To Beat or Not to Beat:
Beat Gestures in Direction Giving

Mariët Theune and Chris J. Brandhorst

Human Media Interaction
University of Twente
PO Box 217, 7500 AE Enschede
The Netherlands
`m.theune@ewi.utwente.nl, c.j.brandhorst@alumnus.utwente.nl`

Abstract. Research on gesture generation for embodied conversational agents (ECA's) mostly focuses on gesture types such as pointing and iconic gestures, while ignoring another gesture type frequently used by human speakers: beat gestures. Analysis of a corpus of route descriptions showed that although annotators show very low agreement in applying a 'beat filter' aimed at identifying physical features of beat gestures, they are capable of reliably distinguishing beats from other gestures in a more intuitive manner. Beat gestures made up more than 30% of the gestures in our corpus, and they were sometimes used when expressing concepts for which other gesture types seemed a more obvious choice. Based on these findings we propose a simple, probabilistic model of beat production for ECA's. However, it is clear that more research is needed to determine why direction givers in some cases use beats when other gestures seem more appropriate, and vice versa.

Keywords gesture and speech, gesture analysis, beats, direction giving

1 Introduction

When humans speak, they use gestures that "are not random but convey to listeners information that can complement or even supplement the information relayed in speech" [1], p. 228. One type of discourse in which this relation is undoubtedly present is direction giving. To illustrate this, consider two of the main gesture types distinguished by gesture researcher David McNeill [2]. *Deictic gestures* are pointing movements indicating the location of items being referred to. In direction giving, such gestures are often used to indicate the location of landmarks along a route [3]. *Iconic gestures* depict a physical aspect of what is spoken about, such as the shape of an object or the trajectory of a movement. Such gestures are often used to illustrate the shape of landmarks [4].

For another important type of gestures, however, the link with direction giving is less obvious. *Beat gestures* do not convey any semantic content, but reflect discourse structure by marking important words and phrases. Unlike other gestures, they tend to have the same shape regardless of the speech content. McNeill describes their shape as follows:

The hand moves along with the rhythmical pulsation of speech. [...] The typical beat is a simple flick of the hand or fingers up and down, or back and forth; the movement is short and quick and the space may be the periphery of the gesture space (the lap, an armrest of the chair, etc.). The critical thing that distinguishes the beat from other types of gesture is that it has just two movement phases – in/out, up/down, etc. [2], p. 15

In a video corpus of people narrating the events from a Tweety cartoon, McNeill found that beats made up 44,7% of all gestures [2], p. 93. Though the beat ratio may be different for other types of discourse, McNeill's finding shows that beats are frequently used by human speakers, and therefore should not be overlooked when developing gesture models for embodied conversational agents (ECA's): human-like computer characters that can employ gestures and speech to carry out conversations with human users.

In our department we have developed an ECA that can give directions to visitors in a virtual environment [5]. This ECA, called the Virtual Guide, can generate deictic and (simple) iconic gestures, but it has only very limited support for beat gestures. To improve this, we analysed the use of beat gestures in a video corpus of human route descriptions, with the aim of using the results for a simple beat usage model for the Virtual Guide. First, however, we needed to determine which of the gestures in our corpus were beats and which were not.

The research questions addressed in this paper are the following:

1. How can beat gestures be distinguished from other gesture types?
2. At which points in route descriptions do people use beat gestures?
3. Knowing when to use beats, how can this be modelled for the Virtual Guide?

The remainder of the paper is structured as follows. First, in Section 2 we discuss related work on gesture generation for (direction giving) ECA's. In Section 3 we describe our route description corpus. Then, in Section 4 we examine whether beats can be distinguished from non-beats based on their physical properties. In Section 5 we investigate when beat gestures are used during route giving discourse. In Section 6 we propose a simple probabilistic model for the generation of beat gestures, and in Section 7 we end with conclusions and future work.

2   Related Work

NUMACK (the Northwestern University Multimodal Autonomous Conversational Kiosk) is an ECA that can give directions to locations on the Northwestern University campus, using a sophisticated 'multimodal microplanner' for integrated language and gesture generation. The generation of iconic gestures is based on a model by Kopp et al. that links visual properties of objects to gesture features such as hand shape and trajectory [4]. Using this model, new iconic gestures that appropriately reflect the shape of landmarks can be assembled on the fly, instead of using fixed gesture animations as is done by most ECA's (including our Virtual Guide). NUMACK can also generate gestures indicating the location

of landmarks, as described by Striegnitz et al. [3]. However, beat gestures do not appear to be included in NUMACK's gesture repertoire.

A well-known framework for automatic gesture and speech generation for animated characters is BEAT, the Behavior Expression Animation Toolkit [6]. It can be used to animate an ECA based on an input text that is automatically analysed and augmented with suggestions for nonverbal behaviour. This augmentation is done in a "liberal and all-inclusive" fashion: any gesture that is deemed appropriate is suggested and given a priority. Beats are used when introducing new material or when contrasting items and are always given the lowest priority. They are only selected when no higher-priority gestures are available to express the same information (unless they can be overlaid on top of the other gesture). Similar approaches to gesture generation, in which the use of more specific gestures is preferred over beat gestures, include [7–9].

A completely different approach to gesture generation is that by Neff et al. [10], who create statistical models that capture the gesture style of individual speakers based on annotated video material. In their system, gesture choice is based on speaker profiles: probabilistic mappings from semantic tags (capturing aspects of the semantics and communicative function of the verbal message) to gesture types. In this approach, the probability of generating a beat gesture is based on the frequency with which the modelled speaker used beat gestures in combination with a particular semantic tag, as encoded in the speaker's profile.

Most recently, Bergmann and Kopp proposed a data-driven model for integrated language and gesture generation that can still account for systematic meaning-form mappings, where speaker preferences are learned from corpus data [11]. However, like [4], this model is restricted to iconic gestures.

3   Route Description Corpus

Our corpus comprises 16 short movie clips with an average duration of 38 seconds. Each clip shows a person giving an indoor route description in Dutch. All descriptions start from the same point in the building (the point where the direction giver is standing). The movie clips differ in a number of respects:

- Route: two different routes are described. They have the same starting point but a different destination within the same building.
- Camera viewpoint: in 8 movie clips, the direction giver explains the route to the route seeker in a face-to-face dialogue. In the other 8 clips, the route is described to the camera.
- Direction giver: four direction givers were filmed. All were male students or employees in our department, and native speakers of Dutch. Each of them explained both routes twice: first to the route seeker and then to the camera (see the previous point). This resulted in four movie clips for each speaker.

The movie clips were transcribed and segmented into gesture clips using Transana,[1] resulting in a data set of 162 gestures.

---

[1] www.transana.org

4 Distinguishing Beats from Other Gestures

In this section it is examined whether beat gestures can reliably be distinguished using physical properties only. To this end, we annotated the gestures in our video corpus with Beat Filter scores and gesture types.

4.1 The Beat Filter

McNeill's Beat Filter[2] is a method for distinguishing imagistic gestures such as iconics from non-imagistic gestures, i.e., beats [2]. It is a purely formal scoring system, without reference to content or function. It only looks at the kinetics of the gestures. Applying the beat filter to a gesture means giving it a score by adding 1 for each positive answer to the following questions (except question 2). The higher the resulting score, the less likely the gesture is a beat.

1. Does the gesture have more than two movement phases?[3]
2. How many times does wrist/finger movement OR tensed stasis appear in any movement phase not ending in a rest position? (ignore retraction phase, add the number of times to the score)
3. If the first movement is in non-central space: is any other movement performed in central space?[4]
4. If there are exactly two movement phases: is the first phase in a different place as the second phase?

The beat filter was applied by two annotators (the authors) to 154 of the 162 gestures in the corpus. The other 8 gestures were not clearly visible, for example because the speaker turned his back to the camera, and could not be annotated.

4.2 Gesture Types

The Beat Filter does not explicitly group gestures into beats or non-beats; the resulting score only represents the (un)likeliness of a gesture being a beat. To determine the relation between Beat Filter scores and gesture type, all 154 visible gestures from the corpus were independently annotated for gesture type by three annotators: the authors plus a third annotator. The gesture types used were those from McNeill [2]: beats, deictic gestures, iconic gestures, and metaphoric gestures. The latter are like iconics, but describe non-physical, abstract entities, for example shaping the hands like a bowl to illustrate the concept 'group'.

Many gestures do not neatly fit into one of the four above-mentioned gesture categories; they may have features of more than one gesture type, for example because a beat is superimposed on another gesture [2]. Therefore we included the possibility of annotating gestures as belonging to more than one type. In

---

[2] Originally developed by Bill Eilfort.

[3] Movement phases are preparation, stroke and retraction. Beats have no stroke.

[4] The *central space* is the part of the gesture space directly in front of the torso, excluding the hip area and lower [2].

cases when no one dominant type could be established for a particular gesture, it was annotated as a mixed type, e.g., beat/iconic.

In general, the gesture type annotations were based on the gestures' global shape in combination with the speech context, i.e., the words spoken while the speaker was gesturing. For example, if the hands were moved forward in parallel, mimicking a tunnel-like shape when talking about a "hallway", the gesture was annotated as iconic. If the speaker pointed in a certain direction in combination with words such as "left", "right" or "there" this was annotated as a deictic gesture. Beat gestures formed an exception to this: since they have no inherent meaning, they were classified on the basis of their shape alone.

Beat gestures and deictic gestures, which can be somewhat similar in shape, were distinguished based on the amount of extension of the arms (the larger this extension, the more probably it is a deictic gesture), hand shape (extension of the index finger indicates a deictic gesture), and directional aspect in combination with the speech context. Concerning the latter property, we assume that beats are in principle 'directionless', meaning that when making a beat gesture, the hand does not move in the horizontal plane but only in the vertical plane. This is in line with McNeill's characterisation of beats as low-energy gestures with the lowest kinetic complexity [2]. So, if a speaker mentioned a specific direction or landmark while making a somewhat beat-like gesture in the corresponding direction, this was annotated as a deictic gesture, not a beat.

Note this means that gestures were classified as beats only when they had the right shape and there were no indications (e.g., from the speech context) that they were of another type. This 'classification by negation' approach may have led to an underestimation of the number of beats in our data.

4.3 Results

When analysing the results, our first step was to analyse the reliability of the annotations by computing the level of agreement between annotators in terms of the Kappa coefficient. When considering all possible gesture types, agreement between pairs of annotators was quite low (Kappa values ranging between .41 and .44). However, when only considering the distinction between beat gestures and other types of gesture, i.e., when classifying all non-beat gestures as 'other', agreement between annotators was much better with Kappa values of .68, .60 and .57 between annotator pairs. Though not all good according to the strictest scale for evaluating Kappa significance, according to more lenient scales these values indicate at least moderate agreement [12]. In the remainder of this paper, we therefore classify the gestures in our corpus as either beats or 'other' gestures, referring to more specific types only when necessary. For the final type classification we used the type assigned by the majority of the annotators. This resulted in 52 gestures being classified as beats, which amounts to 32,1% of all gestures in our corpus (33,8% of all annotated gestures).[5] This set includes 7 beats that were

_____

[5] The actual percentage of beats in our corpus may be slightly higher, because some of the 8 unannotated gestures could be beats.

classified as beat/other, and 2 what we termed 'multibeats': quick sequences of beats that could not be separated into individual beat gestures.

We found large differences in beat usage between individual speakers. Table 1 shows the total number of gestures per speaker, the number of beats, and also the average number of gestures per word. Note the striking contrast between speakers 1 and 4: the former used few gestures, many of which were beats, while the latter used many gestures, few of which were beats.

Table 1. Gesture use of individual direction givers.

	Gestures	Beats	Gestures/word
Speaker 1	23	10 (43,5%)	.06
Speaker 2	61	24 (39,3%)	.12
Speaker 3	40	12 (30,0%)	.10
Speaker 4	38	6 (15,8%)	.14
Total	162	52 (32,1%)	

For the Beat Filter, agreement on the filter questions was unfortunately very low. The highest agreement between the two annotators was .34 for the answers to question 1. This means that the Beat Filter scores assigned to the gestures in our corpus are very unreliable. Nevertheless, as illustrated by Fig. 1, the Beat Filter does give some indication of the probability that a gesture is a beat: for both annotators, gestures with a low score are more likely to be beats than not. In Fig. 1, the multibeats are shown separately from the other beats. This is because the former are always assigned a relatively high score by the Beat Filter, since these successions of beat moves are seen as one gesture ([2], p. 381).

As can be seen in Fig. 1, several gestures were assigned a low Beat Filter score despite not having been classified as beats. Most of the non-beat gestures with a score of 0 or 1 were annotated (by the majority of annotators) as deictic gestures: 13 out of 19 (68%) for annotator A and 11 out of 20 (55%) for annotator B. This is not surprising, since deictic gestures are fairly similar in shape to beat gestures, as discussed in Section 4.2.

4.4 Discussion

As shown above, human annotators can fairly reliably distinguish beats from other gestures based on a global impression of their shape, but they cannot reliably apply the Beat Filter that was designed to make the same distinction in a more formal way. Moreover, Fig. 1 shows that although lower Beat Filter scores do tend to correspond to higher relative numbers of beat gestures, many gestures with a low beat score are not beats. In most cases these 'other' low scoring gestures turn out to be deictic gestures, which can be very similar in shape to beats. This holds in particular for what we call 'weak' deictic gestures, i.e., deictic gestures on which the speaker did not spend much energy. They are
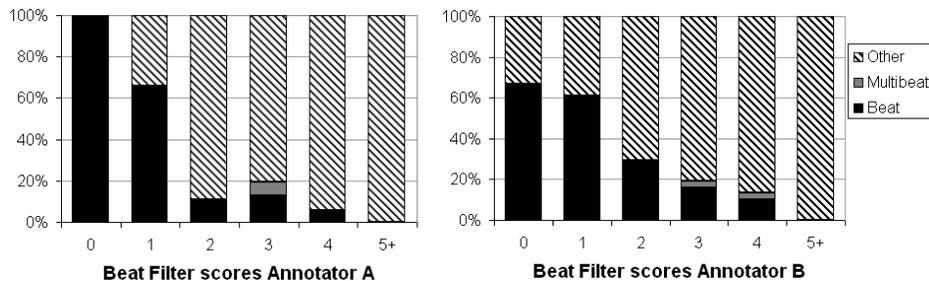
Fig. 1. Beat Filter scores and gesture types

small and quick: the hand only moves a short distance into the direction that is indicated, staying inside the periphery of the gesture space, and there is no tensed stasis or finger movement. Both characteristics are shared by beats and earn zero Beat Filter points for questions 2 and 3. Moreover, other features that do distinguish deictic gestures from beats (arm extension and the presence of a directional component) are not checked by the Beat Filter. In other words, the Beat Filter is not well-suited to distinguish beats from deictic gestures. This can be explained by the fact that the Beat Filter was only designed to distinguish imagistic gestures (iconic/metaphoric) from non-imagistic gestures (beats), and deictic gestures are somewhere in between the two.

To make the Beat Filter better suited for distinguishing between beats and all other gestures it needs to be extended with additional questions that are specifically aimed at filtering out deictic gestures, by checking for directional aspects and arm extension. However, even then it will remain difficult to distinguish beats from 'weak' deictic gestures. In addition, the description of the Beat Filter will have to be improved so the questions cannot give rise to different interpretations by individual annotators, which we assume was one of the causes for the low agreement found in our study.

5   When Are Beat Gestures Used?

This section takes a closer look at the speech context in which beat gestures are used. We identified some important route description concepts and examined by which type of gestures (beats or other) they were accompanied in our data.

### 5.1   Route Description Concepts

Conceptually, the basic elements of route descriptions are **paths**, instructions to move along some pathway, **turns**, instructions to change direction at a choice point, and **landmarks**, mentions of objects along the route that help with navigation, in particular by signalling where turns are to be made [13]. A fourth category distinguished in [13] is that of **location** information, indicating the spatial location of the destination. For our purposes, we have replaced this concept

with two new categories: the more general **spatial information** indicating the spatial location of all route objects (not just the destination) and **destination**, which are direct references to the destination. Additional concept categories we distinguish are **deictic references**, situationally dependent references to points in time and space, and **hesitations**. These are not specific to route descriptions, but they occurred frequently in our corpus in combination with gestures.

Below, we list all concept categories used in our analysis, together with some examples of how these concepts were verbally expressed in our corpus.[6] In some examples, multiple concepts are mentioned in one phrase. Here, the words that were accompanied by a gesture are given in italics, to indicate which concept was marked by the gesture:

– **Paths**: "*through* the corridor, "*past* the lavatories", "*all the way* to the end"
– **Turns**: "turn left", "walk downstairs", "go in that direction".
– **Landmarks**: "very long corridor", "spiral staircase", "windows"
– **Spatial information**: "then you are *near*", "*behind* it we see lots of computers", "the tunnel *on the right*"
– **Destination**: "the East Hall", "the practicum rooms"
– **Deictic references**: "now", "here", "over there", "*that* corridor"
– **Hesitations**: "ehm", "maybe", "I don't know"

For each of the 162 gestures in our corpus, we annotated which concept it accompanied. If the speech context of the gesture did not match any of the categories given above, the concept was classified as **other**.
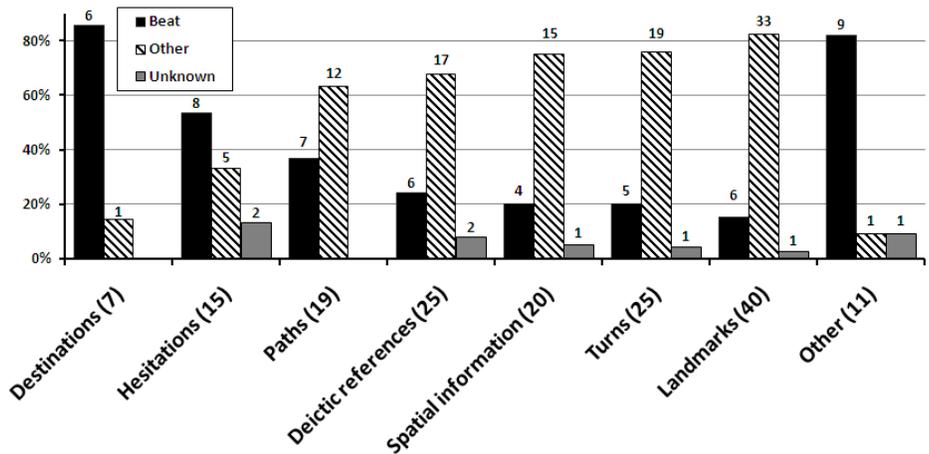
Fig. 2. Concept categories and gesture types for all 162 gestures in the corpus. For 8 gestures the type could not be determined; these are labeled as 'unknown'.

_____

5.2 Results

Figure 2 shows the results of our analysis, where the 8 'unknown' gestures are those of which the type could not be determined (see Section 4). In our corpus, some concepts are more often accompanied by beats than any other gestures. In the first place we find **destinations**: 85,7% of all gestures accompanying the mention of a destination are beats. Beats are also prevalent during **hesitations**. Here, 53,5% of accompanying gestures are beats. Finally, almost all (81,8%) of the gestures accompanying **other** concepts are beats. This category is mainly made up of various discourse structure markers ("*and* because", "*so* I'd say", "which *also* says") and abstract actions ("what you want to *do* is", "then you *see*"). For the remaining concepts, other gestures were used more frequently than beats, with beat frequencies ranging from 36,8% (**paths**) to 15% (**landmarks**).

5.3 Discussion

In our corpus, references to the route **destination** are predominantly accompanied by beats. Presumably this is because these references mostly had the form of proper names rather than descriptions referring to shape or location, meaning that the use of an iconic or deictic gesture was not appropriate in these cases.

We also found a relatively high number of beats accompanying *hesitations*.[7] One possible explanation for this is Krauss' hypothesis that gesturing aids lexical access [14]. Interestingly, Krauss' hypothesis was explicitly limited to 'lexical gestures', i.e., non-beats, while our findings suggest that beats might play a similar role. An alternative explanation is that the beats serve as 'attempt-suppressing signals' indicating that the speaker intends to hold the turn, thus suppressing any interruption attempts by the conversation partner while the speaker is searching for words [15].

For the other categories besides the rest category **other**, beats are in a clear minority. References to spatial locations, directions and landmarks lend themselves well to being illustrated by deictic or iconic gestures, which may explain why beat gestures are only rarely used when expressing these concepts. Still, the fact that beats are used at all, when seemingly more appropriate gestures are available, is somewhat surprising. To shed more light on this issue, we take a closer look within some of the concept categories, inspecting the specific cases in which beats are used. For **deictic references**, it turns out that most beats accompany references to the "here" and "now" of the speaker (4 beats out of 5 gestures) rather than references to concrete, visible locations (2 beats out of 20 gestures). This make sense, since for concrete spatial references beats are less useful than deictic gestures (17 of 20), as the latter may help the hearer to identify the referent. On the other hand, pointing does not have much added value in case of 'here and now' references, which are quite unambiguous. Gestures accompanying these references only seem to be used for marking them as new or

---

[7] The beat ratio for hesitations in our corpus may be relatively high because one speaker uttered relatively many hesitations, mostly accompanied by beat gestures.

otherwise important, and this discourse function can be fulfilled with the least effort by a beat gesture.

On a smaller scale, this 'principle of least effort' also seems to apply to **spatial information**. If we split this category into references to *topological* information and references to *projective* information, cf. [16], we see that beats are used more often for topological information (3 beats out of 10 gestures) than for projective information (1 beat out of 10 gestures). Again, a possible explanation is that deictic gestures have less added value for topological information (references to a region proximal to some object, e.g., "near", "behind") than for projective information (references to a particular direction relative to an object, e.g, "to the left of"), so for topological information speakers are more likely to use the less effortful beat gestures instead. However, in both this and the previous case, our data are too sparse to draw any strong conclusions from them.

Another explanation for the use of beats with 'less obvious' concept categories lies in the notion of information structure. McNeill claims that less informative discourse elements are more likely to be accompanied by beats than by other gestures [2]. This holds for example for anaphoric references to discourse elements that have been previously mentioned. When inspecting the **landmarks** category, we see that our corpus contains 8 anaphoric references to landmarks (within the same route description) that are accompanied by a gesture, and for these the 'beat ratio' is 3 beats out of 8 gestures (37,5%) as opposed to 3 beats out of 32 gestures (9,4%) for first mentions. Though again these data are too low in number to allow any strong conclusions, they do support the information structure explanation for the use of beats in references to landmarks. Another finding that points in this direction is the fact that most beats were found in the second versions of the route descriptions in our corpus. On average, the second versions had about twice as many beats as the first.

6   A Simple Model of Beat Gesture Use

Some gesture generation models for ECA's only select beat gestures when no other gestures are available [8, 9]. In contrast, we propose to give beat gestures the same basic priority as other gesture types. Given the results of statistical corpus analysis, along with the notion that the use of beat gestures also depends on personal style of the individual speakers, the probability that a beat gesture is generated in a certain context can be given by the following formula:

$$P(B|u) = P(B|c_u)m_s$$

where B is the generation of a beat gesture, u is the speech context (a word or phrase to be uttered), c_u is the concept being expressed by the utterance, and m_s is a multiplier for a specific speaker.

This probability function can also be used for other gesture types. It can be easily applied in the Virtual Guide, which already uses a weighted randomization algorithm for gesture selection [5]. It would also be applicable in other

frameworks such as BEAT [6] that assign priorities to possible gestures, which is something a probability can be used for. Note that our proposed data-driven approach is similar in spirit to that of Neff et al. [10], though their model is far more sophisticated. Whether this sophistication also leads to better results than our simple approach or is overly complex for a merely marginally better result is a question that can only be answered when our model has been implemented and tested in practice. To this end, more data have to be gathered to feed the model.

7 Conclusions and Future Work

Our corpus analysis has shown that beat gestures are frequently used within route descriptions. We found that, in line with the literature [2], beats are most often used to mark important concepts in the discourse. In the case of direction giving discourse, the concepts marked by beats tend to be the ones that cannot be easily visualised using other gestures, such as (named) route destinations and topological spatial information. However, beats are also used - albeit much less frequently - with concepts for which other gestures seem a more obvious choice, for example turn directions. These findings can be at least partially explained in terms of information structure: information which is 'discourse-old' is more likely to be accompanied by a beat than by another type of gesture (if a gesture is used at all).

We applied McNeill's Beat Filter on our corpus, to see if we could reliably distinguish beats from other gestures on purely formal grounds [2]. We found that very low agreement between annotators, probably due to different interpretations of the filter questions. To avoid this a more detailed coding manual will be required, defining exactly what counts as movement phases etc. and how borderline cases should be handled. Probably, thorough annotator training will be needed as well. In addition, to make the Beat Filter more useful it should have additional questions to distinguish between beats and deictic gestures.

We have proposed a probabilistic model of beat gesture use in direction giving in which the likelihood of using a beat gesture to mark certain concepts is derived from corpus data, similar to the approach of Neff et al. [10]. Though this is a step forward compared to the way beats are currently handled in the Virtual Guide, as well as those ECA models where beats have a lower priority than other gestures [6–9], we are aware that the model is still far too simple. In the current version of the model, gesture choice only depends on the concept being expressed, optionally weighted to take speaker preferences into account. In reality, gesture choice is also influenced by other factors, including the newness of the presented information. Nevertheless, we expect that implementing our current model will already increase the perceived naturalness of our direction giving ECA. Before we can test this, however, we need more – and more reliably annotated – corpus data to derive the gesture probabilities needed by the model. Having more data available may also help uncover additional factors influencing direction givers' choice to use beat gestures in certain contexts.

References

1. Iverson, J., Goldin-Meadow, S.: Why people gesture when they speak. Nature **396** (1998) 228
2. McNeill, D.: Hand and Mind: What Gestures Reveal about Thought. The University of Chicago Press, Chicago, IL/London, UK (1992)
3. Striegnitz, K., Tepper, P., Lovett, A., Cassell, J.: Knowledge representation for generating locating gestures in route directions. In Coventry, K., Tenbrink, T., Bateman, J., eds.: Spatial Language in Dialogue. Oxford University Press (2009) 147–166
4. Kopp, S., Tepper, P., Striegnitz, K., Ferriman, K., Cassell, J.: Trading spaces: How humans and humanoids use speech and gesture to give directions. In Nishida, T., ed.: Engineering Approaches to Conversational Informatics. John Wiley & Sons (2007) 133–160
5. Theune, M., Hofs, D., van Kessel, M.: The Virtual Guide: A direction giving embodied conversational agent. In: Proceedings of Interspeech. (2007) 2197–2200
6. Cassell, J., Vilhjálmsson, H., Bickmore, T.: BEAT: the Behavior Expression Animation Toolkit. In: Proceedings of SIGGRAPH. (2001) 477–486
7. Hartmann, B., Mancini, M., Pelachaud, C.: Formational parameters and adaptive prototype instantiation for MPEG-4 compliant gesture synthesis. In: Proceedings of Computer Animation 2002, IEEE Computer Society (2002) 111–119
8. Nakano, Y.I., Okamoto, M., Nishida, T.: Enriching agent animations with gestures and highlighting effects. In: Intelligent Media Technology for Communicative Intelligence. (2004) 91–98
9. Olivier, P., Jackson, D., Wiggins, C.: A real-world architecture for the synthesis of spontaneous gesture. In: Proceedings of the 19th annual conference on Computer Animation and Social Agents (CASA). (2006)
10. Neff, M., Kipp, M., Albrecht, I., Seidel, H.P.: Gesture modeling and animation based on a probabilistic re-creation of speaker style. ACM Transactions on Graphics **27**(1) (2008) 5:1–24
11. Bergmann, K., Kopp, S.: Increasing the expressiveness of virtual agents – autonomous generation of speech and gesture for spatial description tasks. In: Proceedings of AAMAS. (2009) 361–368
12. DiEugenio, B.: In the usage of Kappa to evaluate agreement on coding tasks. In: Proceedings LREC. (2000) 441–444
13. Williams, S., Watson, C.: A profile of the discourse and intonational structure of route descriptions. In: Proceedings of Eurospeech. (1999) 1659–1662
14. Krauss, R.M.: Why do we gesture when we speak? Current Directions in Psychological Science **7** (1998) 54–59
15. Duncan, S.: Some signals and rules for taking speaking turns in conversation. Journal of Personality and Social Psychology **23**(2) (1972) 161–180
16. Kelleher, J.D., Costello, F.J.: Applying computational models of spatial prepositions to visually situated dialog. Computational Linguistics **35**(2) (2009) 271–306