Measuring and Reconstructing Pointing in Visual Contexts

Alfred Kranstedt B3, Technology
Andy Lücking B3, Linguistics
Thies Pfeiffer C3, Technology
Hannes Rieser B3, Linguistics
Marc Staudacher B3, Linguistics
CRC 360 "Situated Artificial Communicators", Bielefeld University
alfred.kranstedt@googlemail.com
{andy.luecking|thies.pfeiffer|hannes.rieser|marc.staudacher}@uni-bielefeld.de

Abstract

We describe an experiment to gather original data on geometrical aspects of pointing. In particular, we are focusing upon the concept of the pointing cone, a geometrical model of a pointing's extension. In our setting we employed methodological and technical procedures of a new type to integrate data from annotations as well as from tracker recordings. We combined exact information on position and orientation with rater's classifications. Our first results seem to challenge classical linguistic and philosophical theories of demonstration in that they advise to separate pointings from reference.

1 Background

Dealing with pointing as a linguistic device implies dealing with two poles: On the one hand, pointing is bound up with reference. On the other hand, pointing is not precise.[1] Sources for the first pole can be found in philosophical literature, the second pole is supported by psychological research. Wittgenstein (1958, Blue Book, p. 50) gives away the philosophers' private detail that he "may know where a thing is and then point to it by virtue of that knowledge." Butterworth (2003, p. 25) sums up psycholinguistic investigation in stating that pointing "did not allow precise target localization."[2] Obviously, both positions do not fit together. The commonsense view that we can demonstrate objects seems to conflict with the fuzziness of vector extrapolation between index finger and target. Some years ago we started to hypothesize that the "blur" of pointings can be systematically couched in the geometrical concept of the pointing cone (Kranstedt et al., 2006a), and thereby deliver a model of a pointing's extension. This promises to be useful in both linguistics and artificial intelligence – see (Kranstedt et al., 2006b) for an overview. However, camera-based studies that aimed at delimiting the cone's apex angle suffered from the drawback that two-dimensional video data were too poor to derive exact three-dimensional topologies from. To overcome such limitations we pursue an original methodological approach employing audio, video, and body movement recordings simultaneously in a restricted, task-oriented object identification game setting and augmenting them with human annotation. We present some results gained by the empirical study (Section 2) in Section 3. The results play a prominent role in shaping the subsequent outlay of theorizing in Section 4.

2 Empirical Study

The empirical study involves two participants engaged in a restricted object identification game. This task was derived from earlier studies on the use of pointing gestures in referring (Lücking et al., 2004). Each participant gets a certain role, one is called *Description Giver* (henceforth DG) and the other *Object Identifier* (OI). DG and OI are placed in a CAVE-like environment which incorporates a marker-based optical tracking system with nine cameras (6DOF tracker). The information delivered by the cameras is integrated *via* special software and provides points and orientations in an absolute coordinate system, which origin lies in the center of the CAVE-like environment. We tracked the DG only. He sits on a stool and is equipped with carefully positioned

[1] We restrict ourselves to concrete pointings here. See (McNeill, 1992) for abstract pointings.

[2] See also (Butterworth and Itakura, 2000).

markers for the tracking system measuring arm, index finger, hand, and head movements. It is clocked by a frame (1/25 sec.) so that longer movements deliver more tracking data. In addition, the whole scene is recorded from two different perspectives with digital cameras. Speech is captured with the DG's headset. The whole set-up with the prepared DG can be seen in Figure 1, a screenshot from our video recordings. The special gloves used to track the stretched index finger are displayed in Figure 2. Both OI and DG

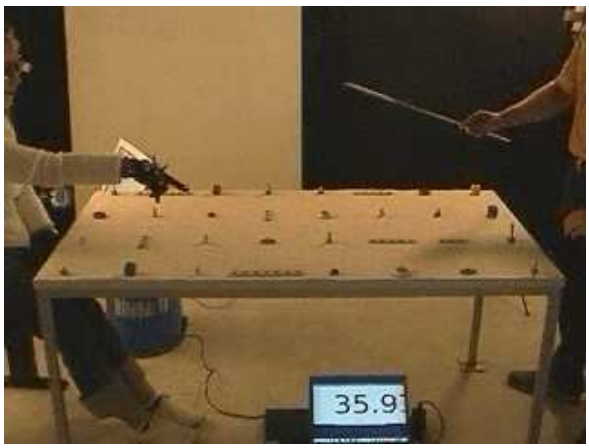


Figure 1: The experimental set-up: The DG sits to the left of the table, the OI stands to the right and has a pointer. The system time needed for synchronizing tracking and recording is displayed on a monitor.

are located around a real table (77.5 × 155.5 cm) with 32 parts of a Lorentz Baufix toy airplane, the experimental domain. The objects' centers were lined up randomly on an underlying grid ensuring that they are laid out equidistantly, see Figure 3. This layout is used for all trials of the study. The outer objects' centers frame an area of 70 × 140 cm. That is, the distance between objects' centers of the same column is 20 cm in neighbouring rows. To exemplify the mapping from rows to distance measures: The distance of the third row from the left, DG's, side of the table is 47.75 cm (2 × 20 cm + 7.75 cm for the outer margin).

Figure 2: Special gloves

2.1 The Realization of the Experiments

The identification game gets instantiated in two variations, differing in the communicative channels (speech and gesture) the DG is allowed to use:

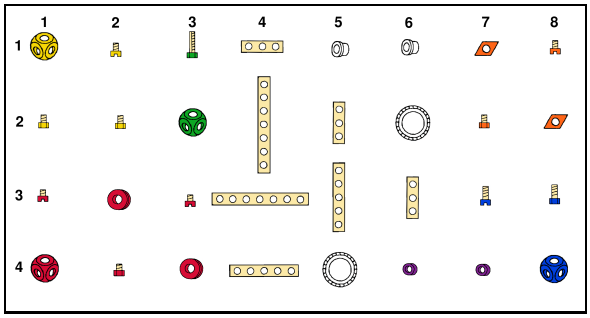


Figure 3: The experimental domain is divided up into eight rows and four columns. It covers an area of 70 × 140 cm. The DG is positioned to the left, the OI to the right of the domain.

- speech plus gesture (S+G Trial);
- gesture only (G Trial).

In each subsetting the DG has to get the object of each of the 32 identification games from the display on the monitor (roughly) in front of him. The order of the objects has been fixed in a presetting. In order to abstract over potential sequence effects, different object presettings have been randomly generated which are iterated over the subsettings and over the whole experimental runs.

The flowchart of the interaction. The interaction between DG and OI is highly restricted to avoid uncontrollable negotiation processes between the interactants. It consists of three formalized steps:

1. Demonstration by DG (bimodal or gestural, according to current subsetting);
2. Interpretation and identification by OI with a pointer only (the referent remains in it's place);
3. Feedback by DG.
   The feedback is restricted to "Ja" (*yes*) in the successful case (accept) and to "Nein" (*no*) in the unsuccessful case (denial). In both cases the identification game terminates and the participants move on, starting with the DG selecting the next object from his display.

2.2 Annotation

46 of the recorded experimental subsettings, 23 with and 23 without speech, enter into analysis. That makes a total of 1472 (46 × 32) demonstrations.

Annotation of the video data has been carried out making use of two software tools, Anvil and Praat. The audio tool Praat[3] was used for the transcription of spoken language, the video films were annotated with the multimodal annotation tool Anvil[4]. Since the concern of the study is pointing, annotation is restricted to DG's first move, that is, to the demonstration act. Annotation is done on several layers (of course, annotating speech is restricted to the S+G Trials):

gesture.phase [*preparation*, *stroke*, *retraction*]; structuring gesture motion according to the trinity established by (McNeill, 1992).

gesture.handedness [*left*, *right*]; for two-handed gestures both values are specified simultaneously.

speech.transcription DG's speech transcribed at the level of words.

speech.number The number of words used in DG's move.

speech.quality [*shape*, *color*, *function*, *position*, *proxy*]; "semantic categories" that are referred to in an utterance (the last one labels taxonomically unspecified nouns, NPs or determiners, like "Ding" (*thing*) or "Das" (*that*) or "Dies Teil" (*this thing*)).

move.referent unique name of object.

move.success [*yes*, α], if the OI could successfully identify the object. Name α of erroneously chosen object otherwise.

Our research interest is the precision of pointing – operationalized in terms of the pointing cone. Accordingly, only those gesture tokens enter into analysis which are purely deictic (showing, e. g., no iconic traits). Furthermore, the success (or failure) of a move should depend on exactly one gesture. We implement this two-step filter in annotation layers, on which annotators have to make suitable decisions:

gesture.validity [*yes*, *no*]; is the gesture a purely deictic one?

move.validity [*yes*, *no*]; Is the game's gesture valid and does the gesture include exactly one stroke?

[3] `http://www.praat.org/`

[4] `http://www.dfki.de/~kipp/anvil/`

As a preliminary test procedure for the reliability of the annotation scheme the interrater-agreement between three raters' annotations of one video on the most versatile layers, namely `speech.quality` and `gesture.validity`, has been calculated. With a value of $AC_1 = 0.9$ for semantic categories and a value of $AC_1 = 0.85$ for gesture classification, both ratings prove to be quite consistent.[5]

2.3 Processing Tracking Data

The geometrical and temporal information assembled in the tracking data files is processed to deliver quantitative models of pointing. Since we have the orientation and the exact position of the DG's head ("cyclop's eye") and the exact position of the index finger as well as of the referred object, we are able to represent pointing beams as vectors. Based on careful qualitative observations of the subjects' pointing behavior, we assume two different yet plausible ways of anchoring and orienting a beam: Firstly, origin and orientation may be given exclusively by the index finger (*index finger pointing*, IFP); secondly, the beam can be anchored in the (tip of the) index finger, but the orientation is determined by projecting a beam from the cyclop's eye (point between the eyes of the DG) through the anchor (*gaze finger pointing* GFP). Thus GFP models the presumed influence of gaze on pointing in a strict way. The "true pointing vector" (if there is such a thing) probably is somewhere in the middle between the extremes defined by GFP and IFP and might be reconstructable by interpolating the two. Using our IADE (*Interactive Augemnted Data Explorer*) framework (Pfeiffer et al., 2006), a tool for recording, analysis and (re-)simulation of multimodal data, both pointing beams can be visualized in simulation videos, as shown in Figure 4. The (extreme) case shown exemplifies that both kinds of pointings can diverge a great deal. The picture also shows the idealized beam. Idealized beams are the straight lines connecting the pointing vector's anchor with the point in space inhabited by the object referred to. Comparing the GFP and IFP beams with their ideal counterparts delivers a measure of pointings' "faultiness". As error estimates we employed two gauges, angular and orthogonal deviation. *Prima facie*, angular de-

[5] Identity of ratings cannot be ascribed to chance on a risk level of $\alpha = 0.01$. AC_1 is the first order agreement coefficient developed in (Gwet, 2001). Most of the other layers have been evaluated extensively in a precursor study.

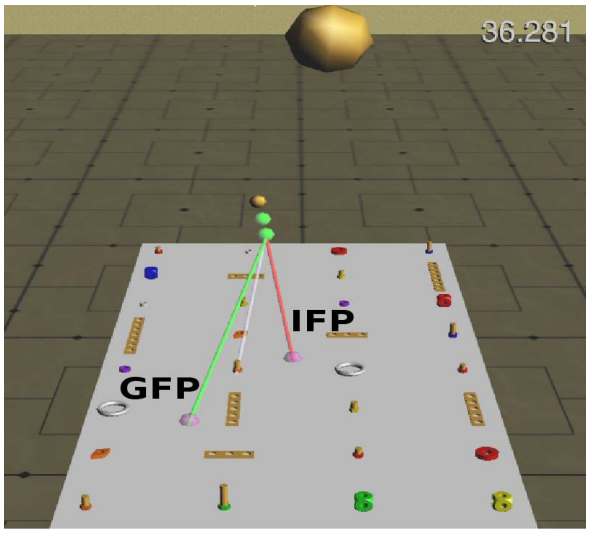


Figure 4: Simulating IFP, GFP, and the idealized pointing beam in between.

viation is more suitable since angles are distance-independent. Angular deviation is calculated as the angle γ spanning between the simulated and the ideal pointing vector. A schematic depiction is given in Figure 5. However, given short distances between anchors and objects, even small variances result in a high angular deviation. As a compara-

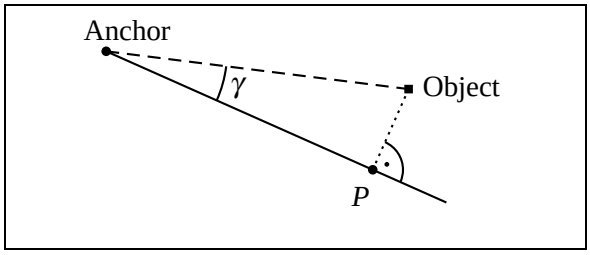
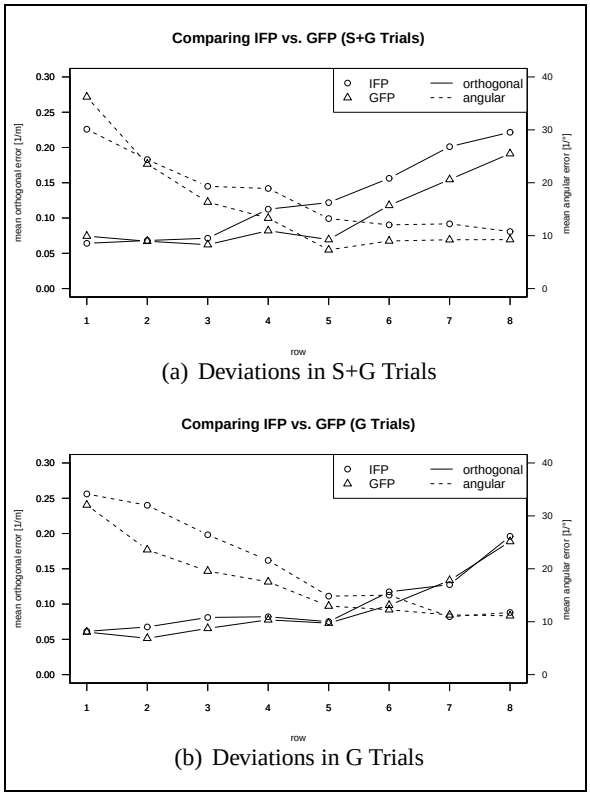


Figure 5: Error estimates for pointing beams: $\gamma =$ angular deviation, $\overline{P\,\text{Object}} =$ orthogonal deviation.

tive value, deviations are measured directly on a meter scale in terms of orthogonal deviation. It is given by the distance between the object's point in space and its orthogonal projection P onto the (prolongation of the) simulated beam.

3 Some Results

Given the outlined measurements we can compare IFP and GFP in terms of preciseness. Plotting the means of their deviations (both orthogonal and angular) against the associated row, the measured IFP and GFP values exhibit a similar envelope, as can be seen from Figure 6, and thus do not permit a preference in either direction. As expected



(a) Deviations in S+G Trials

(b) Deviations in G Trials

Figure 6: Comparing IFP and GFP by means of orthogonal and angular deviation over the rows of the domain.

from the calculations of the error gauges explained above, angular deviation decreases with increasing distance. In opposition, orthogonal deviation rises from row to row. Demonstrations fail their targets – sometimes even by a lot. What do they aim at instead? Plotting the intersection points of tracked demonstrations with the tabletop over the rows of the domain, we get a visual pattern forming "clouds": The impacts of pointing vectors, from IFP as well as from GFP, are distributed around the object to be indicated. The farther the target lies, the more blurred is the shape of the associated scatter-plot, ranging from near circles in the first row to broad and fuzzy regions in the last one. Representative for all plots, Figure 7 shows IFP in G Trials. The ommited ones look quite similar. The plot is based on all DGs' demonstration acts, which, for each object, are averaged by their median. This ensures that each gesture token, be it a long or a short one, makes the same (viz. one) contribution. To make the "clouds"-issue clearer, the areas which are hit by GFP beams stemming from both the S+G and the G Trials are displayed as a bagplot – a bivariate generalization of a boxplot (Rousseeuw et al., 1999) – in Figure 8. The inner

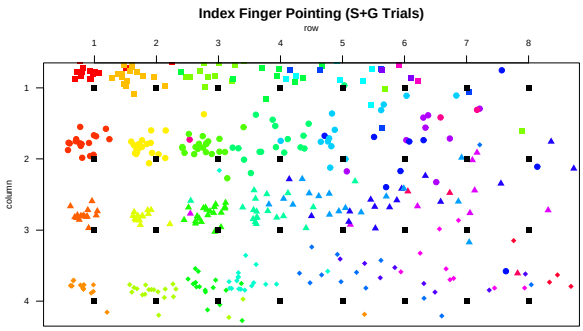


Figure 7: Medial intersections of IFP beams with tabletop in G Trial.

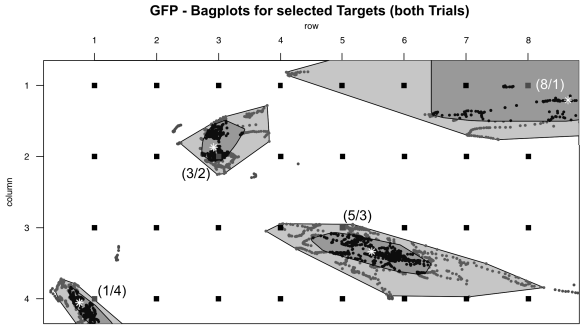


Figure 8: Areas of selected GFP beams. The star marks the median of the cloud, the inner hull frames the data distributed around it.

hull covers 50 % of the data distributed around the "depth median". Using this representation, it can be nicely seen how the clouds grow and get lengthier from row one onwards. In this respect, clouds already exhibit cone-like properties. Those distribution patterns will serve as a basis for us to extrapolate the delineation of the pointing cone from the data (in addition to other parameters and findings of our study – cf. (Kranstedt et al., 2006b, subsec 3.3.4)).

The growing of the clouds may be due to two effects: Firstly, the mean variation of pointing vectors increases naturally with distance; secondly participants *systematically and intentionally* point over the domain when referring to an object in row eight. Thus, they are using what can be called a *gestural hyperbole*. That this behavior is indeed governed by a successful strategy can be seen from Figure 9: The number of identification errors in the G Trials decreases clearly in the last row as compared to the seventh row. There it can also be seen that the participants could identify all objects in the first three rows. The number of failures increases rapidly from the fifth row onwards. Since

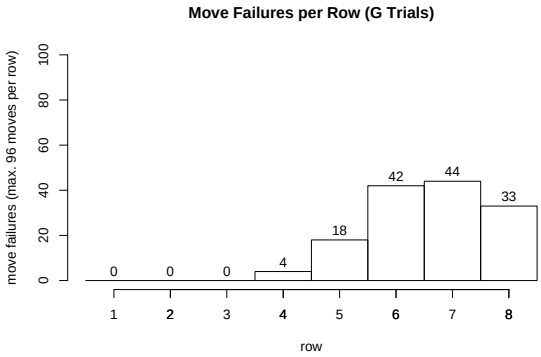
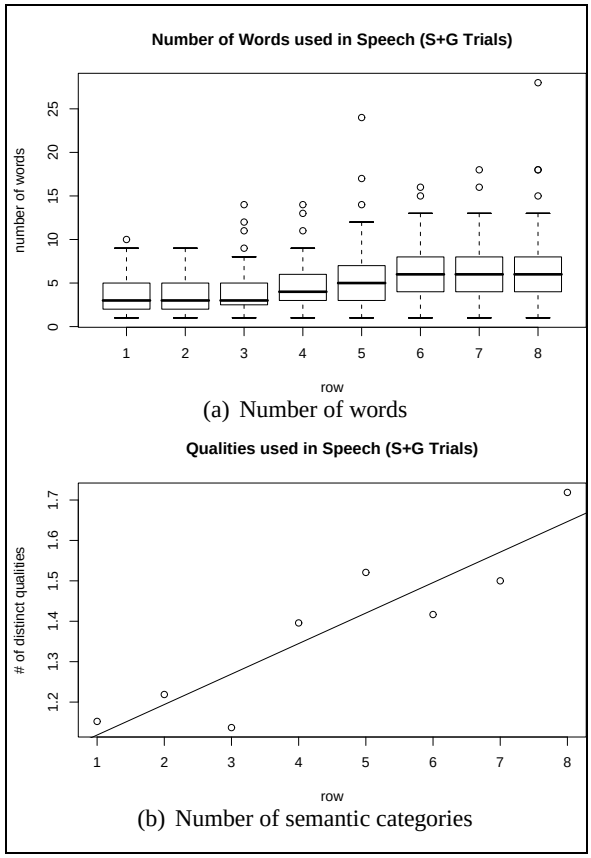


Figure 9: Frequency of identification failures per row.

there are nearly no failures in the S+G Trials we ignore them here.

Considering the S+G Trials, we find two tendencies: 1. The farther away an object is, the more words accompany the gesture; 2. The farther away an object is, the more semantic categories are used to accompany the gesture. Both regularities are depicted in Figure 10.



(a) Number of words

(b) Number of semantic categories

Figure 10: The increasing complexity of co-gestural speech over the rows of the domain.

Since we know about the gestures' loss of discriminatory power wrt distance – this is evident

from the findings presented in Figures 7, 8, and 9 – the tendencies can be ascribed to balancing efforts. The DG compensates his pointing at distance with verbal contributions if he is allowed to, as is the case in the S+G Trials. This can be corroborated by contrasting failures in the G Trials (Figure 9) with the number of words in the S+G Trial (Figure 10(a)): The row in which the number of words increases coincides with the row where the failures increases – both phenomena show up between the fourth and the fifth row. A related increase is shown by the regression line in Figure 10(b) which indicates that the averaged frequency of semantic categories used in referring to objects in the different rows rises from one in row one to nearly two in row eight. Since usually one word expresses a single semantic category, this finding implies that speakers have to use more words if they do not employ a gesture. Indeed, we gained the same result in earlier studies where we expressed it the other way round: *Gestures save words*. It shows the semantic significance of pointings; when referring to objects with gesture and speech people need less words than in referring by speech alone.

Summary of Results. Given a dense domain made up of concrete, equally distributed objects like the one presented here, our findings suggest that pointings can successfully demonstrate objects in the pointer's proximity. However, they seem to do so by delimiting the area the object lies in. The delimiting capacity of pointings diminishes in distance. There, the gesture's spatial cues have to be enhanced by verbal descriptions. The findings are as follows:

- *Pointing is a highlighting (and not a referring) device*. The beams do not meet their targets, they rather encircle them. However, these "clouds" become blurred in the distance. This finding might replace the object-pointing/region-pointing distinction in our earlier work;
- *Pointing breaks down with distance*. Given the density in our setting, pointing starts to get error-prone somewhere between 60.25 and 77.75 cm, which are the distances of the fourth and the fifth row, respectively, measured from DG's end of the table;
- *Distance-dependence of gesture vs speech portions*. Pointings do not permit to single out an object on their own. Distal pointings are accompanied by more complex verbal descriptions. The latter are "more complex" in both numerical and semantic respects. This could not be rendered significant in our earlier studies, but is now in accordance with (van der Sluis and Krahmer, 2004);
- *Partitioning of the pointing domain*. Our earlier investigations suggested a tripartition of the domain into a proximal, a middle, and a distal area. The results presented here suggest a structured domain, too. However, structuring according to the increasing descriptive data would divide the domain into two areas, split somewhere in the middle.

4 On Demonstration: Relating Empirical Data to Theory

If we want to establish a logic of demonstration, we have to deal with at least two questions: Firstly, "What is the logical form (LF) of a demonstration accompanying some expression such as a pure demonstrative or a definite description?" And secondly, "Which models M will be adequate to go with this LF and to provide truth conditions, a suitable notion of entailment and the like?" Since we have to discuss very fundamental things here, we do not want to go into matters of speech act theory, dynamic semantics or sophisticated dialogue theory. Note also that these issues are different from multi-modal integration matters (cf. (Lücking et al., 2006)). For ease of reference, we abbreviate demonstrations, more precisely, their stroke, using $\searrow$.

In order to deal with the LF problem and the M one in a down-to-earth manner let us first recapitulate the empirical findings (referred to as Oi below) which of course do not partition matters into LF-related and M-related ones by themselves. For the start of the discussion we take models M as tupels comprising *inter alia* a domain D.

O1. Empirical domains are structured with respect to DG's proximal and distal relations to targets. Actually, a parameter or index DG should be supplanted by IF or GF-relations indexed by DG;

O2. Demonstrations do not, as a rule, hit their targets;

O3. Demonstrations single out sets of objects rather than single objects;

O4. Proximal demonstrations are distributed around their targets;

O5. Distal demonstrations can encompass other objects besides their target;

O6. The farther away the object demonstrated, the more words accompany the demonstration;

O7. There is a phenomenon of indirect interpretation dubbed *gestural hyperbole*.

How can we account for O1,...,O7? Let us first turn to LF. Here O2 and O3 seem to be of prime importance. Since the extension of a demonstration is not an individual but a region, represented as a set, the LF of a demonstration must not be modelled with a constant but with a one-place predicate. Doing this, a problem arises concerning pure demonstratives like 'this' and their concomitant demonstrations, since $\searrow$ and 'this' are of different type (*predicate* vs. *term*). However, supported by our empirical data and in a way opposed to tradition, we can argue that $\searrow$ does not contribute to the term 'this' as such (and whatever might be used in its place) in *e. g.* 'This is nice' but to the utterance as a whole. As a consequence, we might aim at $[\lambda x(\searrow(x) \wedge \mathit{nice}(x))\mathit{this}]$ to represent the meaning 'This [thing] is demonstrated and is nice'.

Obviously, O1, O4, O5 point into a similar direction and lead us onto issues related to M: While in our setting the extension of a demonstration in the proximal region encompasses only a single object, in the distal region (or in more dense domains) there might be more. This we can accommodate by adding a spatial structure to the model: The model contains a function assigning a coordinate to every object in the domain. Hence we get *distances* between DGs and objects and can do justice to the domain's density. The extension of a demonstration is determined by DG's position, the direction of his pointing, and some pointing domain (in our setting idealized as objects on a surface). To this end, DG's context c determines, *inter alia*, his index finger coordinate (functioning as the anchor point), denoted by c_{IDG}, and the coordinates of his eyes (for orienting the vector in case of GFP), denoted by c_{GDG}. For every gesture occurrence $\searrow_i$ in the context, there is a list of coordinates $[p]_i$ describing the relevant spatial properties of the pointing hand, denoted by $c_{\searrow_i}$.[6] In addition, the pointing domain is represented as a surface s, also part of the context, and denoted by c_s. $\searrow_i$'s intension fixes its extension for every pointing context depending on c_{IDG}, c_{GDG}, $c_{\searrow_i}$, and c_s. It is represented as the function $f : \langle c_{\mathrm{IDG}}, c_{\mathrm{GDG}}, c_{\searrow_i}, c_s \rangle \mapsto \mathit{Ext}(\searrow_i)$ which determines the pointing predicate's extension for all pointing contexts. f is defined in terms of the chosen pointing model, *i. e.* IFP or GFP. So, there is a choice between the two functions *IFP* and *GFP* yielding for every tuple $\langle c_{\mathrm{IDG}}, c_{\mathrm{GDG}}, c_{\searrow_i}, c_s \rangle$ a possibly different pointing cone.[7] The geometrical intersection of this cone with the surface s (*e. g.* the table) yields a region. The collection of the objects in this region *is* the extension $\mathit{Ext}(\searrow_i)$. Moreover, f has the characteristics indicated by the empirical findings, *i. e.* it assigns a smaller extension to pointings in the proximal region and larger extension to pointings in the distal region, extensions having fuzzy borders. It should be clear that from DG's context c a presumably fuzzy partitioning of the domain D into a proximal and a distal subdomain can be reconstructed (*e. g.* that part of the table is distal where there is more than one object in every region pointed at).

Assuming such, the truth conditions for a DG's utterance 'This is nice' amount to ' 'This $\searrow_1$ is nice' is true in context c iff there is exactly one object $o \in D$ such that $o \in \mathit{Ext}(\searrow_1)$ and o is nice.'

If we decide the issue this way, what is going to happen in cases of pointings into the distal region? Well, their felicity will depend on the density of the domain and the meaning of the linguistic information going with the demonstration, which should perhaps have the force of a definite description. This accounts precisely for O5 and O6. If an expression *cum* demonstration turns out to be false wrt the proximal or the distal region, we have to consider a solution along Gricean Pragmatics using the Quality Maxim. The same holds true for the more dramatic O7 cases of indirect interpretation, which are always false on a literal reading.

In sum, if we follow the arguments suggested by the empirical data, we have to separate demonstration from referring, which goes against the prevalent philosophical tradition represented by work from Wittgenstein, Davidson or D. Kaplan. Instead of ending up with two referring terms for

[6] The conceptualisation follows here the work on pure indexicals such as 'I'.

[7] Where c_{GDG} plays no rule for IFP.

the example above, one for 'this' and the other one for ↘ to be related by identity, we get an additional predication, a context-dependent subset of *D*. In a sense, the consequences of the "type shift" of demonstrations from individuals to sets are less dramatic than trying to do without such a shift. Doing without the shift would mean to consider demonstrations as pure referring entities and to treat their non-satisfaction in a neo-Gricean way, perhaps along the lines of Levinson's *Presemantic Pragmatics* (Levinson, 2000).

5 Outlook

To determine the parameters defining function f which assigns extensions to demonstrations in a given context, we have to fix a model for the pointing cone. So a main task in the near future is to derive the delineation of the cone from the empirical data. The concept of a cone and our findings fit well with processing paradigms of pointing represented in (developmental) psychology and linguistics where the function of demonstration is *inter alia* seen in "focusing the attention" (of the addressee). Here as well as in Human Computer Interaction the cone is part and parcel of a precise model for pointing gestures.

However, the empirical findings reported above are difficult to reconcile with traditional philosophical and linguistic theories of demonstration. Therefore we want to compare them to stipulations dealing with demonstration by Wittgenstein, D. Davidson, and D. Kaplan, where the main focus will be "Which paradigmatic cases of demonstration did philosophers found their theories on?"

References

George Butterworth and Shoji Itakura. 2000. How the eyes, head and hand serve definite reference. *British Journal of Developmental Psychology*, 18:25–50.

George Butterworth. 2003. Pointing is the royal road to language for babies. In Sotaro Kita, editor, *Pointing: Where Language, Culture, and Cognition Meet*, chapter 2, pages 9–33. Lawrence Erlbaum Associates, Inc., Mahwah, New Jersey.

Kilem Gwet. 2001. *Handbook of Inter-Rater Reliability*. STATAXIS Publishing Company, Gaithersburg (MD).

Alfred Kranstedt, Andy Lücking, Thies Pfeiffer, Hannes Rieser, and Ipke Wachsmuth. 2006a. Deixis: How to determine demonstrated objects using a pointing cone. In Sylvie Gibet, Nicolas Courty, and Jean-Francois Kamp, editors, *Gesture in Human-Computer Interaction and Simulation*, pages 300–311. Springer, Berlin.

Alfred Kranstedt, Andy Lücking, Thies Pfeiffer, Hannes Rieser, and Ipke Wachsmuth. 2006b. Deictic object reference in task-oriented dialogue. In Gert Rickheit and Ipke Wachsmuth, editors, *Situated Communication*, pages 155–207. Mouton de Gruyter, Berlin.

Stephen C. Levinson. 2000. *Presumptive Meanings*. MIT Press, Cambridge, MA.

Andy Lücking, Hannes Rieser, and Jens Stegmann. 2004. Statistical support for the study of structures in multi-modal dialogue: *Inter*-rater agreement and synchronization. In Jonathan Ginzburg and Enric Vallduví, editors, *Catalog '04—Proceedings of the Eighth Workshop on the Semantics and Pragmatics of Dialogue*, pages 56–63, Barcelona.

Andy Lücking, Hannes Rieser, and Marc Staudacher. 2006. Multi-modal integration. Brandial'06.

David McNeill. 1992. *Hand and Mind—What Gestures Reveal about Thought*. Chicago University Press, Chicago.

Thies Pfeiffer, Alfred Kranstedt, and Andy Lücking. 2006. Sprach-Gestik Experimente mit IADE, dem Interactive Augmented Data Explorer. In *Dritter Workshop Virtuelle und Erweiterte Realität der GI-Fachgruppe VR/AR*, Koblenz. Accepted paper.

Peter J. Rousseeuw, Ida Ruts, and John W. Tukey. 1999. The bagplot: A bivariate boxplot. *The American Statistician*, 53:382–387.

Ielka van der Sluis and Emiel Krahmer. 2004. The influence of target size and distance on the production of speech and gesture in multimodal referring expressions. In *Proceedings of the ICSLP*.

Ludwig Wittgenstein. 1958. *The Blue and Brown Books—Preliminary Studies for the "Philosphical Investigations"*. Harper & Row, New York.