

# Ethical Considerations in the Use of Disfluencies in AI-Generated Speech

Ralph L. Rose<sup>a</sup>

Faculty of Science and Engineering, Waseda University, Tokyo, Japan

Keywords: AI Agents, Natural Language Generation, Disfluencies, Ethics.

Abstract: Disfluency occurs regularly in natural, everyday speech. Such phenomena as silent pauses, filled pauses (*uh*, *um* in English) and repairs occur with regular frequency and training data for natural language speech generation systems may be replete with these. This raises the question whether these items can be used productively in such systems in educational contexts and whether, as non-authentic phenomena, they can be used ethically. This position paper takes the five principles of AI ethics—beneficence, non-maleficence, autonomy, justice, and explicability (Floridi and Cowls, 2019)—as a starting point and proposes the Disfluency Instrumentality Audit as a tool to evaluate the ethical considerations of disfluency in AI-generated speech. The facilitative nature of disfluencies is explained in detail in order to argue for the potential beneficent nature of these phenomena in educational contexts. Sample scenarios are presented and discussed in order to illustrate how the audit might be used to evaluate the ethical considerations. Although presented here as applying to educational contexts, the audit would be applicable to wider contexts involving AI-speech generation for interactive agents.

## 1 INTRODUCTION


Human speech communication inevitably involves some periods during which the communication does not flow smoothly. It may be evidenced by momentary delays in communication such as short silent pauses or longer and more overt delays involving articulations like ‘um’ or self-repair sequences. Such disfluency occurs regularly in everyday speech (e.g., some estimates suggest an average of at least 10 instances per 1000 words for native speakers of English; e.g., Shriberg, 1994). Such disfluency in naturally-occurring speech has been disparaged in the popular literature (see, e.g., Erard, 2008; O’Connell and Kowal, 2004, for reviews). Yet numerous arguments have been made that disfluencies are a normally-occurring part of speech production—one way by which speakers actually construct *fluent* speech (cf., Allwood et al., 1990; Götz, 2013). Furthermore, far from being just meaningless delays in speech, listeners seem to capitalize on these delays, noticing where and in what form they occur in the speaker’s speech, in order to draw inferences about the speaker’s intended meaning. In this sense, they may be *facilitative* to the communication process.

Based on this understanding of disfluency in

speech communication, it is worth considering how disfluencies could be productively used in natural language speech generation to help facilitate the learning process in a student. In fact, a number of applications exist to insert disfluencies into AI-generated speech (cf., Qader et al., 2018; Yang et al., 2020; Dall et al., 2014; Wang et al., 2022; Matsunaga et al., 2022; Ohshima et al., 2015; Pfeifer and Bickmore, 2009). Many of these are presently proof-of-concept applications to test whether such inserted disfluencies may sound natural to listeners.

In any event, the use of such disfluencies to facilitate learning immediately raises certain ethical issues because disfluencies spoken by, say, an AI tutor, are arguably not authentic disfluencies (even if they sound natural), like those produced by a speaker struggling to decide what to say next. The AI tutor is either drawing from scripted speech or generating speech automatically, maybe sentence by sentence, before it is sent to a text-to-speech engine. Thus, there wouldn’t be any processing difficulty motive to insert an ‘um’ for example, or an in-line speech error which the AI tutor would need to cut off and then repair. In this sense then, disfluencies in the AI tutor’s speech might be seen as manipulating the learner.

A brief description of a more concrete scenario helps to illustrate some ethical issues. Consider an ap-

<sup>a</sup>  <https://orcid.org/0000-0002-8449-5666>

plication that generates lengthy responses to prompts for educational purposes (e.g., “Explain the water cycle in nature.”). If a system is generating a response and converting it to speech, there may be moments in the generation process when there are system delays (due to CPU interruptions from other processes, network slow-down, etc.). Presently, these might be reflected to the learner by temporary spinning icons or hourglass images. But what if these were replaced instead with a filled pause, ‘um...’ in the output speech stream. While this may seem to be a correlate of human speech processing, the causes are totally different and the distribution of these pauses may seem unintuitive or even random to the human listener. In this respect, they could become irritating and distracting to the listener. And if the listener were to start to emulate those patterns, it may disadvantage them in later communication with other human interlocutors.

The present work is intended to examine the ethics surrounding the use of disfluencies in AI-generated speech in such scenarios as these. It begins by first reviewing some of the existing literature on ethics in artificial intelligence and then based on it, proposing a simple instrument for evaluating the ethics of disfluencies in AI-generated speech in educational settings, called the Disfluency Instrumentality Audit (DIA). Thereafter, the paper summarizes the literature on disfluencies in communication with a special focus on its facilitative nature to listeners and argues for a careful use of such disfluencies in education. Finally, some sample scenarios are discussed in order to illustrate the application of the DIA in some circumstances.

## 2 ETHICS IN AI: FIVE PRINCIPLES

Discussions of ethics in artificial intelligence are plentiful (see e.g., Jobin et al., 2019; Floridi and Cows, 2019, for recent overviews and links to various industry and academic declarations). In recent years, A consensus seems to have formed in this literature that five basic principles underlie ethical considerations in AI, as follows (summarizing from Floridi and Cows, 2019).

- *Beneficence* revolves around the question of whether the AI system benefits the user and enhances their well-being.
- *Non-maleficence* focuses on whether the system harms the user. This is seen as independent from—and not simply a negation or inverse of—beneficence: A system conceivably may both benefit yet also harm a user. Of particular concern are

preservation of user’s privacy and security.

- *Autonomy* is the question of whether the system affords the user freedom and independence in their use of the system. Even where a user delegates some decision-making to a system, they should be free to regain control.
- *Justice* asks the question of whether the system treats the user fairly, avoiding bias in its treatment of the user.
- *Explicability* refers to whether the system’s actions can be explained and can be explored with respect to two sub-principles: *intelligibility* of the system’s actions and its *accountability*.

The application of these basic principles (or their equivalents) can be seen in some work that focuses on the ethical considerations in natural language generation (cf., Smiley et al., 2017; Henderson et al., 2018). However, at present, there is no work that addresses the specific issue of the use of disfluencies in AI-generated communication. The following section deals with this and proposes a tool for considering the ethics of systems that may use disfluencies.

## 3 DISFLUENCY INSTRUMENTALITY AUDIT

This section applies the five principles to the use of disfluencies in AI-generated communication in order to propose an instrument that may be used to evaluate the ethical considerations. This tool is called the “Disfluency Instrumentality Audit” (hereafter, DIA).

Here, the five principles are applied to the ethical consideration of the use of disfluencies in AI-generated communication in educational contexts. However, it is worth noting that the audit described below could be equally useful in other contexts as well (e.g., customer relations, user assistance systems, or social networking). However, the contexts discussed herein are all presumed to be in education.

Before explaining the audit in detail, it is necessary to identify two technical considerations which have some interaction with the ethical principles. This section will start by discussing these two considerations and then move on to explain the audit scheme in detail.

### 3.1 Technical Considerations

One consideration is the fundamental question of which kinds of disfluencies to use in AI-generated communication. The different types will be explained

in greater detail further below in Section 4, but each type does carry with it some different possible implications and effects on listeners. Silent pauses may be somehow seen as most 'natural' and thus least controversial. Filled pauses, may seem somewhat less natural, especially if the system's pronunciation of the filled pauses is marked. Repairs may seem the most unnatural as listeners might find them implausible and hence jarring. Thus, any system will need to consider first what inventory of disfluencies to use.

A second technical consideration is the manner in which the disfluencies arise in the communicative output. Here are four likely architectures for this.

### 3.1.1 Organically-Produced Disfluencies

If the training data for a machine learning-based natural language generation system contains numerous cases of disfluencies, then the generated output could (and theoretically, should) contain instances of disfluencies. These could be described as organically-produced disfluencies and are simply a reflection of the speech patterns extant in the training data. In this respect, the fact that a disfluency appears is effectively no more remarkable than if the system varies its use of 'soda' and 'coke' for 'soft drink', due to the naturally-occurring variation in the data. (Of course, whether this is pedagogically desirable is important, but is a different question.) In this architecture, ethical consideration of the system would have to focus on the training data and any questions it may raise.

### 3.1.2 Filtered Disfluencies

A machine learning-based system may output organically-produced disfluencies which are then monitored and filtered according to some predefined criteria. In this case, ethical consideration may focus more on the criteria than the training data.

### 3.1.3 Inserted Disfluencies

In some cases, the training data for a system may not contain disfluencies (it may be scripted speech, or the data may have been 'cleaned'). As such, the generated output would not contain disfluencies either. However, an insertion method could be used to intentionally insert disfluencies at specified locations according to predefined criteria. In this case, as above, ethical consideration would focus on the criteria. Note that this case is paralleled by a non-AI case where generated speech is simply manually scripted, with intentionally inserted disfluencies. Here again ethical consideration would focus on the criteria for disfluency insertion.

### 3.1.4 Filtered + Inserted Disfluencies

The final case is rather a hybrid of the previous two: AI-generated output may contain organically-produced disfluencies which are filtered according to some criteria. But, in a complementary fashion, other disfluencies may be inserted according to further criteria. Clearly, then both sets of criteria would need ethical consideration.

## 3.2 Five Principles in the DIA

Following the five principles of Floridi and Cowls (2019), the ethical considerations of the use of disfluencies in AI-generated speech for educational applications can be described as follows (see also Figure 1).

The question of beneficence should be addressed by asking the question whether the disfluencies used in the speech benefit the learner. In the educational context that is under consideration here, an affirmative answer here means that the disfluency somehow helps the user (here, learner) achieve some pedagogical goal. Of course, it's possible that the disfluency has other benefits which are not directly linked to a pedagogical goal, such as influencing the learner toward some positive affective state. To the extent that this facilitates the achievement of a separate pedagogical goal, then this could also be seen as beneficent.

Non-maleficence requires that the disfluency does no harm to the learner and, as further above, the fact that it may benefit the learner is not an ameliorating factor. In the case of disfluencies, an affirmative response here should mean that the disfluency does not have a negative impact (either singly or cumulatively with earlier disfluencies) on the learner such as influencing them toward a negative affective state. If they are distracted or even irritated by the presence of the disfluencies, this may be counterproductive and problematic.

Where there might be a risk of maleficence, the next principle becomes more important. Autonomy means that the learner is not constrained in any way by the disfluency. That is, they retain the agency to control their reaction to the disfluency, perhaps most simply by the capability to disable the feature as a settings option. Further, they should not not feel their speech is somehow restrained due to the AI agent's communication: They shouldn't be made to feel as if they themselves have to produce disfluencies in order to carry on communication with the agent.

The next principle is justice: whether the disfluency treats the learner in an unfair manner. An unfair treatment might be asserted by the user if the system

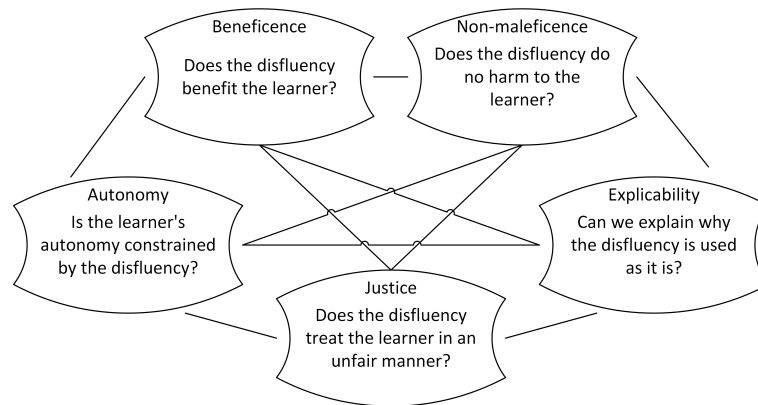


Figure 1: Disfluency Instrumentality Audit (DIA) summary.

seems to the user to be echoing the user’s own use of disfluencies in what might be seen as a disparaging act (e.g., as might be seen among children on the playground). Alternatively, justice might be questioned if the user feels that the AI agent is “talking down” to them—speaking in a condescending manner. Conversely, though, the principle of justice might be satisfied if the user feels a high degree of solidarity with the agent’s speech.

Finally, explicability revolves around the question of whether the use of disfluencies in the generated speech can be explained. In natural human speech, the explanation that such disfluencies occurred because the speaker was trying to decide what to say next or was overwhelmed with simultaneous demands for their thought processes may be taken as sufficient explanations. For AI-generated speech, such an explanation is probably not justified and thus not satisfactory. Instead, it will be necessary to explain the disfluencies with respect to other principles, perhaps as determined by the technical considerations described above. An explicable system should have an intelligible explanation behind it with the human agents that designed the system accountable for how the system uses disfluencies.

## 4 DISFLUENCIES

A starting point in the evaluation of the use of disfluencies in AI-generated speech is the first question of beneficence: Does the use of disfluencies benefit the user? In this section, after first defining some of the various disfluency types that have been examined in the past several decades of disfluency research, an overview of the various facilitative effects of disfluencies will be given to support an argument for the beneficence of their use in educational contexts.

### 4.1 Disfluency Types

Although disfluencies have been studied in many fields for several decades, there is not a widely agreed-upon typology of disfluency types and terminological usage varies significantly. The following definitions are one way of delineating these frequent features of speech (cf., Maclay and Osgood, 1959; Goldman-Eisler, 1961; Kasl and Mahl, 1965; Shriberg, 1994).

- *Silent pause* refers to an interval during which the speaker makes no speech sound. This does not include very short pauses that may be directly related to articulatory functions. Rather, they are silence of ‘excessive’ length. Exactly what counts as excessive has varied in the literature, though most researchers define a cut-off or threshold duration beyond which silence would be classified as a silent pause (e.g.,  $\geq 250$  ms in Goldman-Eisler, 1972).
- *Filled pause* refers to a non-lexical vocalization in a conventional form such as *uh/lum* in English or *e-le-to* in Japanese.
- *Prolongation* is a protraction of the articulation of some portion of a word such as *we-ll* and sometimes accompanied by a vowel change such as when the definite article *the* is pronounced as *thee* (e.g., *give me thee blue one*).
- *(Self-)Repair* is a sequence of speech in which a later chunk of speech should be understood as a correction to the immediately preceding chunk (e.g., *pick up the blue the red one*). Self-repairs most often comprise the repair of just a single word, but in some cases may extend over several words.
- *Repeat* is a sequence of speech in which a chunk of speech is repeated as-is (e.g., *pick up the the*



*the red one*). The repeated chunk may be repeated multiple times before speech resumes.

- *False start* refers to an utterance which is begun and then abandoned in favor of some different utterance (e.g., *is the class ... how long until the class ends?*).

Note that many of these disfluency types can be observed also in pathological cases such as stuttering or aphasia. However, the present work is focused on the non-pathological occurrence of these disfluencies.

## 4.2 Facilitative Disfluencies

The facilitative nature of disfluencies in speech communication has been shown in a number of studies, most notably in recent decades. This section summarizes several observations that have been made. For each observation below, a number of studies have provided evidence, though only a few citations will be given here as samples. For fuller treatments, see the respective sample citations.

One of the earlier known facilitative effects of disfluencies is related to word frequency. Listeners are more likely to judge words following filled pauses as being words with low contextual frequency (e.g., Beattie and Butterworth, 1979; Corley et al., 2006). Thus, while listeners find *nails* in (2a) odd (or unexpected), this oddness is ameliorated in (2b). Conversely, they would find *tongue* in (1b) more odd than in (1a).

- (1a) The hot drink burned his tongue.
- (1b) The hot drink burned his uh tongue.
- (2a) The hot drink burned his nails.
- (2b) The hot drink burned his uh nails.

In a possibly related finding, listeners also judge noun phrases following filled pauses as being more likely to introduce new entities to the current discourse (e.g., Arnold et al., 2003). Hence, listeners are more likely to expect that the *uh* in (3b) will be followed by a new entity and are surprised to hear an already given entity (*blue box*) repeated.

- (3a) Put the red ball on the blue box. Now put the green ball on the blue box.
- (3b) Put the red ball on the blue box. Now put the green ball on the uh blue box.

Researchers have also observed that listeners associate both silent and filled pauses with clause structure. They are more likely to judge a pause as occurring at a clause boundary than at a non-boundary (e.g., Bailey and Ferreira, 2003; Rose, 2019a). Thus, listeners find (4c) very difficult to understand compared to (4a). For similar reasons, they find (4b) the easiest to understand where the disfluency coincides with a

clause boundary. The same effect is found with either filled pauses or silent pauses.

- (4a) While the man hunted the deer ran into the woods.
- (4b) While the man hunted uh the deer ran into the woods.
- (4c) While the man hunted the deer uh ran into the woods.

Listeners also understand a speaker's self-repair faster when it is accompanied by a filled pause (e.g., Brennan and Schober, 2001). Thus, they will understand the instruction in (5b) faster and respond more accurately than in the case of (5a).

- (5a) Put the red the blue ball on the blue box.
- (5b) Put the red uh the blue ball on the blue box.

While the above facilitative effects are clearly related to the immediate processing and understanding of speech, higher level effects have also been observed. For instance, listeners recall information that occurs in a disfluent discourse better than when it occurs in fluent discourse (e.g., Fraundorf and Watson, 2011). Furthermore, at the sociolinguistic level, listeners regard speech containing filled pauses as more polite (e.g., Jansen and Janssen, 2013). This effect has long been discussed in the sociolinguistics and discourse analysis literature where these kinds of hesitation phenomena have been described as *mitigating devices* (cf., Levinson, 1983).

## 4.3 Question of User Design

One issue to note here is the question of user design. That is, it is easy to think that speakers use disfluencies in a volitional manner in order to communicate certain things to the listener such as warning of impending delay (cf., Maclay and Osgood, 1959; Clark and Fox Tree, 2002). However, this question is controversial and remains under debate. Some evidence exists to show that disfluencies are not used intentionally but rather occur symptomatically in the course of language processing difficulties (cf., Finlayson, 2014; Finlayson and Corley, 2012).

## 5 DISCUSSION: SAMPLE SCENARIOS

In order to illustrate the ethical issues described thus far in this paper as well as to provide some concrete examples for consideration, this section imagines three different practical educational scenarios in which disfluencies might be used as facilitative devices and discusses the ethical considerations in terms of the DIA. Each scenario will outline an imaginary

pedagogical situation where an AI system uses disfluencies in its generated speech to facilitate learning. Then the ethics of the situation will be examined using DIA. Note, however, that because the specific technical designs of such AI systems is beyond the scope of this paper, they will not be described here. But each scenario should be seen as encompassing a reasonable range of possible system designs.

### 5.1 Scenario 1: Second Language Learning

The first scenario takes place in the context of second language learning. An AI agent communicates with a learner in a tutor-student arrangement. The tutor guides the learner through pre-designed lesson plans that have specific pedagogical focal points and learning objectives. Interaction takes place in a relatively free form where the tutor may have nominal control of the interaction, but the learner may ask questions freely. As a teaching strategy, the AI system will insert disfluencies (hence, using an insertion or filtering+insertion architecture as described in Section 3.1 above) in order to highlight certain pedagogical points. For example, disfluencies may be inserted to highlight certain new or target vocabulary, to highlight a new or complicated syntactic structure, or around certain key points in order to make those points more salient and memorable to the learner.

Provided that learners could have the autonomy to disable this disfluency insertion feature, this scenario would, in principle, meet the five principles of DIA. The expected benefits are clearly based on the facilitative nature of disfluencies as described in Section 4 and this further makes their inclusion explicable. There appears to be no potential harm to learners and the disfluencies would be fair to learners because exposure to the disfluencies themselves would help to advance their capability to process such disfluencies more easily (cf., Griffiths, 1991).

### 5.2 Scenario 2: Lecture and Discussion

The second scenario takes place in the context of an academic lecture with follow-up discussion. An AI agent delivers a lecture at some length to learners and then interacts with them in a guided discussion based on the content of the lecture. Disfluencies occur in the lecture and interactive speech of the agent in an unsupervised manner (hence, using an organically-produced architecture as described in Section 3.1). The training data for the AI agent includes spontaneous speech data from the same age group as the target learners and the disfluencies (and other speech

patterns) are unfiltered so as to allow them to reflect the learners' peer group and to establish a more casual, familiar tone with the learners.

In this scenario, the agent would likely produce a natural (i.e., average) rate of disfluencies, while also including some common discourse markers associated with the age group (e.g., *you know, like*). However, when considering the five principles of DIA, it becomes clear that the benefit of the disfluencies is doubtful. They would have none of the facilitative effects described in Section 4, except perhaps establishing some solidarity with the learners as an 'in-group' member. Whether this really is beneficial is unknown. Furthermore, although the agent may use disfluencies in a manner that is typical of the training set *as a whole*, this may not necessarily match the disfluency pattern of every individual student. If there is wide discrepancy, a learner might actually regard the speech as condescending. Even if it is similar to their own, they may even fail to recognize it (people are actually not very capable of recognizing disfluent rates as shown in Lickley, 1995), and therefore may endure some harm as a result, or at least feel that the agent was treating them unfairly. Finally, since this approach relies on the organic use of disfluencies, there may be little explicability beyond an appeal to the overall disfluency patterns of the cohort represented in the training data. In short, this approach would seem to fail the ethical standards of DIA.

### 5.3 Scenario 3: Adaptive Tutoring

The third scenario involves an AI tutor that interacts with a learner on a specific topic following a predefined lesson plan. Disfluencies are used in the interaction in an adaptive manner. At first, the tutor uses minimal disfluencies and in a neutral manner. Then, as the conversation progresses and the tutor begins to recognize the individual learner's own disfluency patterns, the agent's own pattern may be adapted to approach and align with the learner's (e.g., as in Finlayson et al., 2010). The intended purpose would be to establish solidarity with the learner in a suitable manner under accommodation theory (Giles and Smith, 1979).

Under the ethical consideration guide by DIA, this case has some overlap with Scenario 2, above. However, the adaptive nature of the use of disfluencies means that there would be greater explicability and perhaps less risk of harm. Still, the actual benefit is questionable even if the use of disfluencies is fair to the learner. At the very least, this system should allow the learner the autonomy to disable these disfluencies if they find it irritable or degrading. Nevertheless, this

scenario is an improvement over Scenario 2, though still remaining in an ethical gray zone.

## 6 GENERAL DISCUSSION

This final discussion section describes some wider issues related to the DIA and disfluencies. While the present paper has focused on the use of disfluencies in AI-generated speech in educational contexts, the DIA would be applicable in wider, non-educational contexts. For instance, it could be used to make ethical decisions about interactive agents in customer relations, user assistance systems, or social networking. Furthermore, while the DIA has focused on speech generation, some disfluencies are used in text-based communication with similar facilitative effects (e.g., Rose, 2019b). Thus, DIA may also prove valuable when evaluating the ethics of disfluency use in chatbots or similar text-based communications.

In Section 4, the facilitative nature of disfluencies was described at length. Much of the work done on this facilitative nature has focused on filled and silent pauses. Further work is yet needed to establish more conclusively that all the other disfluencies show the same basic trends, though it has long been hypothesized that these hesitation phenomena are common in this regard. In the event that later work disproves this hypothesis, then the DIA may need to be revised or limited in its scope.

It should also be noted that some of the facilitative functions ascribed to disfluencies could also be handled by non-disfluent lexical means including such phrases as *well*, *let me see* and *I mean*. Where this is the case, it is proposed that the DIA would still be applicable to the ethical consideration of whether the use of such phrases in AI-generated speech.

## 7 CONCLUSIONS

This paper has outlined ethical considerations for the use of disfluencies in AI-generated speech in educational contexts. In this process, an adaptation of the five principles of Floridi and Cowls (2019) has been made as the Disfluency Instrumentality Audit (DIA). This tool may be used to consider whether certain applications that seek to use disfluencies with learners meets ethical standards. Information has also been given to show how disfluencies may serve a facilitative role in communication and that this may satisfy ethical concerns regarding the beneficence of considered systems. As system designers report on their system designs, it would be valuable if they

would explain how DIA facilitated (or failed to facilitate) the ethical development of their systems. Then, as technology advances and educational applications grow ever more sophisticated and capable (either knowingly or unexpectedly) of using disfluencies in AI-generated speech, it is hoped that the DIA will help system designers, system managers, and others consider carefully whether and how to use such disfluencies in an ethical manner to improve the educational experience of learners in various contexts.

## ACKNOWLEDGEMENTS

This paper is a part of research outcomes funded by Waseda University Grant for Special Research Projects (project number: 2022R-052).

## REFERENCES

- Allwood, J., Nivre, J., and Ahlsén, E. (1990). Speech management - on the non-written life of speech. *Nordic Journal of Linguistics*, 13(01):3–48.
- Arnold, J., Fagnano, M., and Tanenhaus, M. K. (2003). Disfluencies signal thee, um, new information. *Journal of Psycholinguistic Research*, 32(1):25–36.
- Bailey, K. G. and Ferreira, F. (2003). Disfluencies affect the parsing of garden-path sentences. *Journal of Memory and Language*, 49(2):183–200.
- Beattie, G. W. and Butterworth, B. L. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and Speech*, 22(3):201–211.
- Brennan, S. and Schober, M. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, 44(2):274–296.
- Clark, H. and Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, 84(1):73–111.
- Corley, M., MacGregor, L. J., and Donaldson, D. (2006). It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition*, 105(3):658–698.
- Dall, R., Tomalin, M., Wester, M., Byrne, W., and King, S. (2014). Investigating automatic & human filled pause insertion for speech synthesis. In *Proc. Interspeech 2014*, pages 51–55.
- Erard, M. (2008). *Um... Slips, Stumbles, and Verbal Blunders, and What They Mean*. Penguin Random House, New York.
- Finlayson, I. R. (2014). Testing the roles of disfluency and rate of speech in the coordination of conversation. Doctoral, Queen Margaret University, Edinburgh, Scotland, UK.
- Finlayson, I. R. and Corley, M. (2012). Disfluency in dialogue: an intentional signal from the speaker? *Psychonomic Bulletin & Review*, 19(5):921–928.

- Finlayson, I. R., Lickley, R. J., and Corley, M. (2010). The influence of articulation rate, and the disfluency of others, on one's own speech. In Watanabe, M. and Rose, R., editors, *DiSS-LPSS Joint Workshop 2010 - 5th Workshop on Disfluency in Spontaneous Speech and 2nd International Symposium on Linguistic Patterns in Spontaneous Speech*, pages 119–122, Tokyo, Japan. University of Tokyo; September 25–26, 2010.
- Floridi, L. and Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*, 1(1).
- Fraundorf, S. H. and Watson, D. G. (2011). The disfluent discourse: Effects of filled pauses on recall. *Journal of Memory and Language*, 65(2):161–175.
- Giles, H. and Smith, P. M. (1979). Accommodation theory. In Giles, H. and St. Clair, R. N., editors, *Language and social psychology*, pages 45–65. Basil Blackwell, Oxford, UK.
- Goldman-Eisler, F. (1961). A comparative study of two hesitation phenomena. *Language and Speech*, 4(1):18–26.
- Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech*, 15(2):103–113.
- Griffiths, R. (1991). The paradox of comprehensible input: Hesitation phenomena in l2 teacher-talk. *JALT Journal*, 13(1):23–41.
- Götz, S. (2013). *Fluency in Native and Nonnative English Speech*, volume 53 of *Studies in Corpus Linguistics*. John Benjamins Publishing Company, Amsterdam, Netherlands.
- Henderson, P., Sinha, K., Angelard-Gontier, N., Ke, N. R., Fried, G., Lowe, R., and Pineau, J. (2018). Ethical challenges in data-driven dialogue systems. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '18, page 123–129, New York, NY, USA. Association for Computing Machinery.
- Jansen, F. and Janssen, D. (2013). Uw reservering is eh komen te vervallen - experimenteel onderzoek naar het effect van gevulde pauzes in voicemail met slecht nieuws. *Tijdschrift voor Taalbeheersing*, 35(3):237–253.
- Jobin, A., Ienca, M., and Vayena, E. (2019). The global landscape of ai ethics guidelines. *Nature Machine Intelligence*, 1(9):389–399.
- Kasl, S. V. and Mahl, G. F. (1965). Relationship of disturbances and hesitations in spontaneous speech to anxiety. *Journal of Personality and Social Psychology*, 1.
- Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press, Cambridge, UK.
- Lickley, R. J. (1995). Missing disfluencies. In *Proceedings of International Congress of Phonetic Science*, volume 4, pages 192–195, Stockholm.
- Maclay, H. and Osgood, C. (1959). Hesitation phenomena in spontaneous english speech. *Word*, 15:19–44.
- Matsunaga, Y., Saeki, T., Takamichi, S., and Saruwatari, H. (2022). Personalized filled-pause generation with group-wise prediction models. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 385–392, Marseille, France. European Language Resources Association.
- O'Connell, D. and Kowal, S. (2004). The history of research on the filled pause as evidence of 'the written language bias in linguistics' (linell, 1982). *Journal of Psycholinguistic Research*, 33(6):459–474.
- Ohshima, N., Kimijima, K., Yamato, J., and Mukawa, N. (2015). A conversational robot with vocal and bodily fillers for recovering from awkward silence at turn-takings. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 325–330.
- Pfeifer, L. M. and Bickmore, T. (2009). Should agents speak like, um, humans? the use of conversational fillers by virtual agents. In Ruttkey, Z., Kipp, M., Nijholt, A., and Vilhjálmsson, H., editors, *Intelligent Virtual Agents. IVA 2009. Lecture Notes in Computer Science*, volume 5773, pages 460–466, Berlin, Heidelberg. Springer.
- Qader, R., Lecorvé, G., Lolive, D., and Sébillot, P. (2018). Disfluency insertion for spontaneous tts: Formalization and proof of concept. In Dutoit, T., Martín-Vide, C., and Pironkov, G., editors, *Statistical Language and Speech Processing*, pages 32–44, Cham. Springer International Publishing.
- Rose, R. L. (2019a). The structural signaling effect of silent and filled pauses. In Rose, R. and Eklund, R., editors, *The 9th Workshop on Disfluency in Spontaneous Speech (DiSS 2019)*, pages 19–22, Budapest, Hungary. ELTE Eötvös Loránd University; September 12–13, 2019.
- Rose, R. L. (2019b). The structural signalling effect of filled pauses during reading. Poster presentation at CUNY Human Sentence Processing Conference, Boulder, CO, USA.
- Shriberg, E. (1994). Preliminaries to a theory of speech disfluencies. phd, University of California, Berkeley.
- Smiley, C., Schilder, F., Plachouras, V., and Leidner, J. L. (2017). Say the right thing right: Ethics issues in natural language generation systems. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, pages 103–108, Valencia, Spain. Association for Computational Linguistics.
- Wang, S., Gustafson, J., and Székely, É. (2022). Evaluating sampling-based filler insertion with spontaneous TTS. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1960–1969, Marseille, France. European Language Resources Association.
- Yang, J., Yang, D., and Ma, Z. (2020). Planning and generating natural and diverse disfluent texts as augmentation for disfluency detection. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1450–1460, Online. Association for Computational Linguistics.