



The Perception of Overlapping Speech: Effects of Speaker Prosody and Listener Attitudes

Katherine Hilton

Department of Linguistics, Stanford University, USA

k Hilton@stanford.edu

Abstract

Speakers use overlapping speech to achieve a range of interactional moves. Competitive overlaps, or interruptions, challenge an interlocutor's control of the conversational floor, while non-competitive overlaps, like back-channeling and co-constructed discourse, communicate engagement with the conversation and ratify the interlocutor's right to be speaking. Being able to evaluate the intentions behind moments of overlap is critical for interlocutors, as well as researchers seeking to model human-human interaction. Researchers have analyzed the acoustics of overlapping speech in order to understand what determines whether an overlap is heard as competitive or non-competitive. They have overwhelmingly found that prosodic prominence plays an important role; incoming overlaps with higher pitch and intensity are more competitive or interruptive. However, no research has directly tested whether and how listeners use prosodic cues to evaluate moments of overlap. Furthermore, much of the current research on classifying overlapping speech ignores listener variability. The present study uses a perception experiment with 500 participants to test the effects of speaker prosody and listener attitudes on the evaluation of overlapping speech. The results demonstrate that prosodic prominence does significantly affect evaluations of overlapping speech, but it is mediated by the listener's own interactional style and attitudes toward overlapping speech.

Index Terms: overlapping speech, perception, prosody, interactional style, conversation

1. Introduction

Researchers analyzing human interaction have deliberated over whether overlaps, or moments when more than one person talks at the same time, are normal elements of cooperative interactions or signs of discord and dysfunction. Early models of conversation and turn-taking assumed the latter and characterized optimal exchanges as containing neither gaps nor overlaps [1]. These assumptions have been challenged by research which shows that, although it is typical for one person to speak at a time, overlapping speech is quite common, especially at turn boundaries [2, 3].

One reason that overlapping speech is so common, despite listeners' ability to reliably predict the end of a speaker's turn before the onset of a pause [4], is that interactants intentionally use overlaps to accomplish a wide range of communicative goals. Some uses of overlap challenge the current speaker's control of the conversational floor, while others seem to ratify that speaker's turn and show engagement with what they are saying. How people involved in an interaction come to interpret the communicative functions of overlapping speech, whether as a challenge or affirmation, has consequences for the progression

of the interaction and the relationship between interlocutors. The present study uses a perception experiment to analyze the ways that prosody and general attitudes toward overlapping speech shape listeners' evaluations of overlaps and the interactions where they occur.

2. Classifying and evaluating overlaps

Speech scientists and computational linguists have sought to model the dynamics of overlapping speech in human-human interaction for a variety of technological applications, such as improving virtual conversational agents [5]. Specifically, much of this research focuses on identifying the acoustic and temporal cues that distinguish turn-competitive overlaps from non-competitive ones. Overlapping speech is turn-competitive when one speaker initiates overlap in order to take control of the conversational floor. Early work on turn-taking proposed that the timing of an overlapping utterance determined its turn-competitiveness. An incoming utterance is turn-competitive if it begins after the onset of another speaker's utterance but before the transition relevance place (TRP) which signals a possible completion point of that turn [6, 7, 8]. Other work emphasizes the role of prosody in marking an overlap as turn-competitive. In particular, [9] argues that incoming overlaps are characterized as turn-competitive when they are relatively high in pitch and intensity, regardless of where they occur with respect to another speaker's utterance. More recent work demonstrates that the position of an incoming utterance, in addition to its acoustic properties, is important for signaling turn-competitiveness [10, 11, 12]. For example, [11] finds that turn-competitive overlaps can occur in any position but have different acoustic properties depending on where they occur.

Despite the value of these contributions to our understanding of overlapping speech, important gaps remain. The authors in [11] point out that there is a need for perceptual studies which directly test the hypotheses put forth by production studies on overlapping speech. Perception experiments have the potential to shed light on whether and how listeners use prosodic and temporal cues, among others, to evaluate the turn-competitiveness of overlapping speech. [11] also notes that, although focusing on the turn-competitiveness of overlaps has been fruitful, this binary classification might be too simplistic, potentially obscuring the full range of social actions speakers achieve through overlaps. For example, [13] explores several different functions of overlap, such as choral repetitions, collaborative completions, and requests for or offers of clarification or confirmation, which are not merely non-competitive but cooperative in that they build rapport and facilitate communication

Another important aspect of overlapping speech which has been left out of recent quantitative work on overlaps is the great

deal of interspeaker variation. Speakers differ from each other in the amount of overlap they use as well as the social significance they attribute to different types of overlap [14, 15, 16, 17]. What counts as turn-competitive for one speaker might be cooperative for another, and moreover, there is variation in whether turn-competitive overlaps are considered positive or negative occurrences. Among speakers with a “high-engagement conversational style” who use overlaps frequently, an interlocutor’s failure to engage in persistent turn-competitive overlaps can come across as “uncooperative, withholding, even sulking” [13, p. 125].

3. Research questions

These insights from conversation analysis and the lack of perception studies on overlapping speech raise questions that I seek to address in this paper. The first question is whether and how listeners use prosodic cues to evaluate the social actions of an overlap. For example, if an overlap occurs turn-medially before the TRP, will it be perceived by listeners as turn-competitive regardless of its phonetic design? And if the pitch height and intensity, or the prosodic prominence, of an incoming overlap determines its turn-competitiveness, is there a linear relationship between an overlap’s relative prominence and its perceived competitiveness? In other words, does the perceived competitiveness of an overlap increase incrementally as the overlap becomes louder and higher in pitch? Or is there a threshold for relative pitch and intensity, above which overlaps are more likely to be perceived as competitive? Furthermore, how does interspeaker variation in interactional style and attitude toward overlaps affect the evaluation of the social actions achieved through overlapping speech? Do all listeners, regardless of their personal attitudes and style, generally agree on whether or not specific overlaps seem competitive or cooperative? Or is there a qualitative difference in how some listeners orient to prosodic cues for turn-competitiveness. To address these questions, this paper presents the results of a perception experiment which tests the effects of prosodic prominence and listener attitudes on the evaluation of overlaps.

4. Methods

This study uses the Matched Guise Technique [18] and a between subjects design to test the effects of prosody and listener attitudes on the perception of overlapping speech. 500 native speakers of American English completed a survey distributed through Amazon Mechanical Turk in which they were each asked to listen to a single audio recording of two people interacting. Afterward, they answered a series of questions about themselves and the conversation they heard.

4.1. The stimuli

During the survey, participants listened to one of four 25 s audio clips. The four stimuli contain a back-and-forth interaction between a female (Sarah) and male speaker (Michael). Stimuli are identical except for the timing and prosodic prominence of Sarah’s final utterance, which is illustrated in Figure 1.



Figure 1. Duration of each speaker’s turn in seconds.

In the stimulus which serves as the control for this experiment, there is no overlapping speech; Sarah’s final utterance occurs immediately after Michael’s without a pause. In the other three stimuli, the timing of Sarah’s 1.4 s final utterance (represented as a block with dashed outline in Figure 1 below) was altered in Praat [19] so that it begins 350 ms after the onset of Michael’s final utterance, which also has a duration of 1.4 s. Positioning the incoming overlap more than 120 ms from the onset of Michael’s turn makes it clear to listeners that Michael’s utterance was already in progress when Sarah’s comes in [20]. The overlap also begins well before the end of Michael’s turn or any TRPs [1]. This means that, at least in terms of its position, Sarah’s overlap has the potential to be heard as competitive.

The position of the overlap remains constant across all three of the stimuli which contain overlap, but the prosodic prominence of Sarah’s final utterance varies. Previous research has found that turn-competitive overlaps tend to be marked with increased pitch height and intensity. In order to test the effects of these prosodic cues, I manipulated the pitch height and intensity of Sarah’s final utterance in Praat, but kept the pitch and intensity of Michael’s utterance constant. For the stimulus I refer to as MATCH, the final utterances occur in overlap, but Sarah’s pitch (233 Hz) and intensity (66 dB) remain unchanged from the control stimulus and are comparable to Michael’s. In the third stimulus (HIGH), the average intensity of Sarah’s final utterance was increased to 72 dB, and the level of each pitch point was increased by 10%, causing Sarah’s final utterance to be more prosodically prominent than Michael’s. And in the fourth stimulus (LOW), the average intensity of Sarah’s final utterance was decreased to 58 dB and the level of each pitch point was decreased by 10%, yielding an utterance that is less prosodically prominent than Michael’s.

The stimuli were created from a longer recording of two speakers who were both 30 years-old, white and from the Northeastern U.S. The speakers were asked to act out a loose script about their experiences at a restaurant. A transcript of the recording is presented below:

- S:** Oh have you been to Donatello’s since they reopened?
- M:** No I didn’t even know they closed down.
- S:** Yeah they were on one of those like restaurant makeover reality TV show things.
- M:** Really?
- S:** Oh yeah. Whole new menu. New chef. It’s actually pretty good now.
- M:** Last time I was there the service was slow and it was just gross.
- S:** Oh I know.
- M:** I mean it was really dirty.
- S:** It’s actually really good now.

Using loosely scripted recordings allowed for precise control of the content and structure of the conversation. In creating the script, it was important for the content to be substantial and familiar enough to listeners so that they could imagine the interaction and the speakers [21]. The speakers also take roughly the same number of turns of equivalent length and substance, and there are several turn changes within the dialogue. This ensures that, across the interaction more

generally, neither speaker is seen as having greater control of or claim to the conversational floor. Finally, it was essential to end the recording immediately after the overlap and not include subsequent turns by either speaker. Which speaker took the next turn after the overlap, what s/he said, and how s/he said it would provide the listener with clues about how the interlocutors themselves interpreted the overlap.

4.2. The survey

Participants were asked to listen to the recording of Sarah and Michael interacting and then answer questions about each of the speakers and their relationship. After listening to only one of the four stimuli, participants answered a series of questions about Sarah and Michael on one page and then answered questions about themselves on a separate page. The questions were designed to assess whether participants perceived the particular overlap they heard as competitive or cooperative and whether they generally hold positive or negative attitudes toward overlapping speech. The survey questions are listed below.

4.2.1. Questions about the stimulus

1. How well do you think Sarah and Michael get along?
2. How friendly do Sarah and Michael seem?
3. How engaged in the conversation did each person seem?
4. How likely is it that each person felt listened to?
5. How likely is it that each person felt interrupted?
6. At any point did it seem that Sarah or Michael was trying to interrupt?
7. How much do you think each person was controlling the conversation?

4.2.2. Questions about the survey-taker

8. How would your friends or people close to you describe you: (reserved; chatty; a big talker; quiet)?
9. How comfortable are you in conversations where: (there are times when no one is talking; more than one person talks at the same time)?
10. If someone talks at the same time as you during a conversation how likely is it that they are: (being rude or inconsiderate; interested and engaged in the conversation)?

For each question, participants used a sliding scale from 0-100 to indicate the extent to which they perceived that particular attribute to be true. For questions 2-7, there were separate scales for Sarah and Michael, and for 8-10 there were separate scales for each option in parentheses. To control for the possibility that some participants might complete the survey without listening to the recording, participants were asked to indicate the topic of the conversation from a list of nine options. The 5/500 participants who failed to select *restaurants* had their results excluded from the analysis.

Linear regressions tested the effects of prosodic prominence and attitudes toward overlap on the evaluations of Sarah and Michael's interaction. In each statistical model, the ratings for Questions 1-7 about Sarah and Michael serve as the dependent variables, with separate models for each question. The independent variables included the participants' ratings of their own interactional style and attitudes toward overlaps (Questions 8-10), the stimulus they heard, and the interaction between the

stimulus and ratings from Questions 8-10. Across all models, the control stimulus which contains no overlapping speech was treated as the reference level.

5. Results

The results of this study support the hypothesis that the relative prosodic prominence of an incoming overlap has a significant effect on its perceived competitiveness. The results also demonstrate that a listener's attitudes toward overlapping speech influence their evaluations of interactions that contain overlaps.

5.1. Effects of prosody

Contrary to what some previous studies might predict, the relative prosodic prominence of Sarah's overlap did not exhibit a linear relationship with its perceived turn-competitiveness. Instead, it seems that MATCH was perceived as the most turn-competitive and interruptive. Participants who heard the HIGH stimulus were not more likely to perceive Sarah as trying to interrupt or Michael as feeling interrupted compared to the control. However, in response to the MATCH stimulus in which Sarah and Michael's overlapping utterances were comparable in pitch height and intensity, participants were more likely to perceive that Michael felt interrupted ($p = 0.001$), that Sarah was trying to interrupt ($p = 0.04$), and that Sarah was controlling the conversation ($p = 0.001$). Although listeners did not appear to perceive HIGH as the most interruptive, participants who heard this stimulus were less likely to perceive that Michael felt listened to ($p = 0.03$).

While participants tended to converge on perceiving MATCH as turn-competitive, the response to LOW was varied. In comparison with those who heard the control, participants who listened to LOW were significantly more likely to guess that Sarah was trying to interrupt Michael ($p = 0.01$). But they were also more likely to perceive that Michael was trying to interrupt Sarah ($p = 0.01$), that Sarah felt interrupted ($p = 0.03$), and that Michael was controlling the conversation ($p = 0.03$). These responses are surprising and somewhat difficult to interpret in isolation. Their implications become clear, though, when considered in the context of listeners' variable attitudes toward overlapping speech, discussed in Section 4.3.

5.2. Interaction between prosody and attitudes

Not only did listener attitudes influence the degree to which overlaps were perceived as competitive or cooperative, but they also shaped listeners' interpretations of each stimulus in qualitatively different ways. The effect of listener attitudes is particularly apparent in its interaction with LOW, the stimulus where Sarah's incoming overlap has reduced pitch and intensity. Section 5.1 notes the peculiarity of the reactions to this stimulus – participants seemed to disagree over whether it was Sarah, Michael, or both who were interrupting and being turn-competitive in this stimulus. Perhaps even more intriguing is that, for a specific group of speakers, Sarah and Michael's interaction was perceived more positively in the LOW stimulus than in the control which contains no overlap. Participants who hold favorable attitudes toward overlapping speech in general were more likely to perceive Sarah's low intensity, low pitch overlap as a cooperative sign of engagement, while the reverse is true for participants who hold unfavorable attitudes toward overlapping speech.

In comparison with other listeners, those who reported that they generally view overlapping speech as inconsiderate and rude were more likely to perceive that Sarah was trying to interrupt Michael in the LOW stimulus ($p = 0.04$). Conversely, participants who generally view overlapping speech as a sign of interest and engagement and who reported being comfortable in conversations where more than one person talks at the same time were less likely to perceive Sarah as trying to interrupt Michael ($p = 0.001$) and were less likely to perceive Michael as feeling interrupted ($p = 0.01$). Participants who reported being more comfortable in conversations with overlap were also less likely to perceive that Michael was trying to interrupt Sarah ($p = 0.02$) or that Sarah felt interrupted by Michael ($p = 0.02$). Although there was no main effect of stimulus on participants' perceptions of how well Sarah and Michael get along, there was a significant interaction between this attribute and aspects of the participants' attitudes and interactional styles. Speakers who described themselves as quiet were less likely to perceive that Sarah and Michael get along well and were less likely to perceive Sarah as engaged in the conversation during LOW. In contrast to this, participants who view overlapping speech as a sign of interest and engagement were more likely to guess that Sarah and Michael get along well in the LOW stimulus than in the control with no overlap.

6. Discussion

The results of this study indicate that the relative pitch height and intensity of an incoming overlap influences whether it is heard as turn-competitive or cooperative. A somewhat surprising finding is that the stimulus MATCH was consistently perceived as the most interruptive and not the stimulus HIGH which contains the most prosodically prominent incoming overlap. It could be that this is a peculiarity of the particular stimuli used in this study. Another possibility is that listeners use prosodic cues from both speakers to evaluate the social actions of overlapping speech.

Although the pitch height and intensity of Michael's utterance remained constant across all stimuli, it was relatively low in prominence during the HIGH stimuli when compared with Sarah's utterance but comparable to Sarah's in the MATCH stimulus. The authors in [9, 10] discuss how speakers who are interrupted have two options for responding to the interruption; they can yield the floor or they can assert their control of the floor by increasing their intensity and decreasing their speech rate. Because Michael's interrupted utterance was relatively quieter than Sarah's in the HIGH stimulus, it could have seemed to participants that he was yielding the floor to her. Participants might interpret this apparent decision to yield the floor as a sign that Michael recognized Sarah's right to take a turn and continue her idea from a previous turn. Participants might be less likely to interpret Sarah's overlap as interruptive if the relative phonetic design of Michael's utterance offers cues that Sarah has a right to claim the floor. Even though the HIGH stimulus was not perceived as interruptive, per se, it seems that participants still interpreted it as turn-competitive. Participants listening to this stimulus were less likely to perceive that Michael felt listened to by Sarah. This finding could reflect the perception that Michael's decision to yield the floor left the information he was trying to communicate in the final utterance unacknowledged. By contrast, participants might have perceived the MATCH stimulus to be particularly interruptive because of the comparable intensity between the two speakers. Michael's apparent reluctance to yield the floor could signal

that he did not think Sarah had the right to take control of the floor, thus increasing the perception among study participants that she did not have the right to take a turn and was being interruptive.

The results of this study also demonstrate that the social dynamics of a single moment of overlap are evaluated qualitatively differently by listeners depending on their general attitudes toward overlaps. Evaluations of the LOW stimulus, where Sarah's incoming overlap had reduced pitch height and intensity, exhibited the most inter-listener disagreement over whether or not the overlap was turn-competitive, and if so, whether turn-competition was a good or a bad thing. Participants who are comfortable with overlaps and generally view overlapping speech as a sign of interest and engagement in an interaction tended to perceive the interaction between Sarah and Michael in LOW more favorably than in the control stimulus which contains no overlapping speech. The reverse effect was found for participants who describe themselves as quiet and who generally view overlapping speech as inconsiderate and rude. This finding supports ideas presented in [13] that some speakers prefer interactions with a certain amount of overlapping speech.

7. Future work

The results presented in this paper represent the first step in a larger project investigating the perception of overlapping speech. This ongoing research expands upon the present study and addresses some of its limitations by increasing the number of speakers and dialogues used in creating the stimuli, broadening the range of linguistic factors manipulated across each stimuli, and including additional measures of interactional style and listener attitudes.

8. Conclusion

The findings presented in this study support the notion, raised by [22], that the turn-competitiveness of overlaps is not determined by objective physical properties of speech alone. Rather, turn-competitive aspects of overlaps are subjective construals about speaker motives and who has the right to be speaking at a particular time about a particular topic. This study offers new challenges to speech scientists working on virtual conversation agents and other technical applications of the research on human-human interaction. It is appealing to concentrate efforts on classifying overlaps according to acoustic and temporal properties alone, because these properties are straightforward to measure. However, building more human-like virtual conversation agents will ultimately require us to grapple with the complexity of social attitudes and subjectivity. Continuing to incorporate these difficult to measure properties into our models of human-human interaction offers the promise of enriching future human-computer interactions.

9. References

- [1] H. Sacks, E. Schegloff, and G. Jefferson, "A simplest systematics for the organization of turn-taking for conversation," *Language*, vol. 50, no. 4, pp. 696-735, 1974.
- [2] E. Shriberg, A. Stoelcke, and D. Baron, "Observations on overlap: Findings and implications for automatic processing of multi-party conversation," *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech)*, pp. 1359-1362, 2001.

- [3] M. Heldner and J. Edlund, "Pauses, gaps and overlaps in conversations," *Journal of Phonetics*, vol. 38, no. 4, pp. 555-568, 2010.
- [4] M. Casillas and M.C. Frank, "Cues to turn boundary prediction in adults and preschoolers, in S. Brown-Schmidt, J. Ginzburg, and S. Larsson (Eds.), *Proceedings of SemDial (SeineDial): The 16th Workshop on the Semantics and Pragmatics of Dialogue*, pp. 61-69, 2012.
- [5] D. Reidsma, I. de Kok, D. Neiberg, S. Pammi, B. van Straalen, K. Truong, and H. van Welbergen, "Continuous interaction with a virtual human," *Journal of Multimodal User Interfaces*, vol. 4, pp. 97-118, 2011.
- [6] E. Couper-Kuhlen, *English Speech Rhythm: Form and Function in Everyday Verbal Interaction*, John Benjamins, Amsterdam/Philadelphia, 1993.
- [7] G. Jefferson, "Two explorations of the organization of overlapping talk in conversation: Notes on some orderliness of overlap onset," *Tilburg Papers in Language and Literature* 28, Tilburg University, Netherlands, 1983.
- [8] D.H. Zimmerman. and C. West, "Sex Roles, interruptions, and silences in Conversation," *Language and Sex: Difference and Dominance*, Stanford University Press, Stanford, CA, pp. 105-129, 1975.
- [9] P. French and J. Local, "Turn-competitive incomings," *Journal of Pragmatics*, vol. 7, no. 1, pp. 17-38, 1983.
- [10] B. Wells and S. MacFarlane, "Prosody as an interactional resource: Turn-projection and overlap," *Language and Speech*, vol. 41, no. 3-4, pp. 265-294, 1998.
- [11] E. Kurtić, G.J. Brown, and B. Wells, "Resources for turn competition in overlapping talk," *Speech Communication*, vol. 55, no. 5, pp. 721-743, 2013.
- [12] S.A. Chowdhury, M. Danieli, and G. Riccardi, "The role of speakers and context in classifying competition in overlapping speech," in *INTERSPEECH 2015 – 16th Annual Conference of the International Speech Communication Association, September 6-10, Dresden, Germany, Proceedings*, pp. 1844-1848, 2015.
- [13] D. Tannen, "When is an overlap not an interruption? One component of conversational style," in R.D. Di Pietro, W. Frawley, and A. Wedel (Eds.) *Selected Papers of the First Delaware Symposium on Language Studies*, pp. 119-129, 1983.
- [14] D. Tannen, *Processes and Consequences of Conversational Style*, Pd.D. dissertation, University of California, Berkeley, 1979.
- [15] A. Berry, "Spanish and American turn-taking styles: A comparative study," *Pragmatics and Language Learning*, vol. 5, pp. 180-190, 1994.
- [16] H.Z. Li, "Cooperative and intrusive interruptions in inter- and intracultural dyadic discourse," *Journal of Language and Social Psychology*, vol. 20, no. 3, pp. 259-284, 2001.
- [17] D. Kimura, "Utterance overlap and long silence among the Baka Pygmies: Comparison with Bantu farmers and Japanese university students," *African Study Monographs*, vol. 26, pp. 103-121, 2001.
- [18] W.E. Lambert, R.C. Hodgson, R.C. Gardner, and S. Fillenbaum, "Evaluational reactions to spoken language," *Journal of Abnormal and Social Psychology*, vol. 60, pp. 44-51, 1960.
- [19] P. Boersma and D. Weenink, Praat: doing phonetics by computer [Computer program], version 6.0.15, 2016.
- [20] M. Heldner, "Detection thresholds for gaps, overlaps, and no-gap-no-overlaps," *Journal of the Acoustical Society of America*, vol. 130, pp. 508-513, 2011.
- [21] K. Campbell-Kibler. "The nature of sociolinguistic perception," *Language Variation and Change*, vol. 21, pp. 135-156, 2009.
- [22] A. Bennett, "Interruptions and the interpretation of conversation," *Proceedings 4th Annual Meeting of the Berkeley Linguistics Society*, Berkeley, CA, 1978.