Fluency or Disfluency?

Jens Allwood [1]

[1]*SCCIIL Interdisciplinary Center, University of Gothenburg, Gothenburg, Sweden*
.

Abstract

In this paper, I investigate the concepts of "fluency" and "disfluency" and argue that the application of the two concepts must be relativized to type of communicative activity. It is not clear that there is a generic sense of fluency or disfluency, rather what contributes to fluency and disfluency depends on what type of communication we are dealing with.

The paper then turns to a brief investigation of what makes interactive face-to-face communication fluent or disfluent and argues that many of the features that have been labeled as disfluent, in fact, contribute to the fluency of interactive communication. Finally, I suggest that maybe it is time for a change of terminology and abandon the term "disfluent" for more positive or neutral terminology.

1. Why interesting?

The phenomena that sometimes go under the name of "disfluencies" are a pervasive feature of human communication. In written communication, they are to some extent edited out on the basis of normative criteria. In spoken and gestural communication, they are, however, a regular part of the ongoing flow. It seems unlikely that such a common, regular phenomenon is only "dysfunctional" or "dis-functional" or has no function at all. Rather, it seems to have functions that are interesting in themselves and deserve further study.

An ancillary reason for an interest in "disfluencies" is the question of whether artificial dialog systems in virtual agents or robots should be devoid of this feature. This is related to the more general question of what features a dialog system should have. Are "disfluent" features desirable or not desirable in a dialog system? Is what is "disfluent" constant across different human types of communication? All these questions lead back to the question of the nature of fluent and disfluent communication.

2. What is fluent varies with type of communication

A first observation we can make is that the ideals of fluency vary with type of communication. Fluency in written language involves writing in a manner, which is easy to read, making use of full sentences and judicious punctuation, while fluency in spoken language involves clear pronunciation, audibility and clear relevant gestures. In addition, spoken language ideals of fluency are different in different social activities. Fluency in public speaking involves such things as not presupposing context not shared, making good use of what could possibly be shared, being clear, holding attention, evoking interest and positive emotions, being audible and visible, while fluency in interactive (small) talk, with friends, involves such things as making efficient use of the much larger amounts of shared background information available, as well as being flexible and open for interactive cooperation in co-constructing content which, in turn involves such things as being able to change one's mind and having time to think.

Usually, disfluency varies with fluency, so that what is seen as "disfluent" can be seen as the negation of what is seen as "fluent", that is, not being clear, audible, presupposing as shared what is not shared etc.

So, what is fluent or disfluent in written language is not necessarily fluent or disfluent in interactive face-to-face communication and vice versa. Nor is what is fluent or disfluent in public speaking necessarily fluent or disfluent in private friendly interactive face-to-face communication and vice versa.

Finally, we may note a related use of the term "fluent" in connection with learning a new language. We talk about "fluency in a foreign language", referring to the ability to find words and use grammar easily. For a discussion of other aspects of fluency and disfluency, see Lickley (2015).

3. Fluency and disfluency in interactive communication

3.1 A model of interactive communication

Let us now consider some of the features of fluency and disfluency in interactive communication. We will take as our point of departure the model of interactive embodied communication proposed in Allwood, Grammer, Kopp and Ahlsén (2006) (see figure 1).

The model shows how interactive communication involves at least two communicators (A and B), forming a dynamic system of co-activation involving several different levels of awareness.

Like in Kahneman (2011), the model distinguishes processes on a high level of awareness, that are slower and involve responses based on evaluation and deliberation, from processes on a low level of awareness, that are faster and involve reactions based on more automatic appraisal and cognition.

The processes on higher levels of awareness are related to the processes on lower levels of awareness through a gradient, the specific nature of which needs to further investigated.
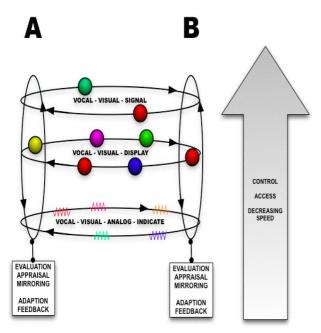

Figure 1. General model of embodied communication.

As a start of such an investigation, three levels of awareness are distinguished in production:

(i) *indicate* – the lowest level of awareness and conscious control. This involves being informative to an interlocutor without any communicative intention, e.g. through vocal features that indicate age, gender or dialect,

(ii) *display* – an intermediate stage of awareness and conscious control, involves intentionally expressing information (for an interlocutor), e.g. an emotion like joy or sorrow,

(iii) *signal* – the highest stage of awareness and control, involves expressing information for an interlocutor in such a way that the interlocutor should notice that the information is being expressed for him/her.

The three levels of awareness are connected and interact so that a feeling of joy, initially automatically "indicated" in intonation or facial gestures can become more aware and then more intentionally "displayed" and finally also

intentionally "signaled" through a verbal utterance like "Great that you could come". Other processes go the other way and connect impulses on a high level of awareness with more automatic reactions on lower levels

Also, on the recipient side, higher levels of awareness and control are integrated with lower levels of awareness and control. Automatic fast processes of perception, reaction and appraisal are connected with and can influence slower processes of evaluation, deliberation, planning and response and going the other way slower processes can influence the faster less aware processes.

Both in production and reception, the processes can be sequential and simultaneous.

In interactive communication, vertical processes connecting higher levels of awareness and control with lower levels of awareness and control, interact with horizontal processes, connecting interlocutors with each other, on different levels of awareness, so that we both influence and are influenced by others on several levels of awareness. Interactive communication, in this way, forms a partly self-organizing system with vertical and horizontal subsystems.

The horizontal system (interactive communication management (ICM, Allwood, 2013) involves many interactive communication components, the most important being the feedback system, whereby interlocutors give each other multimodal feedback (mostly visual and auditory) concerning perception, understanding, emotional and other attitudinal reactions. The information given in the feedback system can be indicated, displayed or signaled. This also means that the means of expression can range from more or less conventionalized vocal verbal expressions, like *yes*, *no*, *mm*, (Lindblad and Allwood, 2013) or gestured verbal expressions, like head nods or head shakes, to less conventionalized, so called "conversational grunts" (Ward, 2006).

The vertical system (own communication management (OCM, Allwood, 2013), similarly, involves many components, two of the most important being:

(i) mechanisms for planning and selection of expressions and their combination (lexicon and grammar), for short "choice mechanisms", and

(ii) mechanisms for on-line modification and change of ongoing production, for short "change mechanisms" (see Allwood, Nivre and Ahlsén, 1990).

Like in ICM, OCM processes can be indicated, displayed or signaled, leading to means of expression that can be more or less conventionalized, ranging, from fully conventionalized hesitation words like *eh* and facial gestures (to gain time) to displayed and indicated

such means, including also processes allowing on-line change management, ranging from signaled explicit negation to more ad hoc indicated means.

The two systems are integrated, so that many expressions can function both in vertical and horizontal processes, e.g. a hesitation expression can give feedback to an interlocutor (ICM), while also gaining time for a speaker to plan and select appropriate means of expression (OCM).

The important thing in all cases is that all processes, (both OCM and ICM) should be means of joint sharing of content and sometimes also explicit co-construction of content.

3.2 Fluency in interactive face-to-face communication

Achieving fluency in interactive face-to-face communication involves achieving at least the following goals:

(i) Being able to communicate while taking context and your interlocutor(s) into account, i.e., not belaboring what is given by context and being sensitive to simultaneously indicated, displayed and signaled vocal and gestural feedback, which is conventionalized to varying extents.

(ii) Being able to hold the floor in order to plan and select what you want to express.

(iii) Being able to manage, e.g. change what you are communicating in such a way that your interlocutor can follow you.

(iv) Being able to keep, yield, give, assign, take, and accept turns.

(v) Being able to actively listen, react and respond by giving vocal and gestural feedback regarding perception, understanding, emotional and other attitudes.

(vi) Being able to co-construct content with your interlocutor, often using short and relevant utterances and gestures.

3.3 What is disfluency?

Let us now define "communicative disfluency" in the following manner.
"Communicative disfluency = Something in the communicative performance that disturbs the flow of communication". For a discussion of different definitions and characterizations of "disfluency", see Eklund (2004).

Some examples of what has been proposed as "communicative disfluencies" include:

(i) Mechanisms for hesitation or clarification, like *eh* or *I mean*, lengthening, pausing or self-repetition, which all have the effect of holding the floor.

(ii) Mechanisms for changing the expression or content of what you are communicating.

(iii) Short words, phrases to give feedback.

(iv) Stammering.

With the possible exception of stammering, we can now raise the question: are these really examples of disfluencies? Are they not rather examples of phenomena that are needed to make interactive communication fluent? Even for stammering, we might wonder if this phenomenon for a particular individual in a particular state might not be what is required to communicate.

Another way of approaching the "disfluent" phenomena exemplified above is to ask if they are fluent or disfluent in all types of communicative activity. It seems fairly clear that most of them would be "disfluent" in written language, if we are not trying to capture authentic speech in writing. It also seems clear that many of them might be disfluent in many types of public speaking. But this does not mean that they are disfluent in interactive (small) talk, where it is important that you are able to hesitate, change your mind, repeat for clarity, be flexible and non-categorical, and give continuous unobtrusive feedback. It seems fairly clear, that many of the functional means for achieving these goals have been labelled as "disfluencies", since they have no role in the kind of fluency required in written language or public speaking, but are concerned with the "communication management" (both ICM and OCM) required in fluent interactive communication.

My claim is thus that many "disfluencies" really are examples of mechanisms that are required for rational, efficient interactive communication, especially making use of processes on lower levels of awareness.

This justifies the question: Is the term disfluency (dysfluency) never appropriate? Two cases may be distinguished:

(i) Looking at one type of communication from the point of view of another, e.g. looking at interactive face-to-face communication from the point of view of written language (this is seldom, if ever, appropriate).

(ii) Comparing the ideal-normative function and goals of a particular communicative activity with actual performance, e.g. mistakes in spelling

or grammar in written language or exaggerated stammering or overlong pauses in interaction where a faster tempo was expected (this can be appropriate and be the basis for attempts at change).

4. Can terminology be changed in science?

Sometimes terminology changes in science. Usually, this signals a change of perspective or that an earlier view is seen as inappropriate or incorrect. "Phlogiston" disappeared and "oxygen" took over, when we changed our views of how what we now think of as oxidation, takes place. "Alchemy" became "chemistry", as part of an attempt to purge the field of practices considered to be less scientific. Charles Sanders Peirce changed the name of his philosophy from "pragmatism" to "pragmaticism" – "a name so ghastly that nobody will use it", when he was dissatisfied with some of the uses made of his philosophy. There are many other examples. Change of terminology is not uncommon.

Maybe it is time to change the terminology; abandon the term "disfluent" for more positive or neutral terminology, except in a few, well defined cases where really the goals of a particular communicative activity are not being met. For these cases, perhaps the word "dysfluency" could be used.

5. Conclusion

I have tried to argue that the notions of fluency and disfluency need to be relativized to type of communication. I have also argued that some interactive communicative practices that might seem "disfluent" from the perspective of public speaking or written language, in fact, in interactive communication, in most cases, are the opposite, i.e. features that help interactive communication become more fluent and efficient. Finally, I have also suggested that it might be good if our common terminology for the phenomena discussed, reflected this.

References

Allwood, J. 2013. A multidimensional activity based approach to communication. In I. Wachsmuth, J. de Ruiter, P. Jaecks, P. & S. Kopp (eds.) *Alignment in Communication*. Amsterdam: John Benjamins, 33-55.

Allwood, J., K. Grammer, S. Kopp & E. Ahlsén. 2006. A framework for analyzing embodied communicative feedback in multimodal corpora. LREC 2006 Workshop on Multimodal Corpora.

Allwood, J., J. Nivre & E. Ahlsén. 1990."Speech Management: on the Non-Written Life of Speech". *Nordic Journal of Linguistics* 13(1):3-48.

Eklund, R. 2004. Disfluency in Swedish human–human and human–machine travel booking dialogues. PhD thesis, *Linköping Studies in Science and Technology*, Dissertation No. 882, Department of Computer and Information Science, Linköping University, Sweden.

Kahneman, D. 2011. *Thinking Fast and Slow*. Farrar, Straus and Giroux.

Lickley, R. J. 2015. Fluency and disfluency. In M. Redford (Ed) *The Handbook of Speech Production*. Wiley-Blackwell, 445-469.

Lindblad, G. & J. Allwood. 2013. Prosodic expressions of emotions and attitudes in communicative feedback. In J. Allwood, E. Ahlsén, P. Paggio, C. Navarretta & K. Jokinen, K. (eds.) Proceedings of the 4th Nordic Symposium on Multimodal Communication, November 15-16, Gothenburg, Sweden, *NEALT Proceedings* 21:70-76.

Ward, N. 2006. Non-lexical conversational sounds in American English. *Pragmatics and Cognition* 14(1):129-182.