:
X = Y X consists of Y
* One or more occurrences of the element
=> Discrete transition
() Optional
```

Description of coding procedure, if any

First the boundaries of the whole gesture unit are determined. Then the fine structure is described.

Creation notes, i.e. who wrote the coding scheme (contact details), when, and in which context?

Sotaro Kita (Max-Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, kita@mpi.nl)
This scheme is based on the systems described in:
McNeill, D. (1992). Hand and Mind. University of Chicago Press, Chicago.
Kendon, A. (1972). Some relationships between body motion and speech. In: Siegman, A., Pope, B. (Eds.). Studies in Dyadic Communication. Pergamon Press, New York.
Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In: Key, M. R. (Ed.). The Relation Between Verbal and Nonverbal Communication. Mouton, The Hague.

4.6.5 Usage

Origin of the coding scheme and reasons for creating it

Coding of co-speech gestures and signs.

How many people have used the coding scheme and for what purposes?

Gesture researchers at the MPI and elsewhere.

How many dialogues/interactions have been annotated using the coding scheme?

Many

Has the coding scheme been evaluated?

Yes, coding scheme has been evaluated with 2 researchers from the MPI applying the categories from written instructions. Coding was performed on the basis of digitised video. For annotating a program called MediaTagger was used. The coding scheme yielded good inter-coder reliability.

Is the coding scheme language dependent (which language(s)) or language independent?

Language independent.

Is there tools support for using the coding scheme or API for editing/parsing coded descriptions? In which language?

MediaTagger, Eudico.

4.6.6 Accessibility

How does one get access to the coding scheme?

Via Sotaro Kita.

Is the coding scheme available for free or how much does it cost?

Available for free.

4.6.7 Conclusion

How well described is the coding scheme?

Well described in an article (see section 2).

How general and useful is the coding scheme?

The scheme claims to be general enough to describe the phase pattern of gestures.

4.7 MPML - A Multimodal Presentation Markup Language with Character Agent Control Functions

4.7.1 Description header

Main actor

IMS: Ulrich Heid (heid@IMS.Uni-Stuttgart.DE)

Verifying actor

LIMSI: Jean-Claude MARTIN (martin@limsi.fr)

Date of last modification of the description

11 October 2001

4.7.2 Reference

Web site

Material on MPML, including a tool for viewing and a 15 seconds movie with a multimodal presentation can be found at http://www.miv.t.u-tokyo.ac.jp/MPML/en/2.0e

Short description

MPML is a markup language for multimodal output of animated agents on web sites (speech, movement of the agent as such, pointing gestures). MPML is designed to encode the movements, pointing gestures and speech (speed, voice type, start and end time) of a little agent explaining the contents of a web site. The markup language is XML conformant and based on W3C standards.

As MPML is not meant for human/human dialogue, it is of less central importance for the ISLE work. It is however a good example of an XML-based markup language and of a tool set (partly existing partly upcoming) related to the XML format.

One illustrative example of coding

No data available.

References to additional information on the coding scheme

Tkayuki Tsutsui, Santi Saeyor and Mitsuru Ishiyuka: MPML: A Multimodal Presentation Markup Language with Character Agent Control Functions.

Yuan Zong, Hiroshi Dohi, Helmut Prendinger, Mitsuru Ishizuka: Emotion Expression Function in Multimodal Presentation.

4.7.3 Coverage

Which types of raw data are referenced?

As MPML is designed for the creation of presentations with synchronized media, it allows to bundle different output media (playing of speech, display of text and/or graphics, playing of video), along the

lines of the W3C's SMIL specification. However, the possibility to display synchronized data is radically different from an annotation of media files and/or transcriptions. MPML is a markup language, not primarily a coding scheme.

Which modalities is the coding scheme meant to code?

MPML concerns web sites with animated agents. It does not seem to have been used for human/human dialogue; the web presentations described in the papers analysed are all primarily meant for multimodal output, not even for man/machine dialogue.

Which annotation level(s) does the coding scheme cover?

The specification of MPML includes the following kinds of agent behaviour:
the agent's speaking (including language, voice, seed, start and end time
general movements of the agent (including target point and speed of movement)
specific movements of the agent, including pointing gestures.

Which coding tasks has the coding scheme been used for?

MPML has been used, so far, for the encoding of agent behaviour in web sites.

4.7.4 Detailed description of coding scheme

Which header file information is included (meta-data)?

Header information includes general information about the presentation encoded with MPML. The header contains a field for free text input and a field for the layout of the presentation. Note that the actions to be performed by the animated agent are encoded in the body of the MPML specification.

Coding purpose of the coding scheme?

The purpose of MPML is the encoding of agent behaviour in multimodal web sites (thus: multimodal output only).

List and description of phenomena, which can be annotated by the scheme

The phenomena are all meant for multimodal output:
Speech (in terms of chunks of canned speech or speech produced by a text-to-speech-system);
Movement of an agent (e.g. little character moving on the screen);
Pointing gestures: the agent can use its hands to point to an area on the screen.

Description of markup language/markup declaration

MPML is fully XML-conformant and based on standards from the W3 consortium (e.g. SMIL). MPML itself is a layer on top of other XML data, using ca. 30 specific tags defined within MPML and usable, for example, in the viewing tool provided on the MPML web site.

Examples

For a video sequence, see the following URL: http://www.miv.t.u-tokyo.ac.jp/MPML/en/2.0e (mpmlmovie.mpg to be selected).

Description of coding procedure, if any

We are not aware of a specific coding procedure. Reliability is not an issue in the set-up of MPML.

Creation notes, i.e. who wrote the coding scheme (contact details), when, and in which context?

References:
[Takayuki Tsutsui, Santi Saeyor and Mitsuru Ishiyuka]:
MPML: A Multimodal Presentation Markup Language with Character Agent Control Functions.
[Yuan Zong, Hiroshi Dohi, Helmut Prendinger, Mitsuru Ishizuka]:
Emotion Expression Function in Multimodal Presentation.

4.7.5 Usage

Origin of the coding scheme and reasons for creating it

MPML was created to support the construction of multimodal web sites; the purpose of MPML is to allow users to encode the voice and animation of an agent guiding a web site visitor through a web site.

How many people have used the coding scheme and for what purposes?

Information not available.

How many dialogues/interactions have been annotated using the coding scheme?

Information not available.

Has the coding scheme been evaluated?

Information not available.

Is the coding scheme language dependent (which language(s)) or language independent?

In so far as MPML is a meta-scheme which allows for the insertion of any kind of speech and language data, MPML is by definition language independent.

Is there tools support for using the coding scheme or API for editing/parsing coded descriptions? In which language?

At the point in time when the paper by [Zong et al. 2000] was written, an editor for MPML was under construction; moreover there exists a playing tool for viewing MPML presentations.
In addition, the authors of MPML have a converter from MPML to the command language of a given system, as well as an XSL plug-in for a web browser.
Since MPML is only of limited interest for ISLE given the narrow application domain and the small amount of actual gesture encoding, we do not think that there is need, in a future version of D-11.1 of ISLE, to deal with the MPML tools mentioned.

4.7.6 Accessibility

How does one get access to the coding scheme?

MPML, as well as the pertaining tools and samples are available for free at the web site mentioned above.

Is the coding scheme available for free or how much does it cost?

See above.

4.7.7 Conclusion

How well described is the coding scheme?

The description we have in textual form is rather general. But there is a DTD for MPML available online.

How general and useful is the coding scheme?

The scheme is quite limited in scope; this is however in line with its purpose. Given that the objective of MPML differs from that of a general coding scheme for the annotation of multimodal dialogues, the overall relevance and usefulness for ISLE are limited.

4.8 SmartKom Coding scheme

4.8.1 Description header

Main actor

Norbert Reithinger (norbert.reithinger@dfki.de)

Verifying actor

LIMSI: Jean-Claude MARTIN (martin@limsi.fr)

Date of last modification of the description

3. September 2001

Date of last verification of the description

12 February 2002

4.8.2 Reference

Web site

www.phonetik.uni-muenchen.de/Bas/BasHomeeng.html

Short description

The goal of the SmartKom project (http://www.smartkom.org) is to develop a system for natural interactivity, using speech, gestures, and facial expression as interaction media. As development, training, and test data, a corpus of WOZ dialogs is collected and annotated.

One illustrative example of coding

Figure 4.8.1. The figure shows the 4 view video stream of all video channels that are recorded during the WOZ experiments. On the left top position the output of the mimics camera can be seen. The top right is the posture of the subject. The left lower shows the view of the gesture camera and the right lower shows the graphical surface.

The gestures are stored in a separate file from the transliterations. An example file is as follows:

```
; DVD: 14
; Dialog: w058_pk
; zuletzt bearbeitet am: 16.7.01
; Aufnahme-Qualitaet: ok
; Anmerkungen: --
; Erst-Labeling: Bernd
; Korrektur: Bernd
; End-Korrektur: Silke
; VPK: ABC
GES: 331520     16640   I-Geste   I- tipp +          Zeige li Hand   links oben          Treffer   342400  1280
GES: 1203200    32000   U-Geste U - überleg - k      Zeige li Hand   links unten
GES: 1264640    17920   I-Geste   I - tipp -         Zeige li Hand   ~Das+f"unfte+Element   links   unten
        Treffer 1278720
GES: 3376000    26240   U-Geste U - les – p          Zeige li Hand   links unten
```

References to additional information on the coding scheme

Silke Steininger: Transliteration of Language and Labeling of Emotion and Gestures in SmartKom. LMU, März 2001, SmartKom Report 1

Silke Steininger: Labeling Gestures in SmartKom - Concept of the Coding System LMU, März 2001, SmartKom Report 2

Silke Steininger, Bernd Lindeman, Thosrsten Paetzold: Labeling von Gesten im Mensch-Maschine Dialog – Gesten- Kodierkonventionen SmartKom Version 2. LMU, März 2001. SmartKom Technisches Dokument 14.

The last technical document describes the currently valid annotation scheme and procedure.

4.8.3 Coverage

Which types of raw data are referenced?

The video and audio data are referenced from the transliteration. The transliteration and the gesture label files are time aligned

Which modalities is the coding scheme meant to code?

The gesture coding covers three types of hand gestures

- Interactive
- Supporting
- Others

The coding scheme does not exactly code the morphology of the gesture, but rather the function of a gesture.

Which annotation level(s) does the coding scheme cover?

The SmartKom annotation covers prosody, gesture, and mimics. In this description we limit ourselves to gestures.

Which coding tasks has the coding scheme been used for?

The scheme was used to code multi modal information in the SmartKom WOZ dialogs.

4.8.4 Detailed description of coding scheme

Which header file information is included (meta-data)?

The header contains information about the DVD number, the dialogue number, the last modification date, recording quality, remarks, initial labeller, correcting labeller, and speaker identification.

Coding purpose of the coding scheme?

The purpose is to provide information about the intentional information contained in a gesture.

List and description of phenomena, which can be annotated by the scheme

The gesture types are

- Interactive (I-Geste): they are used to solve a task, e.g. deictic gestures on elements of the screen.
- Supporting (U-Geste): they don't relate directly to the screen's content, but are used to prepare the cognitive process of thinking, verbalizing, and acting.
- Others (R-Geste): they are labelled to all gestures not related to the task-solving process, e.g. scratching the skin.

Each gesture has sub-types, like encircling gesture, pointing gesture, or continuous stroke.

Description of markup language/markup declaration

Each gesture description start with GES. Then, separated by a tab, the label file contains

- Onset time of the gesture
- Duration of the gesture
- Label
- Morphology
- Reference word(s) in the transliteration file
- Object
- Onset stroke
- Duration of stroke
- Remarks

An example is a gesture, where the subject pointed on the screen ("I-tipp") used the left hand ("Zeige li Hand") and hit the object ("Treffer"). The gesture stroke started with millisecond 342400 and lasted 1280 milliseconds.

```
GES: 331520     16640   I-Geste  I- tipp +          Zeige li Hand     links oben        Treffer   342400  1280
```

Examples

See above for an example.

Description of coding procedure, if any

The labellers use the tool "Interact" (www.mangold.de). The labellers know the annotation manual (SmartKom technical document 14, version 2) and are trained. Each labelling is checked by a second labeller.

Creation notes, i.e. who wrote the coding scheme (contact details), when, and in which context?

See above. The SmartKom project is described at www.smartkom.org, the recordings and annotations are produced by the Bavarian Archive for Speech Signals (www.phonetik.uni-muenchen.de/Bas/BasHomeeng.html).

4.8.5 Usage

Origin of the coding scheme and reasons for creating it

For the multi-modal interaction system SmartKom, annotated multimedia data is needed that contains annotation of gestures and mimics to train and to test the system modules for gesture and mimics recognition.

How many people have used the coding scheme and for what purposes?

The exact number is not available. The annotations are done at one site.

How many dialogues/interactions have been annotated using the coding scheme?

Currently 33 dialogues of the SmartKom WOZ corpus have been annotated. All dialogues collected for this corpus will be annotated.

Has the coding scheme been evaluated?

No evaluation yet.

Is the coding scheme language dependent (which language(s)) or language independent?

The scheme should be language independent.

Is there tools support for using the coding scheme or API for editing/parsing coded descriptions? In which language?

Currently, there are only project internal tools available to process the files. The annotation tool is commercial and should be evaluated for ISLE report 11.1.

4.8.6 Accessibility

How does one get access to the coding scheme?

Currently, the technical documents are for project internal use only.

Is the coding scheme available for free or how much does it cost?

The availability is determined by BAS.

4.8.7 Conclusion

How well described is the coding scheme?

The coding scheme is described in a 29-page manual (German) and can be used to train annotators.

How general and useful is the coding scheme?

The scheme is useful in the field of natural interactivity research, especially for multi-modal interaction systems. It does not describe every detail of the morphology of gestures, but concentrates on the functional aspects of gestures in the system set-up.

4.9 SWML (SignWriting Markup Language)

4.9.1 Description header

Main actor

IMS: Ulrich Heid (heid@IMS.Uni-Stuttgart.DE)
We would like to thank Antônio Carlos da Rocha Costa, Pelotas, Brazil, for his helpful comments on an earlier version of this section and for additional information he provided, beyond what hat been analysed initially by the author.

Verifying actor

LIMSI: Jean-Claude MARTIN (martin@limsi.fr)

Date of last modification of the description

11 October 2001

4.9.2 Reference

Web site

http://swml.ucpel.tche.br/swml.htm

Short description

SWML is a markup language for the SignWriting system. It is XML-based, which makes it technically interesting for ISLE. SignWriting can in some sense be compared with HamNoSys: SignWriting focuses on gesture components and sequences, facial expressions, etc. and is to be seen as a markup language for computational treatment of documents written in SignWriting symbols. For example, the Brazilian Sign language LIBRAS has been written in SignWriting and computationally treated (queried) via its SWML rendition.

One illustrative example of coding

See the paper by Roche Costa/Dimuro 2001, http://swml.ucpel.tche.br

References to additional information on the coding scheme

DTD: http://swml.ucpel.tche.br
See also the links given in [Rocha Costa/Dimuro 2001].

4.9.3 Coverage

Which types of raw data are referenced?

Since SWML is an encoding system for a writing system, it is not meant to make reference to external media (like video); at least we have this impression from the documentation.

Which modalities is the coding scheme meant to code?

SWML codes utterances in sign languages written in the SignWriting System.

Which annotation level(s) does the coding scheme cover?

SignWriting comprises not only the "linguistic" aspects of sign languages (e.g. hand shapes, positions etc.), but also facial expression, dynamic aspects of gestures (smoothness etc.) and "punctuation"; these can all be rendered in SWML, as well.

Which coding tasks has the coding scheme been used for?

Transcription of utterances in the SignWriting System: processing and storage of sign language documents written in the SignWriting System, as well as the insertion of sign language texts into HTML documents.

4.9.4 Detailed description of coding scheme

Which header file information is included (meta-data)?

Creation of the document by:
- a SWML-aware editor;
- a sign language database.

Coding purpose of the coding scheme?

To support exchange and interoperability of SignWriting aware software.

List and description of phenomena, which can be annotated by the scheme

SWML text includes:
- sign boxes (which contain signs, i.e. sets of symbols);
- text boxes (alphanumeric text).

For details see [Rocha Costa/Dimuro 2001].

Description of markup language/markup declaration

There is a DTD for SWML in [Rocha Costa/Dimuro 2001]. cf. also the URL http://swml.ucpel.tche.br/dtd-version1.0-draft2.htm

Note that handling of graphical symbols involves the conversion of XML to Graphics formats. In [Rocha Costa/Dimuro 2001], SVG is proposed as a possible upcoming standard.

A full set of symbols for use in web applications is available in GIF; VML is also under development. SWML encoded SignWriting can be presented in any graphical format (GIF, JPEG, SVG, VML, etc.).

For details on the GIF inventory, see the following URL: http://swml.ucpel.tche.br/sss-1995/index.htm

Examples

See [Rocha Costa/Dimuro 2001]: URL: http://swml.ucpel.tche.br .

Description of coding procedure, if any

Since SWML is the computational encoding of SignWriting, the actual linguistic coding procedures are part of SignWriting, not of SWML. Inherent to SWML is the encoding procedure for the graphical symbols of SignWriting.

We do not go back to SignWriting here as the details of sign languages are beyond the scope of the present report (for details cf. http://www.signwriting.org)

Creation notes, i.e. who wrote the coding scheme (contact details), when, and in which context?

[Rocha Costa/Dimuro 2001]

4.9.5 Usage

Origin of the coding scheme and reasons for creating it

SWML is an XML-based representation system (and pertaining software) for the SignWriting system developed by Valerie Sutton, for the Center for Sutton Movement Writing ; SWML is thus one way of computationally encoding Sign Writing in XML. It is comparable to the XML encoding of HamNoSys which is being developed at Hamburg, partly in the ViSiCAST project.

How many people have used the coding scheme and for what purposes?

No information available.

How many dialogues/interactions have been annotated using the coding scheme?

No information available.

Has the coding scheme been evaluated?

No information available.

Is the coding scheme language dependent (which language(s)) or language independent?

SWML is dependent on SignWriting. SignWriting itself is not language dependent (and thus, SWML inherits from it its language independence); SignWriting does not have a linguistic stance, but is intended to represent movements; instead of linguistic entities it focuses on gesture sequences, facial expressions, etc. (cf. lanceWriting). The examples cited in the SWML documents come from American work in support of deaf people and from LIBRAS, the Brazilian Sign language.

Is there tools support for using the coding scheme or API for editing/parsing coded descriptions? In which language ?

SWML comes with a searching and a matching procedure described by [Rocha Costa/Dimuro 2001]. There is a query tool for written signs; as the signs are renditions of the SignWriting symbols, SWML and the query facility provide a search functionality for documents in SignWriting. No need for inclusion in D-11.1, as the topic is too far away from the tasks of ISLE. in addition, there are online converters at http://swml.ucpel.tche.br/what_achieved.htm

4.9.6 Accessibility

How does one get access to the coding scheme?

Cooperation with interested parties is sought; cf. the URL: http://swml.ucpel.tche.br

Is the coding scheme available for free or how much does it cost?

See above.

4.9.7 Conclusion

How well described is the coding scheme?

The information accessible to us is sufficient for the purpose of writing the present short summary.

How general and useful is the coding scheme?

As in the case of HamNoSys, the objectives of SWML and those of SignWriting are rather far away from those of ISLE. There is a relevant technical aspect of SWML: it is represented in XML in the form of a graphical/geometrical inventory of symbols; for these, a matching algorithm is also given and a query tool developed on that basis.

4.10 TUSNELDA Corpus Annotation standard

4.10.1Description header

Main actor

IMS: Ulrich Heid (heid@IMS.Uni-Stuttgart.DE)

Verifying actor

We would like to thank Andreas Wagner of Tübingen University for his very valuable comments and updates on an earlier version of this section.

Date of last modification of the description

Friday, 28 December 2001

4.10.2Reference

Web site

http://www.sfb441.uni-tuebingen.de/tusnelda-online.html

Short description

The TUSNELDA annotation standard is a corpus encoding standard that takes into account the needs of linguistic research using a variety of linguistic data structures for a variety of languages.

One illustrative example of coding

Figure 4.10.1. Transcription of a comic picture (source: Goscinny/Uderzo - Asteriks u Belgiji).

```
<figure id="s35b5" entity="belgiji/s35b5.bmp">
    <figtrans>
      <sp who="Obeliks">
```

```
      <spokenpar>
        Gde da na&dstrok;em belu zastavu ?
        <marked type="deic-loc">Ovde</marked> je sve pusto !
      </spokenpar>
      <situation>
        <keywords>
          <term>open hands <term>
          <term>slightly bent</term>
        </keywords>
      </situation>
    </sp>
    <sp who="Asteriks">
      <spokenpar>
        <marked type="deic-loc">Tamo</marked> je neki mališan !
      </spokenpar>
        <situation>
          <keywords>
            <term>forefinger</term>
            <term>stretched out</term>
          </keywords>
        </situation>
    </sp>
  </figtrans>
</figure>
```

References to additional information on the coding scheme

Cf. "Creation notes, ... references to literature" (p. 61).

4.10.3Coverage

Which types of raw data are referenced?

Most of the description published has to do with text and the annotation of dialogues. For dialogues, as well as for the transcription of dialogues in comics, situational coding as well as gesture coding has been included into TUSNELDA. TUSNELDA assumes the availability (and module-wise inclusion) of linguistic annotations, such as part of speech annotations. The integration of morphological and syntactic annotation schemata is currently in preparation.

Which modalities is the coding scheme meant to code?

Data sources are spoken dialogue, comics (i.e. picture sequences and text corresponding to spoken dialogue), including an annotation of gesture.

Which annotation level(s) does the coding scheme cover?

Levels are transcriptions, names of speakers, as well as a situational annotation in terms of key words and deictic gesture labels.

Which coding tasks has the coding scheme been used for?

TUSNELDA is used as an annotation standard common to several corpora for linguistic research on different languages (currently including German, Bosnian/Serbian/Croatian, Russian, Spanish and Portuguese) and for analysing different linguistic phenomena such as modal verbs in German, politeness phenomena in Russian, deictic expressions in Serbian etc. The interesting aspect for standardization in TUSNELDA is that it is supposed to be usable for all these purposes, and to support as much as possible automatic tools for annotation and query.

4.10.4 Detailed description of coding scheme

Which header file information is included (meta-data)?

A number of philological tags have been introduced or modified. In [Wagner/Kallmeyer01] the changes with respect to CES and TEI are discussed in detail and listed.

Coding purpose of the coding scheme?

TUSNELDA is aimed at philological work on corpora from many different languages, including the annotation of text-and-image-sequences, e.g. from comics.

List and description of phenomena, which can be annotated by the scheme

As situational characteristics are encoded by keywords (or plain text, respectively) which can be flexibly chosen by the annotator, TUSNELDA is generally open to the annotation of any gesture and situational phenomenon. Currently, for the annotation of deictic gestures, a fixed set of classifying keywords is used. This classification differentiates between five different ways of pointing with the hand(s) ("forefinger", "thumb", "open hand", "open hands", and "holding forth") in combination with three different postures of the arm ("bent", "slightly bent" and "stretched out"). The corresponding deictic expressions are classified as "deic-dem", "deic-loc", "deic-pres", "deic-qual", "deic-quant", or "deic-temp", respectively.

Description of markup language/markup declaration

Detailed description of the coding scheme: in general, TUSNELDA is derived from the CES. About one third of the TUSNELDA definition differ from the respective CES definitions or do not have equivalents in CES. A little less than two thirds of the TUSNELDA definitions have been taken over without change from CES. Emphasis in TUSNELDA is on mechanisms for empirical philological research, but the compatibility with the basic devices of CES has been a major aspect as well. TUSNELDA is in SGML. The encoding language is XML compatible.

Examples

See the coding manual at URL: http://www.sfb441.uni-tuebingen.de/c1/tusnelda-guidelines.html

Description of coding procedure, if any

So far, no details of the coding procedures have been published, likely because coding is still ongoing at an experimental level.

Creation notes, i.e. who wrote the coding scheme (contact details), when, and in which context?

[Wagner/Kallmeyer 2001] Andreas Wagner, Laura Kallmeyer: "Der TUSNELDA-Standard -- Ein Korpusannotierungsstandard zur Unterstuetzung linguistischer Forschung", in: Henning Lobin (Ed.): *Sprach- und Texttechnologie in digitalen Medien. Proceedings der GLDV Fruehjahrstagung 2001*. 28.

- 30. Maerz 2001, Giessen, (Giessen, Universitaet Giessen) 2001, 252 -- 262; also: http://www.uni-giessen.de/fb09/ascl/gldv2001

[Kallmeyer et al. 2001] Laura Kallmeyer, R. Meyer, Andreas Wagner: "Guidelines for the TUSNELDA Corpus Annotation Standard", to appear; also: http: //www.sfb441.uni-tuebingen.de/c1/tusnelda_guidelines.html

4.10.5Usage

Origin of the coding scheme and reasons for creating it

The TUSNELDA coding scheme has been developed within the Sonderforschungsbereich 441, at Universität T bingen, and it is being developed as an internal standard to a number of projects which deal with corpus-based linguistic research.

How many people have used the coding scheme and for what purposes?

No details about the number of users and the number of interaction annotated could be gathered so far. However, several projects of the above mentioned SFB are using TUSNELDA or will start using it, among others for transcribing spoken dialogues and for transcribing interactions in comics.

How many dialogues/interactions have been annotated using the coding scheme?

At the current stage, 12 comics (Serbian) and 2 video transcriptions (Portuguese) have been annotated (as well as a larger number of audio transcriptions, newspaper interviews, and other written texts). All these texts can be accessed via a query interface from the above mentioned web page.

Has the coding scheme been evaluated?

An evaluation of the coding scheme has not yet been published, and likely, it will only be available at a later point in time, when several projects will have used the coding scheme. The coding scheme in itself is not language dependent, because it basically covers interaction between speakers and gesture, and it leaves, very much in line with the idea of a metaschema, space for the inclusion of language specific linguistic annotation schemes.

Is the coding scheme language dependent (which language(s)) or language independent?

TUSNELDA is designed to be applicable to many languages, and it is being used on sample texts from Germanic, Romance, and Slavic languages. In the future, non-European languages (Tibetan, Warao) will be captured as well.

Is there tools support for using the coding scheme or API for editing/parsing coded descriptions? In which language?

Any XML-aware tool is usable for handling TUSNELDA. It is planned to adapt and/or develop annotation and query tools which take into account peculiarities of TUSNELDA, to provide user-friendly access to the data.

4.10.6Accessibility

How does one get access to the coding scheme?

The coding scheme and the annotated texts of TUSNELDA are accessible via the above mentioned web page.

Is the coding scheme available for free or how much does it cost?

The coding scheme is freely available.

4.10.7Conclusion

How well described is the coding scheme?

There is a coding manual that provides a detailed description of all parts of the annotation standard (http://www.sfb441.uni-tuebingen.de/cl/tusnelda-guidelines.html).

How general and useful is the coding scheme?

The developments of TUSNELDA is likely not yet as ripe yet as those described in the other chapters of this report. However, since some work on multimodal corpus creation is ongoing, it seemed useful to integrate a brief description of early developments, and of activities which have recently started. As XML, the CES and the work of the TEI are used in these activities, general compatibility of these new developments with the upcoming ISLE proposals is ensured, at least as long as an XML-based representation is followed.

4.11 General description of coding schemes for prosody, gestures and speech

4.11.1 Description header

Main actor

DFE: Joaquim Llisterri (Joaquim.Llisterri@uab.es) and María Jesús Machuca Ayuso (maria@liceu.uab.es).

Verifying actor

NISLab: Malene Wegener Knudsen (mwk@nis.sdu.dk), Laila Dybkjær (laila@nis.sdu.dk) and Niels Ole Bernsen (nob@nis.sdu.dk)

Date of last modification of the description

June 21st, 2001.

4.11.2 Reference

Resources web sites

Discourse Annotation Tools, Shared Tools and Resources, DRI (Discourse Resource Initiative): http://www.georgetown.edu/luperfoy/Discourse-Treebank/tools-and-resources.html

Gesture Annotation: Tools and Data, Linguistic Annotation, LDC (Linguistic Data Consortium): http://morph.ldc.upenn.edu/annotation/gesture/

Selection of documents, ISLE Metadata: http://www.mpi.nl/world/ISLE/documents/docs_frame.html

References to additional information on the coding scheme

Arndt, H., Janney, R.W.: Intergrammar: Toward an integrative model of verbal, prosodic and kinesic choices in speech. Berlin: Mouton de Gruyter, 1987.

Bauer, H.R.: Frequency Code: Orofacial Correlates of Fundamental Frequency. Phonetica, 44, pp. 173-191, 1987.

Bertrand, R., Boyer, J., Cavé, C., Guaïtella, I. and Santi, S.: Voice and gesture relations in interaction situations: some prosodic and kinesic aspects of back-channel. In Elenius, K. and Branderud, P. (Eds.): Proceedings of the XIIIth International Congress of Phonetic Sciences. Stockholm, Sweden, 13-19 August, 1995. Vol.2. pp. 746-749, 1995.

Bertrand, R. and Casolari, F.: Approche prosodique et pragmatique des modulations. In Actes des XXIes Journées d'Étude sur la Parole, Avignon. 1996.

Bolinger, D.: Intonation in American English. In Hirst, D. and Di Cristo, A. (Eds.): Intonation Systems. A Survey of Twenty Languages. Cambridge: Cambridge University Press. pp. 45-55, 1998.

Braffort, A., Gherbi, R., Gibet, S., Richardson, J. and Teil, D. (Eds.): Gesture-Based Communication in Human-Computer Interaction. In proceedings of the International Gesture Workshop, GW'99, Gif-sur-Yvette, France, March 17-19, 1999 Proceedings. Berlin: Springer Verlag (Lecture Notes in Computer Science, 1739), 1999.

Cassell, J.: Nudge Nudge Wink Wink: Elements of Face-to-Face Conversation for Embodied Conversational Agents. In Cassell, J. *et al.* (Eds.): Embodied Conversational Agents. Cambridge, MA:

MIT Press, 1999. Can be downloaded from:
http://gn.www.media.mit.edu/groups/gn/publications/ECA_Cassell.chapter.to_handout.pdf

Cassell, J., McNeill, D. and McCullough, K.E.: Speech-Gesture Mismatches: Evidence for One Underlying Representation of Linguistic and Non-Linguistic Information. Pragmatics and Cognition 7, 1, pp. 1-33, 1999. Can be downloaded from:
http://gn.www.media.mit.edu/groups/gn/publications/prag&cog_handout.pdf

Cassell, J., Pelachaud, C., Badler, N., Steddman, M., Achorn, B., Becket, T., Douville, B., Prevost, S. and Stone, M.: Animated Conversation: Rule-Based Generation of Facial Expression, Gesture and Spoken Intonation for Multiple Conversational Agents. In Proceedings of SIGGRAPH '94. (ACM Special Interest Group on Graphics), 1994b. Can be downloaded from:
http://gn.www.media.mit.edu/groups/gn/publications/siggraph94.pdf

Cassell, J. and Stone, M.: Living Hand to Mouth: Psychological Theories about Speech and Gesture in Interactive Dialogue Systems. In AAAI 1999 Fall Symposium on Narrative Intelligence, 1999. Can be downloaded from: http://gn.www.media.mit.edu/groups/gn/publications/cassell-stone_AAAI99.pdf

Cassell, J., Stone, M., Douville, B., Prevost, S., Achorn, B., Steedman, M., Badler, N. and Pelachaud, C.: Modeling the Interaction between Speech and Gesture: In Ram, A. and Eiselt, K. (Eds.): Proceedings of the Sixteenth Anuual Conference of the Cognitive Science Society. Lawrence Erlbaum Associates. pp. 153-158, 1994a. Can be downloaded from:
http://gn.www.media.mit.edu/groups/gn/publications/cogsci94.pdf

Cavé, C., Guaïtella, I., Bertrand, R., Santi, S., Harlay, F. and Espesser, R.: About the relationship between eyebrow movements and f0 variations. In ICSLP 96, The Fourth International Conference on Spoken Language Processing. October 3 - 6, Wyndham Franklin Plaza Hotel, Philadelphia, PA, USA. Vol 4. pp. 2175-2178, 1996. Can be downloaded from:
http://www.asel.udel.edu/icslp/cdrom/vol4/613/a613.pdf

Cavé, C., Guaïtella, I. and Santi, S.: Fréquence fondamentale et mouvements rapides des sourcils: une étude pilote. Travaux de l'Institut de Phonétique d'Aix 15, pp. 25-42, 1993.

Cruttenden, A.: Intonation. Cambridge: Cambridge University Press (Cambridge Textbooks in Linguistics), 1986.

Cutler, A., Dahan, D. and van Donselaar, W.: (1997): Prosody in the comprehension of spoken language: a literature review. Language and Speech 40, pp. 141-201, 1997.

Dybkjær, H. and Heid, U.: Towards annotated dialogue corpora. Final Report of the ELSNET Transition Phase Dialogue Annotation Action, 29 November 1996. Can be downloaded from: http://www.nis.sdu.dk/publications/papers/elsnet-da-96/elsnet-da-96.html

Guaïtella, I.: Interaction entre l'activité gestuelle et l'activité vocale dans la communication: éléments théoriques et méthodologiques. In Actes des XXèmes Journées d'Etude sur la Parole. Trégastel, 1994.

Guaïtella, I.: Étude des relations entre geste et prosodie à travers leurs fonctions rythmique et symbolique. In Actes du XIIème Congrès International des Sciences Phonétiques. 19-24 août 1991, Aix-en-Provence, France. Aix-en-Provence: Université de Provence, Service des Publications. Vol. 3. pp. 266-269, 1991.

Guaïtella, I., Cavé C. and Santi, S.: Relations entre geste et voix: le cas des sourcils et de la fréquence fondamentale. In Actes du Colloque Images et Langages, Multimodalité et Modélisation Cognitive. Paris. pp. 261-268, 1993.

Guaïtella, I. and Santi, S.: Pragmatic approach to the kinesic and prosodic modes of communication. In 1990 International Pragmatics Conference. Abstracts. Barcelona. International Pragmatics Association - Universitat Autònoma de Barcelona - Universitat de Barcelona, 1990.

Kaneko, T. and Ishizaki, S.: The Multi-Modal Dialogue Corpus. From The Second International Conference on Cognitive Sciences and the 16th Annual Meeting of the Japanese Cognitive Science Society, 27-30 July 1999, Tokyo, Japan, 1999. Can be downloaded from: http://www.sccs.chukyo-u.ac.jp/ICCS/olp/p3-14/p3-14.htm

Kendon, A.: Movement coordination in social interaction: some examples described. In Weitz (Ed.): Nonverbal Communication: Readings with Commentary. Oxford: Oxford University Press, 1974.

Magnuson, J. S., Dahan, D., Allopenna, P. D., Tanenhaus, M. K. and Aslin, R. N.: Using an artificial lexicon and eye movements to examine the development and microstructure of lexical dynamics. In Gernbacher, M. A. and Derry, S.J. (Eds.): Proceedings of the Twentieth Annual Conference of the Cognitive Science Society. Mahwah, NJ: Erlbaum. pp. 651-665, 1998.

McNeill D.: Hand and Mind: What gestures reveal about thought. Chicago: University of Chicago Press. 1992.

Pelachaud, C. and Prevost, S.: Sight and Sound: Generating Facial Expressions and Spoken Intonation from Context. In Conference Proceedings of the Second ESCA/IEEE Workshop on Speech Synthesis. September 12-15, 1994. Mohonk Mountain House, New Paltz, New York, USA. pp. 216-129, 1994. Can be downloaded from: http://www.fxpal.xerox.com/people/prevost/pdf%20papers/esca.pdf

Poyatos, F.: Nonverbal communication and translation. Amsterdam: John Benjamins Publishing Co., 1997.

Prevost, S.: Contextual Aspects of Prosody in Monologue Generation. In IJCAI Workshop on Context in Natural Language Processing. Montreal, 1995. Can be downloaded from: http://www.fxpal.xerox.com/people/prevost/pdf%20papers/ijcai-context95.pdf

Purson, A., Santi, S., Bertrand, R., Guaïtella, I., Boyer, J. and Cavé, C.: The Relationships between Voice and Gesture: Eyebrows Movements and Questioning. In Eurospeech'99, 6th European Conference on Speech Communication and Technology. September 5-9, 1999, Budapest, Hungary, 1999.

Quazza, S. and Garrido, J.M.: Prosody. In Klein, M. (Ed.): Supported Coding Schemes. MATE Deliverable D1.1. LE Telematics Project LE4 – 8370. July 1998. Can be downloaded from: http://technovoice.cselt.it/voce/chap6/chap6.html
http://www.ims.uni-stuttgart.de/projekte/mate/mdag/pd/pd_1.html

Santi, S., Guaïtella, I., Cavé, C. and Konopczynski, G. (Eds): ORAGE'98, ORAlité et Gestualité: communication multimodale, interaction. Paris, L'Harmattan, 1998.

Steininger, S.: Transliteration of Language and Labeling of Emotion and Gestures in SmartKom. In Workshop Proceedings of the Second International Conference on Language Resources and Evaluation: Meta-Descriptions and Annotation Schemes for Multimodal/Multimedia Language Resources. Athens, Greece. pp. 49-51, 2000. Can be downloaded from: http://www.phonetik.uni-muenchen.de/Forschung/Publications/steininger_ISLE_00.ps;
http://www.mpi.nl/world/ISLE/documents/papers/Steininger_paper.pdf

Thompson, L.A. and Massaro, D.W.: Evaluation and integration of speech and pointing gestures during referential understanding. Journal of Experimental Child Psychology, 42, pp. 144-168, 1986.

4.11.3 Coverage

According to Poyatos (1997), speech can be defined as a triple audiovisual structure made up basically of words, paralanguage and kinesics. Gesture and speech would arise together from an underlying representation with two different modalities: visual and linguistic; therefore, the relationship between gesture and speech is essential to the production of meaning and to its comprehension (McNeill, 1992; Cassell et al., 1994a).

It has been shown that when speech is ambiguous or in a speech situation with some noise, listeners do rely on gesture cues (Thompson and Massaro, 1986). Cassell et al. (1999) also mention the fact that listeners attend to information conveyed in gesture when that information supplements or even contradicts the information conveyed by speech. Due to the evidence of synchronization between gesture and speech, prosodic information in the transcriptions of speech corpora can become very relevant to determine the function of some types of gesture.

Types and functions of gestures in relation to speech

Although the aim of this report is not a typology of gestures, it seems interesting to mention that in some classifications reference to prosodic aspects can be found. As far as movements of the head and facial expressions are concerned, when discussing the syntactic function of those gestures, Cassell et

al. (1994b) remark that raising the eyebrows, nodding the head or blinking can appear on accented syllables or in a pause. Moreover, a type of hand gesture known as beats (small formless waves of the hand) is described as occurring simultaneously with emphasized words (Cassell et al., 1994b).

Prosodic parameters and gestures

A description of the relationship between prosodic parameters and gestures may take into account not only the type of gestures, but also the different phases in their implementation.

As far as the phases of gestures are concerned, three parts can be distinguished: the preparation of the gesture, the most energetic part of it and the relaxation of the gesture.

According to Cassell et al. (1994b), these parts are related to intonational phrases (i.e. boundary tones or the points in the utterance where prosodic boundaries do occur). The preparation starts just before or just at the beginning of the intonational phrase and finishes just before the next gesture in the intonational phrase or the nuclear stress of the phrase. The stroke phase of the gestures tends to co-occur with (or just before) the phonologically most prominent syllable in the utterance or with the nuclear stress (Kendon 1974; McNeill 1992; Cassell et al. 1994b). The relaxation occurs by the end of the intonational phrase.

Prosodic parameters related with types of gesture are shown in Table I, summarizing findings from Cruttenden (1986), Guaïtella (1991), Cavé et al. (1993), Bertrand et al. (1995) and Cassell et al. (1994 a, b).

In general, rising tones and related gestures involve an increase in tension, whereas falling tones and related gestures involve a decrease in tension (Cruttenden, 1986).

<table>
<tr><td>Gesture parameters</td><td>Prosodic parameters</td><td></td></tr>
<tr><td>Head movements</td><td>F0 variation</td><td rowspan="2">The activity of gesture parameters is related to intonation.
A decrease of fundamental frequency (F0) and intensity is related with a change of direction of the look.</td></tr>
<tr><td>Eye movements</td><td>Syllabic segmentation</td></tr>
<tr><td>Eyebrow movements</td><td>F0 variation</td><td rowspan="2">There is a systematic change of F0 movements related with a fast movement of the eyebrows.
A correlation between the range of the F0 curves and the amplitude of eyebrow movements can be established.</td></tr>
<tr><td>Hand movements</td><td>Intensity</td></tr>
</table>

Figure 4.11.1. A description of the relation between gestural and prosodic parameters.

Nevertheless, relation between prosody and gesture seems to be language dependent. Bolinger (1998), for example, emphasizes that differences between American English and British intonation are not in the configurations of the fundamental frequency (F0) contour but in frequency and pragmatic choice. As a consequence, Bolinger suggests that American English intonation should be studied in relation to the entire American gesture setting. For example, a higher pitch is associated with higher positions of the eyebrows, shoulders and often hands and arms.

Proposals to annotate prosodic information related to gestural information

A proposal for prosodic annotation has been put forward in part 2.1.1 of ISLE deliverable D11.2: Requirements Specification for a Tool in Support of Annotation of Natural Interaction and Multimodal

Data, including four levels of annotation: prosodic units, prosodic phenomena, phonetic correlates of prosodic phenomena and linguistic phenomena.

Work on the relationship between gesture and prosodic parameters reviewed in this report suggest that the following elements can be related to gesture: F0 movements and shape of the F0 contour, F0 prominences, F0 range, boundary tones or intonational phrases, (nuclear) sentence stress, intensity changes, emphasis and pauses. All of them can be mapped with elements included the annotation proposal put forward in D11.2, as summarized in Figure 1.6.2.:

Elements in the proposed annotation scheme (D11.2)	Prosodic elements related to gesture as mentioned in the reviewed literature	Gesture information as mentioned in the reviewed literature
Prosodic units		
Syllable		Eye movements
Intonation group	Intonational phrases	Gesture preparation Gesture relaxation
Breath group	Pauses	Raising of the eyebrows Blinking Nodding of the head
Sentence	Sentence stress	Stroke phase of the gesture
Prosodic phenomena		
Pitch accent	Syllable stress F0 prominences	Raising of the eyebrows Blinking Nodding of the head
Boundary tones	Boundary tones	Gesture preparation Gesture relaxation
Sentence stress	(Nuclear) sentence stress	Stroke phase of the gesture
Emphatic stress	Emphasis	Hand movements
Prosodic correlates		
F0 values	Fo prominences	Head movements Eyebrow movements Hand movements
F0 maximum and minimum	F0 range	Eyebrow movements
F0 movements	F0 movements and shape of the F0 contour	Head movements Eye movements Eyebrow movements
Intensity	Intensity changes	Eye movements Hand movements
Pauses	Pauses	Raising of the eyebrows Blinking Nodding of the head

Figure 4.11.2. Mapping of the prosodic annotation scheme proposed in D11.2 and prosodic correlates of gesture found in the reviewed literature.

Rather than defining a new scheme for annotation of prosody in relation to gesture information, it seems that using the suggested proposal for prosody – which is based on the practices found in the most widespread annotation systems as reviewed in Quazza and Garrido (1998) – together with the proposal for gesture annotation and an adequate synchronisation between both types of annotation would be sufficient to allow the study of the interaction between prosody and gesture.

5 Lesser Known/Used Gesture Coding Schemes

5.1 LIMSI TYCOON scheme for analysing cooperation between modality

5.1.1 Description header

Main actor

LIMSI-CNRS : Jean-Claude MARTIN (martin@limsi.fr)

Date of last modification of the description

9 August 2001

5.1.2 References

Web site

http://www.limsi.fr/Individu/martin/

Martin, J.C., Grimard, S., Alexandri, K. (2001) On the annotation of the multimodal behaviour and computation of cooperation between modalities. *Proceedings of the workshop on «Representing, Annotating, and Evaluating Non-Verbal and Verbal Communicative Acts to Achieve Contextual Embodied Agents»,* May 29, 2001, Montreal, in conjunction with The Fifth International Conference on Autonomous Agents. pp 1-7 http://aos2.uniba.it:8080/aa-ws.html

Illustrative sample picture or video file

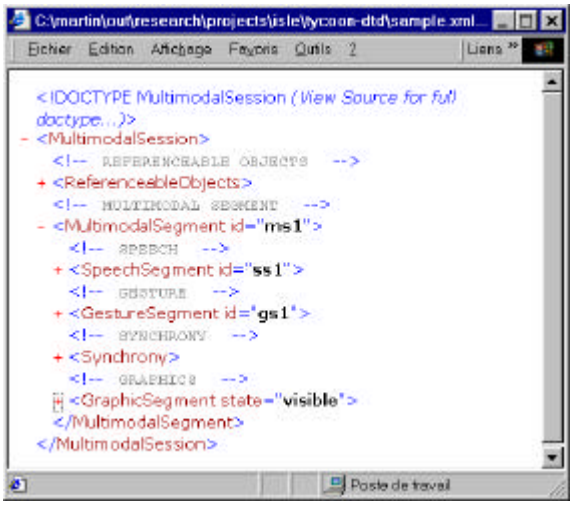


Figure 5.1.1. Example of the XML annotation of a sample command observed in the SRI corpus [1].

Figure 5.1.2. The "referenceable objects" section of a multimodal annotation.

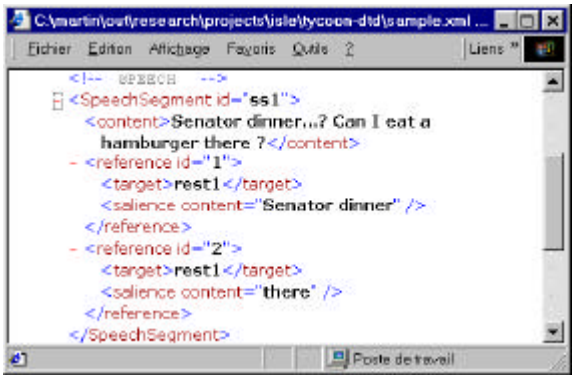


Figure 5.1.3. A speech segment (*"Senator dinner ... ? can I eat a hamburger there ?"* which contains two references.

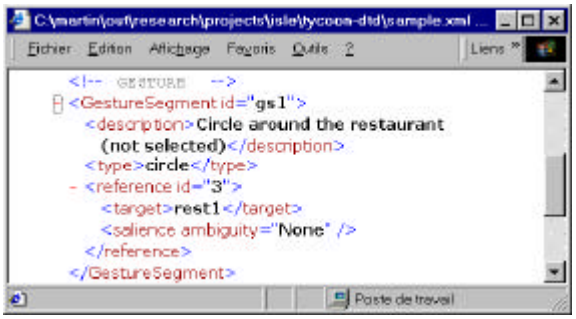


Figure 5.1.4. A gesture segment including a reference to the object *rest1*.

5.1.3 Description

Our goal is to ease the computation of metrics of multimodal behaviour from video corpora. The metrics we are interested in are based on the TYCOON theoretical framework for studying multimodality. Thus, the coding scheme we propose features the annotation of available referenceable objects and the annotation of references to such objects in each modality. Pieces of information enabling the computation of salience values associated to referred objects are also included in the coding scheme.

The logical structure of the coding scheme is defined as follow:

A corpus of multimodal behaviour is annotated as a multimodal session

A multimodal session includes one referable objects section and one or more multimodal segments

A *multimodal segment* is made of a speech segment, a gesture segment, the annotation of temporal relation between these two segments and a graphics segment

We have implemented this coding scheme as a Document Type Definition (DTD) for defining the generic structure of multimodal behaviour annotations. Such annotations are done in the eXtensible Markup Language (XML). We will take the example of the XML annotation of a sample multimodal command observed in the SRI corpus (Cheyer at al. 1998). Such an annotation is composed of a *ReferenceableObjects* section describing the graphical objects the user is able to refer to, and a *MultimodalSegment* section composed of four sub-sections: speech, gesture, synchrony, and graphics (Figure 1). The first section contains annotation about the referable objects the user may refer to such as restaurants, hotels (Figure 2). This section about the referable objects is followed by one or several multimodal segment sections. Each multimodal segment section may contain annotations about speech, gesture, synchrony or the state of the graphics modality. Both speech and gesture annotations may contain annotation of references to objects (Figures 3 and 4).

The TYCOON-DTD has already been applied to the annotation of 40 multimodal segments coming from 5 different corpora.

5.2 W3C Working Draft on Multimodal Requirements for Voice Markup Languages

5.2.1 Description header

Main actor

LIMSI-CNRS : Jean-Claude MARTIN (martin@limsi.fr)

Date of last modification of the description

9 August 2001 (draft is dated July 2000)

5.2.2 References

Web site

http://www.w3.org/TR/multimodal-reqs

Illustrative sample picture or video file

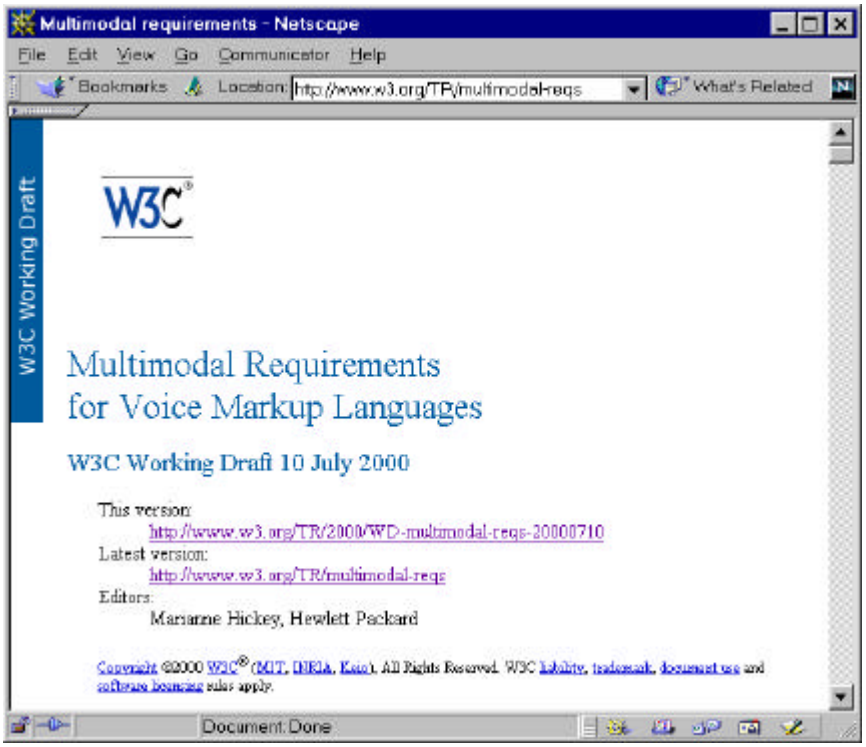


Figure 5.2.1. W3C web page on a preliminar study about the integration of multimodality in voice XML.

5.2.3 Description

This description is neither about a coding scheme or about annotation of observed multimodal behaviour. Instead it is a working document about the requirement that multimodal markup language for web browsing application should follow. Still it is about coding of multiple modalities, and it is from the W3C consortium. As such I think it deserves a short description.

Multimodal browsers allow users to interact via a combination of modalities, for instance, speech recognition and synthesis, displays, keypads and pointing devices. The Voice Browser working group is interested in adding multimodal capabilities to voice browsers. This document sets out a prioritised list of requirements for multimodal dialog interaction, which any proposed markup language (or extension thereof) should address. The focus is on multimodal dialog where there is a small screen and keypad (e.g. a cell phone) or a small screen, keypad and pointing device (e.g. a palm computer with cellular connection to the Web).

The suggested requirements are classified as follow:

General requirements: Scalable across end user devices, Complimentary use of modalities, Seamless synchronization of the various modalities,

Input modality requirements: Audio Modality Input, Sequential multi-modal Input, Uncoordinated Simultaneous Multi-modal Input, Coordinated, Simultaneous Multi-modal Input, Input modes supported, Extensible to new input media types, Semantics of input generated by UI components other than speech, Modality-independent representation of the meaning of user input, Coordinate speech grammar with grammar for other input modalities, Time window for coordinated multimodal input, Composite meaning, Support for conflicting input from different modalities, Context for recogniser, Time stamping,

Output media requirements: Audio Media Output, Sequential multimedia output, Uncoordinated Simultaneous Multi-media Output, Coordinated Simultaneous Multi-media Output, Synchronization of multimedia with voice input, Temporal semantics for synchronization of voice input and output with multimedia, Visual output of text, Media objects supported by SMIL, Media-independent representation of the meaning of output, Time stamping

Architecture, Integration and Synchronization points: Reuse standard markup languages, Mesh with modular architecture proposed for XHTML, Detect that a given modality is available, Means to act on a notification that a modality has become available/unavailable, Synchronization points, Interaction with External Components.

5.3 The New England Regional Leadership Non-Verbal Coding scheme

5.3.1 Description header

Main actor

LIMSI-CNRS : Jean-Claude MARTIN (martin@limsi.fr)

Date of last modification of the description

Tuesday, 09 October 2001

5.3.2 References

Web site

http://crs.uvm.edu/gopher/nerl/personal/comm/f.html

Illustrative sample

Non-Verbal Expression : A Checklist of Behavior

> Eye Contact
l. Spontaneous eye contact and eye movement
2. Breaking eye contact
3. Staring too intensely
4. Looking down
5. Looking directly at speaker when speaking
6. Looking directly at speaker when listening
7. Looking away
8. Staring blankly

> Body Posture
l. Slight forward lean
2. Body facing speaker
3. Relaxed posture
4. Relaxed hand position
5. Spontaneous hand and arm movements
6. Gestures for emphasis
7. Touching speaker
8. Relaxed leg position
9. Slouching
10. Fixed, rigid position
11. Physically too close to speaker
12. Physically distant from speaker
13. Arms across chest

14. Body turned sideways

> Head and Facial Movements
l. Affirmative head nods
2. Calm, expressive facial movements
3. Appropriate smiling
4. Expressions matching speaker mood
5. Face rigid
6. Continual nodding
7. Extraneous facial movements
8. Continual smiling
9. Little smiling
10. Cold, distant expression
11. Frowning
12. Overly-emotional reactions

> Vocal Quality
1. Pleasant intonation
2. Appropriate loudness
3. Moderate rate of speech
4. Simple, precise language
5. Fluid speech
6. Monotone
7. Too much effort
8. Too loud

> Distracting Personal Habits
l. Playing with hair
2. Fiddling with pen or pencil
3. Chewing gum
4. Smoking
5. Drinking
6. Tapping fingers or feet
7. Other

5.3.3 Description

This coding scheme looks rather simple and there is no references to any application to a real video corpus. It looks more like a tutorial on communication for leaders. Yet one originality is the “Distracting Personal Habits” section.

6 Practices and best practice

There probably exists a wealth of NIMM annotation schemes out there, far more than those which are described in the present report. Most of them are tailored to a particular purpose and used solely by their creators or at the creators' site. Such coding schemes tend not to be very well described and they tend to be hard to find. This survey itself includes a number of such coding schemes, many of which were created by ISLE NIMM participants or by people known to ISLE NIMM participants, this being the main reason why we were aware of them. Other coding schemes included above are fairly general ones, in frequent use, or even considered standards in their field (cf. below).

Nearly all the reviewed coding schemes are aimed at markup of video, possibly including audio. A couple of schemes are used for static image markup.

The collected material comprises schemes for markup of a single modality as well as schemes for markup of modality combinations. Figure 6.1 provides an overview of the coding schemes reviewed, including the annotation purpose for which they were created. Figure 6.1 leaves out the four unmarked general descriptions in Figure 1.3.1, as these cannot be characterised as coding schemes proper.

In most cases, a coding scheme was created originally because a person or a site had a particular need for annotated data, e.g. related to systems development. Sections 6.1. and 6.2 briefly summarise findings on facial and gesture coding schemes, respectively. Section 6.3 summarises findings on coding schemes evaluation and tools support. Finally, Section 6.4 concludes on the current state of the art as regards coding schemes and tools support.

6.1 Facial coding schemes

We only found about half as many facial coding schemes as gesture coding schemes. One reason may be is that there is a few facial coding schemes which are actually being used by a relatively large number of people. Thus, MPEG-4 is considered a standard and is being widely used. FACS is used by many people as well but is not really well suited for markup of lip movements. ToonFace is good for 2D caricature coding but not for real facial expression annotation. Other reviewed facial expression schemes seem to have been used only by a single person or by a few people.

The facial coding schemes are all language independent and they all focus entirely on facial expression. However, within facial expression they cover a multitude of different features including, e.g., gaze, eye brows, eye lids, wrinkles, and lips. A couple of schemes have baby or children's faces as their target. Most of the schemes focus on adult faces.

6.2 Gesture coding schemes

In the area of gesture, the picture seems more diverse than for facial expression. Where facial expression is often the sole focus point, gesture often seems to be studied along with other modalities, each modality being coded separately. It is only in the field of sign languages that the schemes we looked at focused on gesture alone. Many other gesture schemes were created to study gesture in combination with one or several other modalities with the purpose of supporting the development of a multimodal system. When several modalities are involved, it becomes important to be able to handle interrelationships among phenomena expressed in different modalities. Time alignment would seem basic as the common point of reference for this purpose.

The tag sets used in some of the analysed coding schemes for sign languages are merely symbolic ones. This may be because sign languages themselves convey information coded in the form of symbols which themselves are abbreviated notations for gestures. This is contrary to spoken dialogue annotation where textual annotations of gestures are preferred. The tag sets must be reliable and based on a well-defined vocabulary for describing the gestures covered by the coding scheme. The vocabulary may be organised hierarchically according to, e.g., shapes, positions, or movements of body parts.

Several of the analysed gesture coding schemes are meta-schemes in the sense that they are general information containers (meta-schemes) that can be filled with a concrete coding scheme (e.g. for a particular sign language). Often they also define a way in which to link gesture with other modalities.

There is a trend towards using XML across the gesture schemes analysed. The creators of less recent schemes tend to announce plans or ongoing projects for conversion into XML (HIAT, HamNoSys: ViSiCAST), and the more recent ones (e.g. SWML, TUSNELDA) were designed from the start for being represented in XML. XML appears to be a well-suited representation format for the targeted kinds of data. Among other things, it allows for synchronisation of tracks in an easy and efficient way.

Intended for markup of	**Name**	**Purpose of creation**
Gaze	The alphabet of eyes	Analyse any single item of gaze in videotaped data.
Facial expression	FACS (facial action coding system)	Encode facial expressions by breaking them down into component movements of individual facial muscles (Action Units). Suitable for video or image.
	BABYFACS	Based on FACS but tailored to infants.
	MAX (Maximally Discriminative Facial Movement Coding System)	Measure emotion signals in the facial behaviours of infants and young children. Suitable for video or image.
	MPEG-4	Define a set of parameters for defining and controlling facial models.
	ToonFace	Code facial expression with limited detail. Developed for easy creation of 2D synthetic agents.
Gesture	HamNoSys	Designed as a transcription scheme for various sign languages.
	SWML (SignWriting Markup Language)	Code utterances in sign languages written in the SignWriting System.
	MPI GesturePhone	Transcription of signs and gestures.
	MPI Movement Phase Coding Scheme	Coding of co-speech gestures and signs.
Speech and gesture	DIME (Multimodal extension of DAMSL)	Code multimodal behaviour (speech and mouse) observed in simulated sessions in order to specify a multimodal information system.
	HIAT (Halbinterpretative Arbeitstranskriptionen)	Describe and annotate parallel tracks of verbal and non-verbal (e.g. gesture) communication in a simple way.
	TYCOON	Annotation of available referable objects and references to such objects in each modality.
Text and gesture	TUSNELDA	Annotation of text-and-image-sequences, e.g. from comic strips.
Speech, gesture, gaze	LIMSI Coding Scheme for Multimodal Dialogues between Car Driver and Co-pilot	Annotation of a resource which contains multimodal dialogues between drivers and co-pilots during real car driving tasks. Speech, hand gesture, head gesture, gaze.
Speech, gesture and body movement	MPML (A Multimodal Presentation Markup Language with Character Agent Control Functions)	Allow users to encode the voice and animation of an agent guiding a web site visitor through a web site.
Speech, gesture, facial expression	SmartKom Coding scheme	Provide information about the intentional information contained in a gesture.

Figure 6.1. A number of reviewed coding schemes and their purposes.

There are as yet no real standards for gesture markup. HamNoSys seems to be the most frequently used among the schemes we looked at as regards gesture annotation-only. For gesture in combination with other modalities there are many schemes – mostly used by few people - but no standardisation.

The picture provided by the survey of a proliferation of home-grown coding schemes is supported by the 28 questionnaires included in ISLE deliverable 8.1, asking people at a multimodal interaction workshop, e.g., which coding scheme(s) they had used or planned to use for data markup. Some respondents did not answer the question at all or had not made any decision yet. However, in 15 cases the answer indicated that a custom-made scheme would be, or was being, used. Only a few respondents also mentioned more frequently used annotation schemes such as TEI, BAS, or HamNoSys.

6.3 Evaluation of coding schemes, tool support for coding schemes

With respect to coding scheme evaluation, the pattern observed for the two groups of facial and gesture coding schemes, respectively, is again pretty different. Apart from ToonFace and the lesser known schemes for which we do not have any information on this point, the facial coding schemes have all been evaluated. ToonFace is being used by several people who report on ease of use but that is all we have found as regards evaluation of this particular scheme. Only a couple of the gesture coding schemes, again not counting the lesser known ones for which we do not have any information on evaluation, have been evaluated and only on a small scale. All the other gesture coding schemes have either not been evaluated or no information on evaluation has been found. This confirms the conclusion that gesture coding schemes have a longer way to go before standardisation can be achieved than is the case for the facial coding schemes.

Only when it comes to tool support is there no significant difference between the two groups of schemes. In most of the cases surveyed there is some kind of tools support for using a coding scheme or for processing the results of using that coding schemes. Only in three cases is there is no tools support at all. With respect to the lesser known coding schemes we do not have information on any available tools.

6.4 Conclusion - still a long way to go

It may safely be concluded from the present survey that there is still a very long way to go before we will be able to code natural interactive communication and multimodal information exchange in all their forms, at any relevant level of detail, generally or exhaustively per level, and in all their cross-level and cross-modality forms. This is already true for the coding of speech at several important levels of abstraction, such as dialogue acts and co-reference, as concluded in the MATE report on multi-level annotation of spoken dialogue, cf. MATE deliverable D1.1 which is available at the MATE web site at mate.nis.sdu.dk. When we move to considering facial coding, we do find a number of general and substantially evaluated coding schemes for different aspects of the facial expression of information (eyes, facial muscles), but it seems clear that we still need a number of higher-level coding schemes based on solid science for how the face manages to express cognitive properties, such as emotions, purposes, attitudes and character. In the general field of gesture, the state of the art is even further from the ideal described above. General coding schemes, as opposed to schemes designed for the study of particular kinds of task-dependent gesture, are hard to find at all, except for the special field of sign languages, and the state of evaluation of particular schemes is generally poor. Finally, when it comes to cross-modality coding, no coding scheme of a general nature would seem to exist at all.

A key to progress, it would seem, is the availability of general-purpose coding tools for natural interactive and multimodal behaviour. Such tools do not yet exist, as shown in the ISLE coding tools state of the art report D11.1, but their existence could mean a breakthrough in the scientific study of how humans express information through intriguingly complex and massively coordinated use of multiple modalities and at multiple levels of abstraction within each modality involved.

Acknowledgements

We gratefully acknowledge the support of the ISLE project by the European Commission's HLT Programme. We would also like to thank coding schemes creators for their willingness to make descriptions of their annotation schemes publicly available in this report as well as for the time they have given in communicating with the European ISLE NIMM team.