Debunking the stigma of "disfluency": Perception of interpolation in naturalistic speech

Jonathan Him Nok Lee and Anna Papafragou (University of Pennsylvania)

Disfluency refers to interruptions in the fluent speech stream, such as filled pauses (*um/uh*). Disfluency impacts online sentence comprehension (Arnold et al., 2003). Other work has argued that disfluency is the consequence of cognitive difficulties, limited knowledge, or lying (King et al., 2018). Similarly, it has been suggested that disfluent speakers are judged as "less credible, less trustworthy, less easy to understand, and less expert" (Charoenruk & Olson, 2018). However, in these studies, the experimental stimuli were generally lab-recorded single sentences read aloud: interpolations were made distinctly salient. Additionally, the experiments often presented biased speaker identities, e.g., participants were told that the speakers were dishonest (King et al., 2018). In naturalistic speech, "disfluency" is ubiquitous and even speakers normally perceived as fluent produce so-called "disfluencies" (henceforth, interpolations) (Zhang, 2020). Moreover, interpolations may not be evaluated negatively if they reflect thoughtfulness of speech or effort in recalling facts. We posit that the prior perception of interpolations may be influenced by tasks, and naturalistic interpolations, in reality, are neutral for the evaluation of speakers. In two experiments, we used naturalistic stimuli to evaluate the perception (Exp.1) and social judgment (Exp.2) of *um/uh* interpolations to test these hypotheses.

Exp.1: Perception. We selected 10 audio stimuli of naturalistic speech from corpora (5 women, 5 men speakers; 5 formal, 5 casual speech), each 15-20 seconds long (m = 50.80 words), that contained 4 filled pauses (*with-filled-pauses*) in accordance with the naturalistic rates of filled pauses (Zhang, 2020) (e.g., stimulus in Table 1). We next digitally removed the filled pauses to create a *filled-pauses-removed* version. Additionally, we included 18 naturally produced *no-filled-pauses* stimuli from the corpus. In a control study, undergraduate English native speakers from a US university (n = 50) rated the *filled-pauses-removed* and *no-filled-pauses* stimuli on a scale from 1 (not smooth, not fluent, with many disruptions) to 7 (smooth, fluent, without disruption). Mixed-effects regression showed no significant differences between the ratings (p = .60; Fig. 1A): suggesting that our edited stimuli sounded natural. In the main study, another group of students from the same sample (n = 100) followed the same procedure but rated the *with-filled-pauses* and *no-filled-pauses*. Results showed no significant differences between the ratings (p = .07; Fig. 1B): native speakers did not perceive the speech with multiple interpolations as disfluent.

Exp.2: Social judgment. In a 2x2 design, we crossed interpolation in the critical auditory stimuli from Exp.1 (*with-filled-pauses* vs. *filled-pauses-removed*) and speaker expertise (expert vs. novice) as described in text presented before the auditory stimulus (Table 2). We explored whether speaker evaluations would vary with the presence/absence of interpolations, in ways that might interact with speaker's knowledge. New undergraduate participants (n = 120) rated competence, warmth, and trustworthiness, each evaluated with 6 questions involving a 1-7 Likert scale (McCroskey & Teven, 1999). Figure 2 shows the judgment results. After parceling the measurement items into 9 variables, Structural Equation Modeling (Figure 3) showed that experts were judged to have significantly higher competence (p = .01) and warmth (p = .04) than novices but no significant differences in their trustworthiness (p = .07). By contrast, interpolation did not have significant effects on any of the constructs, nor did it interact with expertise (all ps > .05).

Discussion. Our findings argue against the idea that interpolation creates negative biases towards the speakers. First, interpolation is not necessarily perceived as "disfluency" (Exp.1). Second, speaking with interpolations does not indicate reduced competence, trustworthiness, or warmth, even for speakers who are generally thought to be less capable (i.e., novices as opposed to experts). We conclude that our understanding of linguistic meaning (including social meaning) has much to gain from the experimental study of speech generated beyond the labs.

Table 1. Sample stimulus (with-filled-pauses vs. filled-pauses-removed).

(Um), when we have to prove something like specific intent, **(um)**, or we have to prove identity, because he's disputing identity, at least one of the cases. **(Um)**, then, we are allowed to bring in prior similar conduct, **(um)**, where he acted in an identical way, or where the victims were in a similar situation. (55 words)

Table 2. Sample text for different speaker expertise.

Expert	Rebecca is an experienced Philadelphia attorney. A client is currently involved in a lawsuit. Rebecca is offering an opinion on the case.
Novice	Rebecca is working as an intern in a law firm. A client is currently involved in a lawsuit. Rebecca is offering an opinion on the case.

Fig. 1. Results of Exp.1.

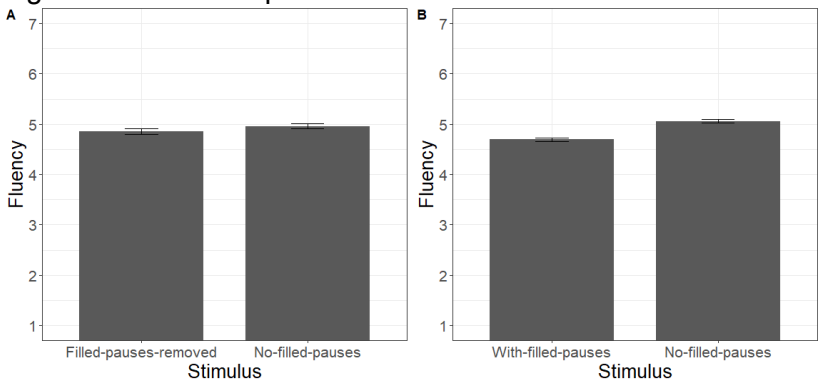

Fig. 2. Results of Exp.2.

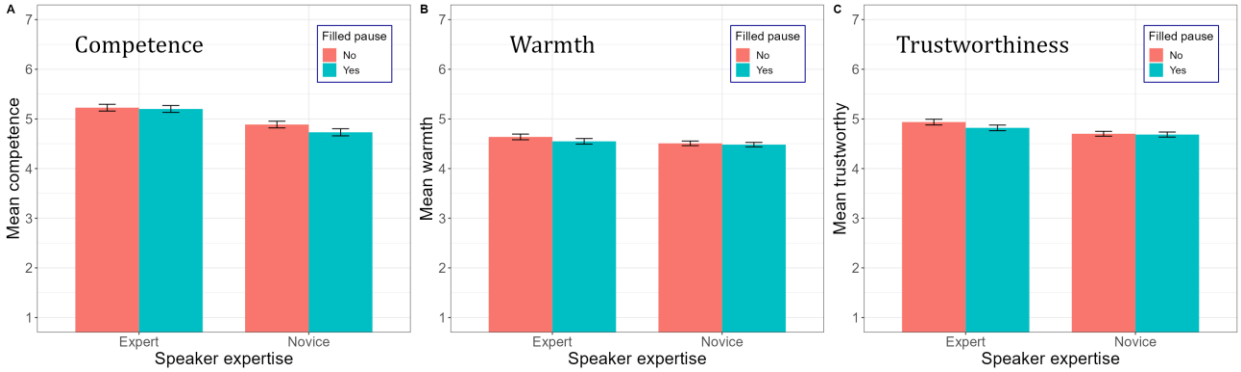

Fig. 3. Path diagram of the Structural Equation Modeling for Exp.2.