Speech Disfluency Guides Social Choices Through Inferred Readiness

Disfluency refers to interruptions in the fluent speech stream, such as filled pauses (*um/uh*). Prior work attributed disfluency to speaker uncertainty (Brennan & Schober, 2001), while others attributed it to cognitive difficulties (e.g., lexical retrieval difficulties; Arnold et al., 2003). Evaluatively, disfluent speakers are judged as less competent and trustworthy (Charoenruk & Olson, 2018). However, it remains unclear what specific speaker attributes listeners track during a conversation to form these impressions. Although some work hypothesized that disfluent speakers may seem less ready (Lickley, 2015) and more careful (Womack et al., 2012), there is currently no evidence that listeners actually perceive disfluent speakers as such, nor is it known whether such judgments depend on speaker expertise, a factor known to strongly shape social judgments (Pornpitakpan, 2004). We examined whether listeners inferred readiness, carefulness, and certainty from filled pauses, how these inferences were modulated by speaker expertise, and whether they drove future social decisions.

Exp.1: Attribute judgment. We extracted 10 spontaneous audio stimuli from corpora (5 female speakers, 5 casual speech), each 15-20 seconds long (M = 50.80 words), containing 4 filled pauses (*FP*) in accordance with naturalistic rates (Zhang, 2020) (e.g., Table 1). We next digitally removed the filled pauses to create a *No-FP* version. A 2x2 design crossed Filled Pause (*FP* vs. *No-FP*) with Speaker Expertise (Expert vs. Novice), which was described by text prior to presenting each auditory stimulus (Table 2). A total of 360 undergraduates rated the speaker on one of three attributes (readiness, carefulness, or certainty) using a 1-7 Likert scale (Table 3). As illustrated in Fig.1, Structural Equation Modeling (SEM) confirmed that filled pauses significantly reduce perceived readiness (p = .01) and certainty (p = .002), irrespective of expertise. However, we observed an interaction for carefulness (p = .04): filled pauses offered a social benefit for experts, who were judged to be more careful than novices when they produced filled pauses (p = .01) but not when they were fluent (p = .75). We attribute this positive bias to the higher reward valence of experts, which amplifies the perceived positivity of violating fluency expectation (Burgoon, 1993). The mixed (positive and negative) judgments are not uncommon in social evaluation, reflecting that these attributes are orthogonal and evaluated on distinct dimensions – a conclusion supported by the lack of correlation among them in our data.

Exp.2: Future decision. To test whether disfluency impacts social choices, we presented the same stimuli to 120 new participants using an implicit measure modeled after Fairchild et al. (2020). Participants were asked on a 1-7 Likert scale: “How likely would you be to ask this person again (as opposed to someone else) in a similar situation?” Mixed-effects regression (Fig.2) showed that participants were more likely to avoid disfluent speakers (p = .04). We further conducted a cross-sample SEM mediation analysis aggregating data from both experiments. Carefulness was trimmed from the model for better estimates and specification. Resulting model (Fig.3) showed that the decision to avoid was mediated by perceived readiness (p = .04) but not by certainty (p = .22). The direct effect of Filled Pause on Decision was non-significant (p = .12), indicating that listeners’ choices were not directly driven by the perception of disfluency but by the mental process of inferring a social attribute – specifically, reduced readiness.

Discussion. Our findings suggest that the evaluation of disfluency is not purely a bottom-up perception process but also involves top-down integration of speaker knowledge (Exp.1). Furthermore, by extending the research scope to decision-making, our findings challenge existing models that primarily link disfluency to speaker uncertainty. Instead, the results align with psycholinguistic views that filled pauses are signals of cognitive difficulty. Listeners appear to interpret interpolations as real-time cues of reduced readiness, and it is this inference of production difficulty that determines the speaker’s social utility as a future conversational partner.

Table 1. Transcript of sample audio stimulus (*FP* vs. *No-FP*) in Exps.1-2.

Well, with the superconductors, **(uh)** that's gonna make it, you know, so economical. **(Uh,)** we should have an enormous supply of it. They'll be able to **(uh)** grow circuitry that is so complicated. They'll be able to get these fantastic **(uh)** circuits, like say for a whole television set just on one chip. (53 words)

Table 2. Sample text for different levels of speaker expertise in Exps.1-2.

Expert	Jim is a professor of electrochemistry. He is expressing his view on the future of superconductors.
Novice	Jim is a college student. He is expressing his view on the future of superconductors.

Table 3. Social evaluation measure used in Exp.1.

Attribute	Item
Readiness	Disorganized in this conversation 1 2 3 4 5 6 7 Organized in this conversation Unprepared for this conversation 1 2 3 4 5 6 7 Prepared for this conversation Unready for this conversation 1 2 3 4 5 6 7 Ready for this conversation
Carefulness	Thoughtless in this conversation 1 2 3 4 5 6 7 Thoughtful in this conversation Careless in this conversation 1 2 3 4 5 6 7 Careful in this conversation Incautious in this conversation 1 2 3 4 5 6 7 Cautious in this conversation
Certainty	Uncertain in this conversation 1 2 3 4 5 6 7 Certain in this conversation Unsure in this conversation 1 2 3 4 5 6 7 Sure in this conversation Not confident in this conversation 1 2 3 4 5 6 7 Confident in this conversation

Fig.1. Results of Exp.1.

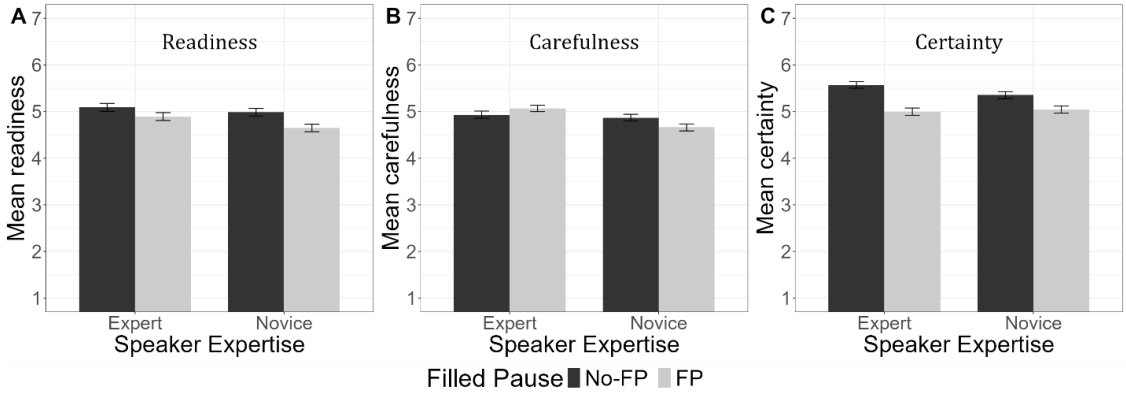

Fig.2. Results of Exp.2. (1 = *Unlikely to ask this person again*. 7 = *Likely to ask this person again*).

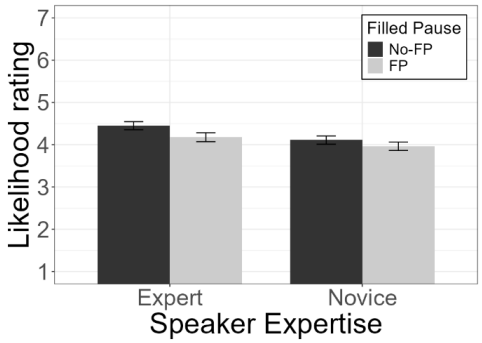

Fig.3. Mediation analysis for Exps.1-2.