DOI: 10.17469/O2109AISV000007

MARTINA ROSSI

Gender influence on phonetic turn-taking cues at potential transition locations in German

Phonetic cues have been found to play a fundamental role in the turn-taking mechanism by helping to signal to the interlocutor(s) the intentions of the current speaker for the upcoming turn. Since previous sociolinguistic research described the existence of gender specific behaviors in interactions, it could be the case that interlocutors' genders might influence the way different speakers use turn-taking cues to signal their conversational intentions. This research aims at investigating the influence of the gender of the speaker and the gender of the interlocutor on phonetic turn-taking cues, i.e., F0 movements, intensity and segmental duration, towards potential transition locations in two-party conversations between German native speakers. The results suggest that both the speakers' and the interlocutors' genders might influence the way potential transition locations are phonetically marked in conversations.

Keywords: turn-taking, German, interaction, phonetic variation, gender influence.

1. *Introduction*

Conversational interactions are a fundamental part of human social behavior. On a daily basis, individuals are involved in face to face or mediated conversations, producing about 1200 talk spurts amounting to around 2 to 3 hours of speech (Mehl, Vazire, Ramírez-Esparza, Slatcher & Pennebaker, 2007; Levinson, Torreira, 2015). It is clear from even the most casual exchanges that people in conversation do not speak at random: in general, it can be expected that there will be one person talking at a time and that the interlocutor(s) will wait for them to finish speaking to launch their turn, usually trying to keep silent gaps and speech overlaps reduced to a minimum (Sacks, Schegloff & Jefferson, 1974). In order to achieve a smooth exchange of turns, speakers seem to be able to predict the approach of a potential transition location (PTL from now on), i.e., a point in the current speakers' turn when speaker change becomes a possibility (Schiffrin, 1987; Transition Relevance Places, Sacks et al. 1974; Potential Turn Boundaries, Zellers, 2016), which allows them to start planning their next conversational move. This can either be to take up the next turn, to remain silent and let the speaker continue, or to produce a non-interrupting backchannel to signal attention. In fact, interlocutors do not wait for the current speaker to finish with their turn before deciding what to do next; they start encoding their following turn while the current one is still ongoing (Levinson, Torreira, 2015), so that they are able to respond appropriately at the appropriate time. Evidence in favor of early planning within the current turn is given by the discrepancy between the language production system's latencies for the encoding of a single word and of a simple short clause, which are respectively of around 600 ms (Indefrey, Levelt, 2004; Schnur, Costa & Caramazza, 2006; Indefrey, 2011) and 1500 ms (Griffin, Bock, 2000), and the average duration of a silent gap in between turns across several languages, which amounts to around 200 ms (Stivers, Enfield, Brown et al., 2009; Heldner, Edlund, 2010).

Predictions about the approach of a PTL and about what will come after it are made by interlocutors on the basis of turn-taking cues in the current speaker's turn. Previous studies have identified several communicative means that are used by speakers to signal their conversational intentions for the next turn: gestural (Hadar, Steiner, Grant & Rose, 1984; Paggio, Navarretta, 2011; Edlund, Beskow, 2007, 2009; Mondada, 2007; Zellers, Gorisch, House & Peters, 2019), linguistic (Ford, Thompson, 1996; Local, Walker, 2012; Levinson, 2013; Levinson, Torreira, 2015) and phonetic (Yngve 1970; Local, Kelly & Wells, 1986; Hjalmarsson, 2011; Gravano, Hirschberg, 2011). Among the phonetic cues, the variation of F0, intensity and segmental duration have been found to significantly contribute to the turn-taking mechanism in several languages (e.g., Gravano, Hirschberg, 2009; Zellers, 2016; Brusco, Vidal, Beňuš & Gravano, 2020; Ishimoto, Teraoka & Enomoto, 2017), including German (Kohler, 1983; Selting, 1996; Niebuhr, Görs & Graupe, 2013; Peters, 2006; Dombrowski, Niebuhr, 2005). In particular, PTLs followed by a speaker change tend to be marked by either a rising or falling intonation and by a decrease in intensity, while turn holds are preceded by level F0 contours and higher intensity profiles. Backchannels also seem to be preceded by determinate sets of prosodic cues, such as regions of low pitch and some cases of uptalk in English and Japanese spontaneous conversations (Ward, Tsukahara, 2000), and final rising intonation (Skanze, Schlangen, 2009), together with higher intensity values (Gravano, Hirschberg, 2011) in task-oriented interactions in Swedish and in English. Mixed patterns of variation towards PTLs are found for segmental duration. While some studies give evidence for pre-boundary lengthening before turn yields (e.g., Local et al., 1986, for British English; Gravano, Hirschberg, 2011, for American English; Niebuhr et al., 2013 for German), others find increased segmental duration before turn holds and faster speech rate before speaker changes (e.g., Koiso, Horiuchi, Tutiya, Ichikawa & Den, 1998; Zellers, 2016; Brusco et al., 2020).

1.1 Gender variation in conversational behavior

Sociolinguistic research on verbal interactions has revealed the existence of social variation in the way turn-taking takes place: in particular, the interlocutors' genders appear to have a strong role in conversational behavior. Tannen (1994, 1998) proposed that, as a result of their different social and cultural background, men and women tend to use and interpret linguistic strategies such as interruptions, taciturnity and indirectness in contrasting ways. For instance, all-female talk has been observed to be often characterized by a collaborative floor with co-constructed utterances (Edelsky, 1993), cooperative overlaps and frequent minimal responses (Menz, Al-Roubaie, 2008; Stubbe, 2013), while men interacting with each other tend to stick to the one-speaker-at-a-time model, with either long uninterrupted turns or short rapid ones, where overlaps are rare and perceived as deviant (Coates, 2004). Thus, while in some contexts conversational strategies like interruptions might be used to show support and achieve cooperation, in others they can also be employed to reinforce asymmetry and establish dominance over the other speaker. Asymmetrical interactional patterns have been often observed in mixed-gender conversations, with one speaker saying too much or too little, interrupting the other or virtually withdrawing from the conversation (e.g., Waara, Shaw, 2006; Coates, 2004; Zimmerman, West, 1975). The configuration of same-gender and mixed-gender interactions, however, does not always follow these exact patterns: results from different studies give oftentimes seemingly inconsistent results, since the influence of speakers' genders on their turn-taking behavior seems to be mediated by other social and contextual factors, such as the conversational goal and interactional setting, institutionalized roles and gender identity salience (Plug, Stommel, Lucassen, Dulmen & Das, 2021).

The majority of the scientific literature on gender variation in turn-taking has strongly focused on pragmatics, addressing, for example, talkativeness, the use of tentative language, and interruptions. It is not clear, however, if the similarities and the differences between genders also extend to the use they make of turn-taking cues, such as phonetic ones, to signal to the interlocutor their conversational intentions of offering the next turn or continuing to speak. We are aware of the overall differences in F0, intensity and segmental duration in men's and women's speech (e.g., women tend to have a higher mean F0 than men [Weirich, Simpson, 2019], men have higher conversational intensity levels than women [Gelfer, Young, 1997] and slightly lower durational values as well [Pépiot, 2014]). However, the possible influence of gender on the phonetic variation of these features as turn-taking cues has, to the best of the author's knowledge, never been empirically investigated in any language. Thus, the analysis presented in the current paper aims at proposing a first sociophonetic exploration of turn-taking by testing if, and how, social variables such as the gender of both interlocutors in a dialogue might have an effect on the acoustic cues that speakers use to signal their intentions for the next conversational turn.

2. *Methods*

In order to offer more insight into the variation of phonetic turn-taking cues in spontaneous interactions in German and to investigate the possible influence of gender on such variation, a dataset composed of two-party conversations between German native speakers, balanced for the gender of the interlocutors, was annotated, and values for F0, intensity and segmental duration were extracted at different test locations approaching a PTL.

2.1 Dataset

The two-party conversations analyzed for this study are part of the German sub-corpus of the DUEL Multi-lingual Multimodal Dialogue Corpus (Hough, Tian, De

Ruiter, Betz, Kousidis, Schlangen & Ginzburg, 2016). The DUEL corpus consists of naturalistic, face-to-face conversations based off loosely-directed tasks assigned to the participants. The tasks were specifically designed to provide subjects with a theme to discuss without, at the same time, being constrained in how to develop their interactions (*ibid*). In particular, in the portions of dialogues analyzed for the present research, the speaker pairs dealt with either the "Dream Apartment" or the "Film Script" tasks, in which participants had to imagine and plan out the organization, furniture and decoration of a hypothetical shared apartment or imagine and describe a movie scene centered on an embarrassing situation, respectively. The two interactants in each dyad were recorded on separate channels, which allows the phonetic analysis of speech even when the participants talk in overlap with each other. For the present study, 6 dyads have been annotated and analyzed, for a total of 12 different speakers, all German native speakers and (at the time of the recording) students at the University of Bielefeld, aged 21-28. The dyads were selected with the aim of having both speakers of the same gender in conversation with each other, as well as subjects of different genders interacting with each other. The dyads thus included in the dataset for the study are 2 male-to-male (MM) conversations, 2 female-to-female (FF) conversations and 2 mixed-gender (MF) conversations (see Tab. 1). The first 5 minutes of the tasks have been taken into consideration for this analysis.

2.2 Annotation

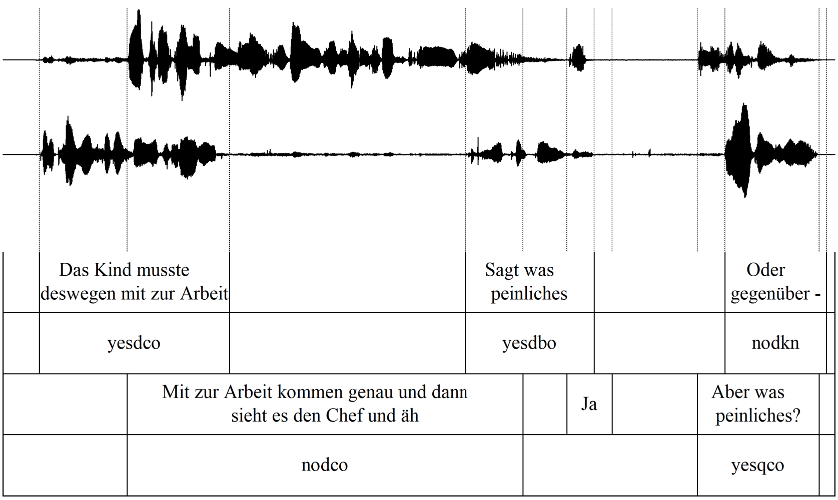
The annotation was carried out in *Praat* (Boersma, Weenink, 2022; Boersma, 2001) and it started with the manual annotation of PTLs (Schiffrin, 1987). As described by Sacks et al. (1974), such locations in conversations, defined by them as Transition Relevance Spaces, arise at the end of Turn Constructional Units, when speaker change becomes possible, though not mandatory. Following Zellers' (2016) classification of Potential Turn Boundaries, these locations were identified not only after syntactically complete utterances, but also after syntactically incomplete utterances which functioned as semantically or pragmatically complete in the local context of the interaction. Such locations were identified in the data using the right boundaries of the "X-utts" tier provided in the DUEL corpus, which contained the segmentation and the orthographic transcription of speakers' utterances within each turn (annotated in the "X-turns" tier). The guidelines for the segmentation of the utterances given to the annotators of the DUEL corpus follow the notion of "slash unit" described by Meteer et al. (1995), in which an utterance should be comprised by maximally a complete sentence or a smaller unit, which may not be syntactically complete but is judged as complete in context by the annotators (Hough et al., 2016). Turns in the "X-turns" tier, instead, included all continuous stretches of speech by one speaker until the other one takes up the floor, or up until a silent gap after which either the current speaker continues talking, or the interlocutor takes up the next turn. By assuming the right boundaries of the "X-utts" tier as PTLs, it was possible to include in the analysis instances considered by the annotators as complete in context, but that were not necessarily followed by a speaker change or by a silent gap.

Once PTLs were identified, they were annotated using a set of labels (Feindt, Rossi & Zellers, 2021; Rossi, Feindt & Zellers, 2022a) describing the utterance by the current speaker and what came after it (see Fig. 1):

- Completeness label: indicated the syntactic/semantic completion of the utterance in context ("yes" for complete utterances and "no" for incomplete utterances);
- Sentence Type label: described the form of the utterance ("d" for declaratives, "q" for questions and "t" for tag questions);
- Sequential Structure label: indicated the conversational action by the current speaker or by the other participant which followed the PTL, i.e., the other participant took up the next full turn (speaker changes, labelled with "c", also referred to as turn yields), the current speaker held the floor (keeps, labelled with "k", also referred to as turn holds), the other participant produced a minimal, non-interrupting response (backchannels, labelled with "b");
- Transition Type label: described the way in which the conversational action following the PTL took place, i.e. with a silent gap ("g"), with a speech overlap ("o"), in a smooth way, without any perceivable silences or speech overlaps (no-gap-no-overlap, "n"); in particular, only those silences longer than 120ms were annotated as gaps, and only those stretches of overlapped speech longer than 120ms were labelled as overlaps, as it has been shown that gaps and overlaps shorter than 120ms are not perceived as such by listeners (Heldner, 2011). Transitions occurring with possible silences or overlaps shorter than 120ms were considered as no-gap-no-overlaps.

For each utterance, words, syllables and segments were also annotated.

Figure 1 - *Example of the annotation of PTLs in Praat, from one of the dialogues in the dataset. The first part of the label(s) refers to the completeness in context ("yes" or "no"), the following letter refers to the sentence type ("d" for declarative, "q" for questions, "t" for tag questions), the third part of the label refers to the sequential structure that followed ("c" for speaker change, "k" for keep and "b" for backchannel), and the final letter describes the transition type ("g" for gaps, "o" for overlaps, "n" for no-gap-no-overlaps)*



2.3 Data extraction

As a part of a wider research project aimed at investigating the location and the extension of the transition space in conversation and testing the hypotheses related to the relevance for turn-taking of time windows against that of phonological categories (issues that will not be discussed in the current paper), the present analysis observes the variation of the phonetic parameters using three different time-points approaching a PTL as reference (Feindt et al., 2021). The phonetic parameters investigated are F0, intensity and segmental duration, and the three test time-points approaching a PTL are located at 500 ms before the end of the utterance, at 200 ms before the end of the utterance, and at the end of the utterance. The phonetic values were extracted automatically using a *Praat* script from utterances with a duration of 1 second and up, so that it was possible to extract data from all three test locations without them being too close to the start of the utterance (Feindt et al., 2021). F0 readings, extracted using *Praat* with the settings for semitones (st) above 1Hz, were then normalized with the individual speaker's baseline in order to exclude the influence of physiological factors. Using a sample of F0 datapoints in semitones extracted from 2 minutes of clear speech for each subject, the individual baseline was calculated as 0.75 times the first quartile of the data, following Zellers and Schweitzer (2017) and Zellers (2021). Intensity was extracted in decibels (dB) and normalized with the speakers' mean (Ludusan, Dupoux, 2015). Values for segmental duration were also extracted at three different test intervals approaching the PTL: the average segmental duration, in milliseconds (ms), was extracted over the last 500 ms and over the last 200 ms (i.e., from the offset of speech to 500 ms before that and from the offset to 200 ms before that) and, finally, the duration of the last segment at the end of the utterance was extracted. This way, it was possible to observe how average segmental duration varied towards PTLs. No normalization was carried out at this stage for segmental duration; however, the randomness of the segments included in the analysis allows us to make general preliminary observations about the variation of segmental duration approaching PTLs, excluding the potential influence of the different segment types.

3. *Analysis and results*

A total of 489 PTLs with the related phonetic values at the three test locations were extracted. The qualitative and quantitative analysis of the data was carried out in *R* (RStudio Team, 2020). Linear mixed effects models with the subject as a random factor were used for the quantitative analysis of the data using the *R* package lmerTest (Kuznetsova, Brockhoff & Christensen, 2017).

From a first qualitative exploration of the data, differences between the behavior of the male and female speakers in conversation appear. For instance, the numbers of speaker changes and keeps in FF conversation were equal: female speakers tend to yield the turn as much as they hold it (78 changes and 70 keeps); on the contrary, in MM conversation there is a bigger disproportion between turn holds and speaker changes, with male speakers keeping the floor for longer, and more frequently than they are ceding it (44 changes and 92 keeps). Moreover, it was observed that smooth transitions (i.e. no gap no overlap) occurred more frequently in same-gender conversation, thus 26% of the transitions are smooth in MM conversation and 34% in FF conversations, while only 16% of transitions were smooth in mixed-gender conversations (see Tab. 2). Silent gaps occur more frequently in MM conversations than in FF and MF ones, while speech overlaps are observed more frequently in FF dyads (see Tab. 1).

Table 1 - *Transition type (gaps, overlaps, no-gap-no-overlaps) distribution in same-gender ("MM" for male-to-male, "FF" for female-to-female) and mixed-gender ("FM" for female-to-male) dyads*

	gaps	*overlaps*	*no-gap-no-overlaps*
MM	61% (97)	13% (20)	26% (41)
FF	39% (64)	28% (46)	34% (56)
MF	56% (93)	27% (45)	16% (27)

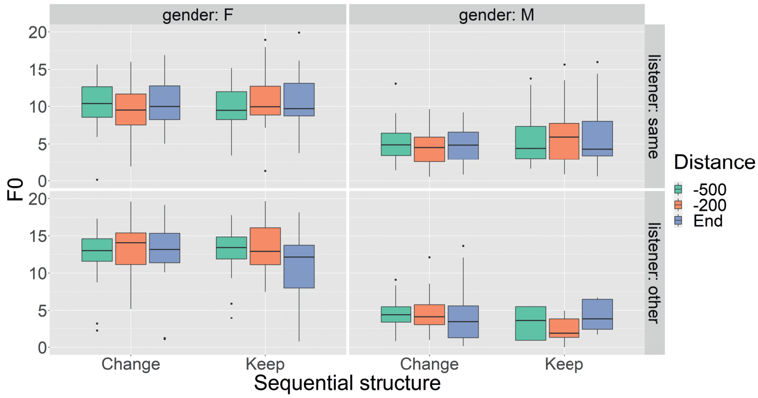
The analysis reported below focusses on the pre-final phonetic marking of syntactically/semantically complete utterances in declarative form preceding speaker change and keep cases, and how the gender of the interlocutors influences that.[1] The PTLs thus considered for the following analysis are 239, including 131 followed by speaker change cases and 108 by turn holds, constituted by syntactically/semantically complete utterances in declarative form.

3.1 F0 values

The variation of the normalized F0 values towards PTLs revealed significant differences both between the speaker changes and keeps, and between male and female interlocutors and the way in which they approach different sequential structures. Male speakers' values at the three test locations analyzed (at the end of the utterance, at 200 ms from the boundary and at 500 ms from the end of the utterance) are closer to their baseline, at around 5 st, while the average values for female speakers are higher, mostly above 10 st.

[1] Due to space constraints, the analysis involving transition types (i.e., gaps, overlap and smooth transitions) will not be reported, and, for the purposes of this study, sentence types such as questions and tag questions are momentarily set aside, in an attempt to limit the sources of prosodic variation. Finally, the exclusion of TRPs preceding backchannels is due to limitations in the current annotation scheme and analysis structure. In fact, several factors have been found to influence backchannel's placement (e.g., the lexical or non-lexical content of the backchannel, or its modality, cfr. Truong, Poppe, De Kok & Heylen, 2011; Ferré, Renaudier, 2017), and would thus have to be considered when targeting their distribution and the cues preceding them. Since such factors are not present in this annotation scheme, nor they would fit into the rest of the analysis, the description that would have resulted would have been biased and ambiguous. Further developments of this research, though, include a more precise annotation of backchannels, as well as their analysis.

Figure 2 - *F0 datapoints distribution (in st, above the speaker's baseline) for speaker change and keep cases at the three test locations ("Distance": -500 ms, -200 ms, End), for female speakers (gender: F) in same (listener: same) and mixed-gender dialogues (listener: other), and for male speakers (gender: M) in same (listener: same) and mixed-gender dialogues (listener: other)*



Considering both the gender of the speaker and the gender of the interlocutor, in speaker change cases the values of male speakers are closer to their baseline, both in MM and MF conversations. In all conditions, the contour created by the three datapoints tends to be flat, though with a certain degree of variability, with the exception of male speakers in MF conversations, where the values of F0 at the End location fall closer to the speakers' baselines, at 2.5 st, while for female speakers they tend to remain higher, closer to the previous ones, at 12.5 st (see Fig. 2). Moreover, in both speaker change cases and keep cases (see Fig. 2), it appears that female speakers are always higher than males, but they are even higher when in conversation with a male speaker; similarly, male speakers in MM conversation show low values, but males in conversation with females are even lower.

Table 2 - *Summary of the linear mixed model for F0 in speaker change cases. A significant three-way interaction is found between the value of the final F0 test location (End), the gender of the speaker and the gender of the interlocutor. The speaker is included as a random factor in the model. Formula: lmer(F0 ~ distance * gender * listener + (1 | speaker))*

	Estimate	*Std. Error*	*DF*	*t-value*	*Pr(>|t|)*
(Intercept)	11.0322	1.6712	9.2173	6.601	8.88e-05 ***
-200: Gender M	2.4714	0.8134	1465.0749	3.038	0.002420 **
End: Gender M	4.4082	0.8134	1465.0749	5.420	6.98e-08 ***
-200 listener other	1.9198	0.8134	1465.0749	2.360	0.018391 *
End: listener other	4.5724	0.8134	1465.0749	5.621	2.26e-08 ***

	Estimate	*Std. Error*	*DF*	*t-value*	*Pr(>|t|)*
Gender M: listener other	-2.2018	4.0219	8.5934	-0.54	0.597998 ns
*-200*gender M: listener other*	-1.5617	1.2932	1465.0749	-1.20	0.227398 ns
*End*gender M: listener other*	-7.1274	1.2932	1465.0749	-5.51	4.20e-08 ***

A linear mixed model shows that the value of F0 at the PTL differ significantly for male speakers in mixed-gender conversations (see Tab. 2). A significant three-way interaction between the distance from the PTL, the gender of the speaker and the gender of the interlocutor is found for keep cases, too: the height of F0 at the –200 datapoint differ significantly from the intercept for male speakers in mixed-gender conversations (see Tab. 3). In these interactions, F0 is rising at the end of the utterances for male speakers, while it is falling for female speakers (see Fig. 2).

Table 3 - *Summary of the linear mixed model for F0 in turn hold cases. A significant three-way interaction is found between the value of the penultimate F0 test location (-200), the gender of the speaker and the gender of the interlocutor. The speaker is included as a random factor in the model. Formula: lmer(F0 ~ distance *gender * listener + (1 | speaker))*

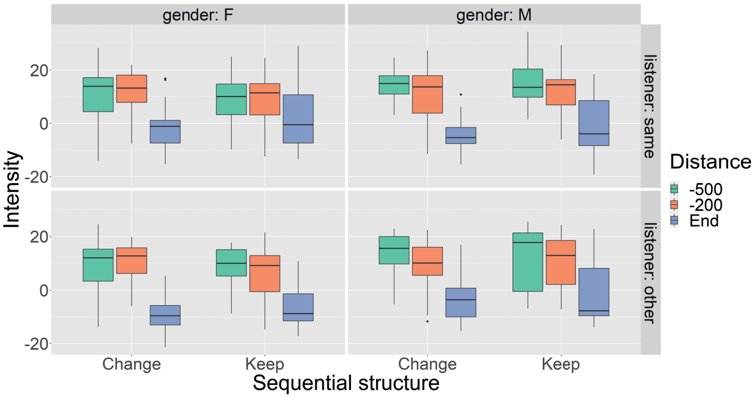
	Estimate	*Std. Error*	*DF*	*t-value*	*Pr(>|t|)*
(Intercept)	9.1696	1.2859	7.3667	7.13	0.000148 ***
-200: Gender M	-0.8469	0.4724	2896.9797	-1.79	0.073108 .
End: Gender M	-1.4471	0.4724	2896.9797	-3.06	0.002209 **
-200 listener other	-0.3076	0.5112	2896.9797	-0.60	0.547383 ns
End: listener other	-2.3349	0.5112	2896.9797	-4.56	5.15e-06 ***
Gender M: listener other	-2.5707	3.6319	7.3266	-0.70	0.500957 ns
*-200*gender M: listener other*	-3.6392	0.9743	2896.9797	-3.73	0.000191 ***
*End*gender M: listener other*	1.3532	0.9743	2896.9797	1.389	0.164965 ns

3.2 Intensity

For each of the sequential structures analyzed, intensity decreases approaching the end of the utterance, with the values of the last test location dropping below the speakers' means. The decrease is higher for PTLs preceding speaker change cases, where it appears that speakers lower their voices much more than before keeps.

In speaker change cases, male speakers appear to be louder than females, except at the end of the turn in MM conversations, where their values drop at –5 db below their means, while the final values for females in FF conversations are closer to their means. On the contrary, female speakers talking to a male conversational partner tend to lower their intensity much more approaching the end of the turn, while males in conversation with a female partner end their turns with a higher intensity, closer to their means (see Fig. 3).

Figure 3 - *Intensity datapoints distribution (in dB, normalized to the speaker's mean) for speaker change and keep cases at the three test locations ("Distance": -500 ms, -200 ms, End), for female speakers (gender: F) in same (listener: same) and mixed-gender dialogues (listener: other), and for male speakers (gender: M) in same (listener: same) and mixed-gender dialogues (listener: other)*



For keep cases, when the speaker held the floor, intensity decreases less towards the PTL. In fact, in these cases the intensity values at the last test location fall below the speakers' means, but they remain closer to it than in speaker change cases. Considering both the speakers' and the interlocutors' gender, however, it is possible to detect an exception in this trend: intensity values at the end of the utterance for male speakers in MF dyads decrease well below their mean, more than it does for female speakers in MF dyads.

For both speaker change cases and keep cases, two separate linear mixed effects models show a three-way interaction between the gender of the speaker, the gender of the interlocutor and the intensity at the final test location (see Tab. 4 and Tab. 5).

Table 4 - *Summary of the linear mixed model for intensity in speaker change cases. A significant three-way interaction is found between the value of the final test location, the gender of the speaker and the gender of the interlocutor. The speaker is included as a random factor in the model. Formula: lmer(intensity ~ distance * gender * listener + (1 | speaker))*

	Estimate	*Std. Error*	*DF*	*t-value*	*Pr(>|t|)*
(Intercept)	10.1865	1.062	9.7709	9.58	2.78e-06 ***
-200: Gender M	-3.2728	0.858	3517.058	-3.81	0.000139 ***

	Estimate	*Std. Error*	*DF*	*t-value*	*Pr(>|t|)*
End: Gender M	-6.6951	0.858	3517.058	-7.80	7.88e-15 ***
-200 listener other	0.1572	0.850	3517.058	0.18	0.853326 ns
End: listener other	-6.7422	0.850	3517.058	-7.92	2.94e-15 ***
Gender M: listener other	1.0235	2.576	9.3741	0.37	0.700063 ns
*-200*gender M: listener other*	-1.8130	1.274	3517.058	-1.42	0.154837 ns
*End*gender M: listener other*	8.8455	1.274	3517.058	6.94	4.57e-12 ***

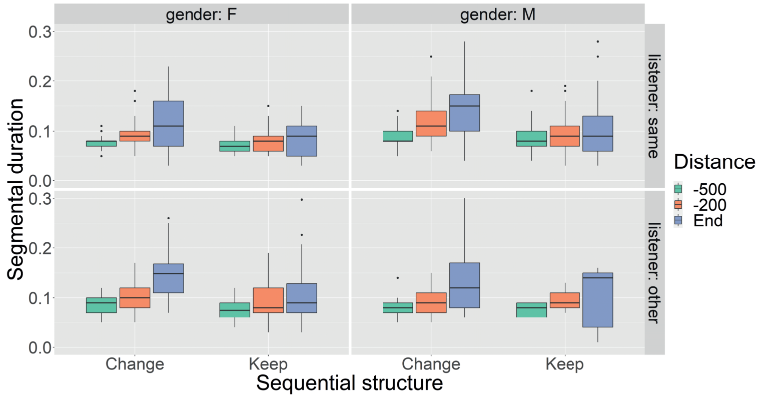
Table 5 - *Summary of the linear mixed model for intensity in turn hold cases. A significant three-way interaction is found between the value of the final test location, the gender of the speaker and the gender of the interlocutor. The speaker is included as a random factor in the model. Formula: lmer(intensity ~ distance * gender * listener + (1 | speaker))*

	Estimate	*Std. Error*	*DF*	*t-value*	*Pr(>|t|)*
(Intercept)	10.3673	2.121	7.5234	4.88	0.00145 **
-200: Gender M	-2.6697	0.943	2896.948	-2.83	0.00468 **
End: Gender M	-7.3551	0.943	2896.948	-7.79	8.69e-15 ***
-200 listener other	-2.7057	1.020	2896.948	-2.65	0.00807 **
End: listener other	-6.9076	1.020	2896.948	-6.76	1.58e-11 ***
Gender M: listener other	-0.4589	5.988	7.4643	-0.07	0.94093 ns
*-200*gender M: listener other*	3.8497	1.945	2896.948	1.97	0.04790 *
*End*gender M: listener other*	8.0387	1.945	2896.948	4.13	3.69e-05 ***

3.3 Segmental duration

Segmental duration is significantly influenced by the distance from the PTL and the different sequential structures that follow the PTL. Duration increases over the three test locations up until the end of the utterance for PTLs preceding speaker change cases. Average final segmental duration is higher also for turn holds, but it remains closer to the average of the previous test location (-200). This indicates that speakers do slow down their speech when approaching a PTL before a keep, but not as much as they do before a speaker change (see Fig. 4).

Figure 4 - *Average segmental duration in ms for speaker change and keep cases over the three test locations ("Distance": -500 ms, -200 ms, End), for female speakers (gender: F) in same (listener: same) and mixed-gender dialogues (listener: other), and for male speakers (gender: M) in same (listener: same) and mixed-gender dialogues (listener: other)*



For speaker change cases, final segmental duration increases to a higher degree for male speakers in MM conversations than for female speakers in FF conversations. In MF dyads, on the contrary, duration at the final test location increases for female speakers, while its average remains closer to the previous locations for male speakers (see Fig. 4).

A linear mixed effects model shows a three-way interaction between the gender of the speaker, the gender of the interlocutor and the duration values at the final test location (Tab. 7).

Table 7 - *Summary of the linear mixed model for segmental duration in speaker change cases. A significant three-way interaction is found between the value of the final test location, the gender of the speaker and the gender of the interlocutor. The speaker is included as a random factor in the model. Formula: lmer(duration~ distance *gender *listener + (1 | speaker))*

	Estimate	*Std. Error*	*DF*	*t-value*	*Pr(>|t|)*
(Intercept)	7.724e-02	6.451e-03	9.551e+00	11.974	4.57e-07 ***
-200: Gender M	1.480e-02	5.137e-03	3.517e+03	2.881	0.00399 **
End: Gender M	4.744e-02	5.137e-03	3.517e+03	9.235	< 2e-16 ***
-200 listener other	4.184e-03	5.091e-03	3.517e+03	0.822	0.41120 ns
End: listener other	3.020e-02	5.091e-03	3.517e+03	5.932	3.29e-09 ***
Gender M: listener other	-1.568e-02	1.564e-02	9.173e+00	-1.003	0.34177 ns
*200*gender M: listener other*	-1.979e-02	7.629e-03	3.517e+03	-2.594	0.00953 **
*End*gender M: listener other*	-5.805e-02	7.629e-03	3.517e+03	-7.609	3.53e-14 ***

In keep cases, as already mentioned, final duration does not increase as much as in speaker change instances. However, this does not seem to be true for male speakers in MF dialogues. In fact, they seem to reach the potential turn end with an increased segmental duration also before turn holds (see Fig. 4). Even if it is possible to detect qualitative differences, in this case, they did not reach statistical significance.

3.4 Summary of the results

The phonetic cues preceding speaker changes and those preceding keep cases show different patterns of variation. Intensity tends to decrease towards the end of the utterance, but to a lesser extent when the current speaker wants to continue talking. Segmental duration increases towards the end, but to a lesser extent, again, when the current speaker has the intention of continuing. F0 values show a high degree of variation, so it is harder to get a clear pattern of variation related to the turn-taking structure. As for the sociophonetic variation of turn-taking cues, there are significant differences in the way male and female speakers in same-gender and mixed-gender conversations mark the test locations for both keep and speaker change cases. In particular, the results obtained using linear mixed models, indicate that, in speaker changes, the interaction between the gender of the speaker, the gender of the interlocutor and the final datapoint, i.e., the end of the utterance, was significant for F0 values (est.: -7.1274; t value: -5.511; $p < .0001$) for intensity (est.: 8.8455; t value: 6.942; $p < .0001$) and segmental duration (est.: -0.05805; t value: -7.609; $p < .0001$). For segmental duration, the interaction of the speaker's and the interlocutor's genders with the "-200" location was significant as well (est.: -0.01979; t value: -2.594; $p < .01$). For keep cases, the variation of F0 values is influenced by the interaction of the gender of the interlocutors and the "-200" location (est.: 1.3532; t value: 1.389; $p < .0001$), and, for intensity, by the gender of the interlocutors and the "-200" (est.: 3.8497; t value: 1.979; $p < .05$) and "End" (est.: 8.0387; t value: 4.133; $p < .0001$) locations.

4. *Discussion*

The results related to the phonetic variation of turn-taking cues in this dataset show patterns that had already been observed in previous research, i.e., speakers tend to maintain their intensity higher (Gravano, Hirschberg, 2009; 2011) and their speech rate faster (Local et al., 1986; Niebuhr et al., 2013) in order to hold the floor. This configuration of the phonetic turn-taking cues approaching PTLs is in line with the idea that speakers might put more effort into marking turn holds than into signaling turn-yielding intentions because speaker change might be "a kind of default option" (Zellers, 2016: 13) in spontaneous conversation. This might be especially relevant when the speaker's utterance is syntactically/semantically complete, as was the case for the utterances considered in this study. In fact, in the presence of syntactic/semantic completion, it might be more likely that the interlocutor will assume that the current speaker has finished with their turn, so a more prominent phonetic marking is needed. Patterns for F0 variation related to turn-taking are less clear. In fact, no recurrent or systematic patterns emerged from the qualitative analysis of the data. This observation is in line with the results obtained by Feindt et al. (2021) who suggest that, since F0 movements in German are not restricted by any phonological function, as it is the case for other languages, e.g., Swedish, speakers in conversation tend to use F0 more freely towards turn ends and make use of their entire F0 range. Moreover, the fact that only syntactically/semantically complete utterances were taken into consideration for the present study gives a further explanation to the high degree of variation obtained from this data sample: in fact, it has been observed that when the upcoming completion of an utterances is signaled by syntactic, semantic or pragmatic means, F0 patterns are less restricted for turn-taking function, and speakers tend to show a higher degree of variability towards the end of their turn (Selting, 1996; Feindt et al., 2021; Rossi et al., 2022a; Rossi, Feindt & Zellers, 2022b).

As for gender variation, a few differences were observed in the phonetic patterns of the parameters analyzed, which were seemingly influenced not only by the gender of the speaker, but also by the gender of the interlocutor. First of all, the datapoints extracted for F0 (in semitones and normalized with the speakers' baseline to exclude the influence of physiological factors) were always higher for female speakers and closer to the baseline for male speakers, a result consistent with other studies on German and other languages (Weirich, Simpson, 2019; Andreeva et al., 2014; Pépiot, 2014; Pépiot, Arnold, 2021). The difference in F0 height was particularly striking in MF conversations, where F0 was closer to the baseline for male speakers than in MM conversations, while for female subjects it was higher in MF than in FF conversations. Moreover, more rises and falls in F0 were observed for male speakers than for female speakers and, in a few occasions, F0 even showed a more consistent patterning for them, as for example in the keep cases in MF dyads (see Fig. 2).

Variation related to turn-taking for intensity and segmental duration, instead, seemed to be employed more consistently by female speakers in both FF and MF conversations, and by male speakers in MF conversations. Interestingly, in fact, male speakers in MF dyads seemed to use these values to a smaller extent to signal their conversational intentions, e.g., they tend to not lower their final intensity as much as it happens in other conditions (e.g., in MM dyads), as well as do not vary their speech rate in the direction it would be expected, or as much as their female conversational partners. On the contrary, male speakers make use of these cues in line with the expectations in MM dyads and, in these cases, their variation related to turn-taking is much more striking than it is for female speakers in FF dyads. Thus, in general, it appears from these observations that female speakers make more use of duration and intensity as turn-taking cues, while it is possible that male speakers may vary more with F0 towards turn ends to signal their intentions. This observation would be in line with the results obtained by Whiteside (1995; 1996) for British English. She finds that, in a reading task, women mark syntactic boundaries by pausing and using phrase-final lengthening, while male subjects rely more on F0 shifts and pause less frequently, which, in a conversational setting, would make it harder for the interactant to take up the next turn (Whiteside, 1995). Moreover, their more consistent use of duration and intensity in MF dialogues may suggest that female speakers in this data sample are doing more work in terms of signaling their conversational intentions than male speakers do. However, similar patterns of variation, even more accentuated, are observed for male speakers in MM conversations. So, in general, it appears that all subjects put more effort in terms of signaling their intentions to either cede or hold the floor when the other interlocutor was a male.

As previously mentioned (see §1.1), the interactional setting should be taken into consideration when discussing gender influence on conversational behavior. Previous studies on talkativeness and assertiveness and their relation to gender identity reported that, in mixed-gender dyads, gender identity was salient for the participants, since the effects of controlled factors were significantly bigger than in same-gender ones (Leaper, Ayres, 2007; Hannah, Murachver, 2007). Applying this to our results may suggest that female speakers in mixed conversations, where gender might have been more salient, tended to take up a more cooperative and facilitative role (Holmes, 1995), focused on keeping the flow of the conversations as smooth as possible by making their turn-taking intentions clear. However, the structure of the exchanges of same-gender dyads suggest some influence of the gender of the two interlocutors, too. For example, a greater number of overlaps were found in female-to-female speech, while a lower number of turn yielding cases accompanied by a higher number of holds was observed for male-to-male dyads, which is similar to the descriptions of same-sex conversations provided by the literature (see §1.1). In fact, it has also been indicated how also experimental activities seem to interact differently with gender, with task-oriented dialogues making interlocutors' genders more relevant than non-structured conversations (Leaper, Robnett, 2011; Leaper, Ayres, 2007). Finally, in this context, the result of male speakers doing more signaling work than female speakers in MM conversation is interesting and unexpected. A possible explanation could be that the goal of the conversation, i.e., discuss with each other to complete a cooperative task, was salient and thus the interactants might have focused on that, and tried to cooperate as much as possible with each other by facilitating turn-taking and putting more effort into signaling their conversational intentions.

5. *Conclusions and outlook*

From this first exploration of the influence of gender on the variation of turn-taking cues, it appears that male and female speakers in this data sample share both commonalities and differences. All the speakers display the characteristics of potential turn ends already investigated for German, with certain patterns of variation of F0, intensity and segmental duration before turn holds and speaker changes. A few differences were observed, however, in the degrees of variation of such cues, especially when the gender of the interlocutor was taken into consideration as well. Even if quantitatively significant, it must be kept in mind that these results concern a relatively small set of data and speaker sample. Moreover, since it seems that no previous study has tested how social and individual variables influence the variation of phonetic turn-taking cues, it is not possible to generalize or make broader claims on the issue. However, these results provide some preliminary evidence for a possible influence of gender on the way male and female speakers mark PTLs in spontaneous interactions in German. Thus, it can be a matter worth of being investigated further, since it might give us some more insight into what contributes to produce the different structures that same-gender and mixed interactions sometimes seem to assume.

Further developments of this research will focus on a larger set of data and will include more phonetic parameters in the analysis, such as voice quality and segmental reduction measures, in order to observe their variation towards PTLs in German and to test the influence of the interlocutors' genders on their use. Then, a better tailored annotation and analysis will focus on backchannels, both vocal and gestural (in the form of head movements) and on the phonetic marking of the speech preceding them. In line with the current one, the investigation of backchannels in German will deal with their form, their distribution in the conversation, and the extent to which the gender of both interlocutors might have an effect on backchannel production and the phonetic cues preceding it.

Acknowledgements

This research was supported by the project "Sound Patterns and Linguistic Structures at the Transition Space in Conversation" funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Council; project no. 444631148). Many thanks to Margaret Zellers and Kathrin Feindt for their valuable suggestions and discussions on the methods and the analysis of the data, to Margareta Ehlers-Karmanova for her work on the segmental annotations, and to the reviewers for their insightful comments and suggestions.

References

Andreeva, B., Demenko, G., Wolska, M., Möbius, B., Zimmerer, F., Jügler, J., Jastrzebska, M. & Trouvain, J. (2017). Comparison of pitch range and pitch variation in Slavic and Germanic languages. *Proceedings of Speech Prosody*, Dublin, Ireland, 20-23 May 2014. https://doi.org/10.21437/speechprosody.2014-143

Boersma, P. (2001). Praat, a system for doing phonetics by computer. In *Glot International* 5:9/10, 341-345.

Boersma, P., Weenink, D. (2022). Praat: doing phonetics by computer [Computer program]. Version 6.2.14. https://www.praat.org/

Brusco, P., Vidal, J., Beňuš, S. & Gravano, A. (2020). A cross-linguistic analysis of the temporal dynamics of turn-taking cues using machine learning as a descriptive tool. In *Speech Communication*, 125, 24-40. https://doi.org/10.1016/j.specom.2020.09.004

Coates, J. (2004). *Women, men and language: a sociolinguistic account of gender differences in language.* Harlow: Pearson Longman.

Dombrowski, E., Niebuhr, O. (2005). Acoustic patterns and communicative functions of phrase-final rises in German: activating and restricting contours. In *Phonetica*, 62, 176-195. https://doi.org/10.1159/000090097

Edelsky, C. (1993). Who's got the floor? In: Tannen, D. (Ed.) *Gender and Conversational Interaction*, 189–227. Oxford: Oxford University Press.

Edlund, J., Beskow, J. (2007). Pushy versus meek-using avatars to influence turn-taking behavior. *Proceedings of Interspeech,* Antwerp, Belgium, 27-31 August 2007. https://doi.org/10.21437/Interspeech.2007-289

Edlund, J., Beskow, J. (2009). Mushypeek: A framework for online investigation of audiovisual dialogue phenomena. In *Language and Speech*, 52(2-3), 351-367.

Feindt, K., Rossi, M. & Zellers, M. (2021) The time course of pitch variation towards possible places of speaker transition in German and Swedish. *Proceedings of Tone and Intonation*, Sonderburg, Denmark, 6-9 December 2021. https://doi.org/10.21437/tai.2021-23

Ferré, G., Renaudier, S. (2017). Unimodal and bimodal backchannels in conversational English. *Proceedings of SEMDial*, Saarbrücken, Germany, 15-17 August 2017. https://doi.org/10.21437/SEMDIAL.2017-3

Ford, C., Thompson, S. (1996). Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns. In E. Ochs, E. Schegloff, S.A., Thompson (Eds.), *Interaction and grammar,* 134–184. Cambridge: Cambridge University Press.

Gelfer, M.P., Young, S.R. (1997). Comparisons of intensity measures and their stability in male and female sneakers. *Journal of Voice*, 11(2), 178-186.

Gravano, A., Hirschberg, J. (2009). Turn-yielding cues in task-oriented dialogue. *Proceedings of SIGDial,* London, UK, 11-12 September 2009. https://doi.org/10.3115/1708376.1708412

Gravano, A., Hirschberg, J. (2011). Turn-taking cues in task-oriented dialogue. *Computer Speech and Language*, 25, 601–634. https://doi.org/10.1016/j.csl.2010.10.003

Griffin, Z.M., Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, 4, 274–279. https://doi.org/10.1111/1467-9280.00255

Hadar, U., Steiner, T.J., Grant, E.C. & Rose, F.C. (1984). The timing of shifts of head postures during conversations. *Human Movement Science*, 3, 237–245.

Hannah, A., Murachver, T. (2007). Gender preferential responses to speech. *Journal of Language and Social Psychology,* 26(3), 274-290. https://doi.org/10.1177/0261927X06303457

Heldner, M. (2011). Detection thresholds for gaps, overlaps, and no-gap-no-overlaps. *Journal of the Acoustical Society of America*, 130(1), 508–513. https://doi.org/10.1121/1.3598457

Heldner, M., Edlund, J. (2010). Pauses, gaps and overlaps in conversations. *Journal of Phonetics,* 38, 555–568. https://doi.org/10.1016/j.wocn.2010.08.002

Hough, J., Tian, Y., De Ruiter, L., Betz, S., Kousidis, S., Schlangen, D. & Ginzburg, J. (2016). DUEL: A Multilingual Multimodal Dialogue Corpus for Disfluency, Exclamations and Laughter. *Proceedings of Language Resources and Evaluation Conference,* Portoroz, Slovenia, 23-28 May 2016.

Holmes, J. (1995). *Women, Men and Language.* London: Longman.

Hjalmarsson, A. (2011). The additive effect of turn-taking cues in human and synthetic voice. *Speech Communication*, 53, 23–35. https://doi.org/10.1016/j.specom.2010.08.003

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Frontiers in Psychology,* 2, 255. https://doi.org/10.3389/fpsyg.2011.00255

Indefrey, P., Levelt, W.J. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92(1), 101-144. https://doi.org/10.1016/j.cognition.2002.06.001

Ishimoto, Y., Teraoka, T. & Enomoto, M. (2017). End-of-Utterance Prediction by Prosodic Features and Phrase-Dependency Structure in Spontaneous Japanese Speech. *Proceedings of Interspeech*, Stockholm, Sweden, 20-24 August 2017. https://doi.org/10.21437/Interspeech.2017-837

Kohler, k. J. (1983). Prosodic boundary signals in German. In *Phonetica*, 40, 89–134.

Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A. & Den, Y. (1998). An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs. In *Language and Speech*, 41, 295–321.

Kuznetsova, A., Brockhoff, P.B. & Christensen, R.H.B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1–26. https://doi.org/10.18637/JSS.V082.I13

Leaper, C., Ayres, M.M. (2007). A meta-analytic review of gender variations in adults' language use: Talkativeness, affiliative speech, and assertive speech. In *Personality and Social Psychology Review*, 11(4), 328-363. https://doi.org/10.1177/1088868307302221

Leaper, C., Robnett, R.D. (2011). Women are more likely than men to use tentative language, aren't they? A meta-analysis testing for gender differences and moderators. In *Psychology of Women Quarterly*, 35(1), 129-142. https://doi.org/10.1177/0361684310392728

Levinson, S.C. (2013). Action formation and ascription. In Sidnell, J. & Stivers, T. (Eds.) *The Handbook of Conversation Analysis*, 103-130. Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781118325001.CH6

Levinson, S.C., Torreira, F. (2015) Timing in turn-taking and its implications for processing models of language. In *Frontiers in Psychology*, 6. https://doi.org/10.3389/fpsyg.2015.00731

Local, J.K., Kelly, J. & Wells, W.H. (1986). Towards a phonology of conversation: turn-taking in Tyneside English. In *Journal of Linguistics*, 22(2), 411-437.

Local, J., Walker, G. (2012). How phonetic features project more talk. In *Journal of the International Phonetic Association*, 42, 255–280. https://doi.org/10.1017/S0025100312000187

Ludusan, B., Dupoux, E. (2015). A multilingual study on intensity as a cue for marking prosodic boundaries. *Proceedings of the International Congress of Phonetic Sciences*, Glasgow, UK, 10-14 August 2015.

Meteer, M.W., Taylor, A.A., MacIntyre, R. & Iyer, R. (1995). *Disfluency annotation stylebook for the switchboard corpus.* University of Pennsylvania.

Mehl, M.R., Vazire, S., Ramírez-Esparza, N., Slatcher, R.B. & Pennebaker,

J.W. (2007). Are women really more talkative than man? In *Science*, 317, 82. https://doi.org/10.1126/science.1139940

Menz, F., Al-Roubaie, A. (2008). Interruptions, status and gender in medical interviews: The harder you brake, the longer it takes. In *Discourse & Society*, 19(5), 645-666. https://doi.org/10.1177/0957926508092247

Mondada, L. (2007). Multimodal resources for turn-taking: pointing and the emergence of possible next speakers. In *Discourse Studies*, 9(2), 194-225.

Niebuhr, O., Görs, K. & Graupe, E. (2013). Speech reduction, intensity, and F0 shape are cues to turn-taking. *Proceedings of SIGDial*, Metz, France, 22-14 August 2013.

Paggio, P., Navarretta, C. (2013). Head movements, facial expressions and feedback in conversations: empirical evidence from Danish multimodal data. In J*ournal of Multimodal User Interfaces*, 7, 29–37. https://doi.org/10.1007/s12193-012-0105-9

Pépiot, E. (2014) Male and female speech: a study of mean f0, f0 range, phonation type and speech rate in Parisian French and American English speakers. In *Proceedings of Speech Prosody*, Dublin, Ireland, 20-23 May, 2014. https://doi.org/10.21437/speechprosody.2014-48

Pépiot, E., Arnold, A. (2021). Cross-Gender Differences in English/French Bilingual Speakers: A Multiparametric Study. In *Perceptual and Motor Skills*, 128(1), 153–177. https://doi.org/10.1177/0031512520973514

Peters, B. (2006). *Form und Funktion prosodischer Grenzen im Gespräch*. Phd Dissertation. Christian-Albrechts-Universität zu Kiel, Germany.

Plug, I., Stommel, W., Lucassen, P.L., Dulmen, S.V. & Das, E. (2021). Do women and men use language differently in spoken face-to-face interaction? A scoping review. In *Review of Communication Research*, 9, 43-79. https://doi.org/10.12840/issn.2255-4165.026

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. http://www.rstudio.com/

Rossi, M., Feindt, K. & Zellers, M. (2022a). Individual variation in F0 marking of turn-taking in natural conversation in German and Swedish. *Proceedings of Speech Prosody* 2022, Lisbon, Portugal, 23-26 May 2022. https://doi.org/10.21437/speechprosody.2022-38

Rossi, M., Feindt, K. & Zellers, M. (2022b). Perception of F0 movements towards potential turn boundaries in German and Swedish conversation: background and methods for an eye-tracking study. *Proceedings of FONETIK*, Stockholm, Sweden, 13-15 June 2022.

Sacks, H., Schegloff, E., Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. In *Language* 50, 696–735.

Schiffrin, D. (1987). *Discourse markers.* Cambridge: Cambridge University Press.

Schnur, T.T., Costa, A., Caramazza, A. (2006). Planning at the phonological level during sentence production. In *Journal of Psycholinguistic Research,* 35, 189–213. https://doi.org/10.1007/S10936-005-9011-6

Skantze, D., Schlangen, D. (2009). Incremental dialogue processing in a micro-domain. *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics,* Athens, Greece, 30-31 March / 1-3 April 2009. https://doi.org/10.3115/1609067.1609150

Stivers, T., Enfield, N.J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J.P., Yoon, K. & Levinson, S.C. (2009). Universals and cultural variation in turn-taking in conversation. In *Proceedings of the National Academy of Science*s, 106, 10587–10592. https://doi.org/10.1073/pnas.0903616106

Stubbe, M. (2013). Active listening in conversation: gender and the use of verbal feedback. In S. Yamazaki and R. Sigley (Eds). *Linguistic Insights – Studies in Language and Communication: Approaching language variation through corpora: A Festschrift in honor of Toshio Saito,* 365-416. Bern: Peter Lang.

Tannen, D. (1994). *Gender and discourse.* New York: Oxford University Press.

Tannen, D. (1998). The relativity of linguistic strategies: Rethinking power and solidarity in gender and dominance. In D. Cameron (Ed.) *The Feminist critique of language: A reader.* 261–279. London: Routledge.

Truong, K.P., Poppe, R. de Kok, I. & Heylen, D. (2011). A Multimodal Analysis of Vocal and Visual Backchannels in Spontaneous Dialogs. *Proceedings of Interspeech*, Florence, Italy, 27-31 August 2011. https://doi.org/10.21437/Interspeech.2011-744

Weirich, M., Simpson, A. (2019). Effects of gender, parental role, and time on infant-and adult-directed read and spontaneous speech. In *Journal of Speech, Language, and Hearing Research, 62(11)*, 4001-4014. https://doi.org/10.1044/2019_JSLHR-S-19-0047

Yngve, V.H. (1970). On getting a word in edgewise. In *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society.

Ward, N.G., Tsukahara, W. (2000). Prosodic features which cue back-channel responses in English and Japanese. In *Journal of Pragmatics*, 32, 1177-1207. https://doi.org/10.1016/S0378-2166%2899%2900109-5

Waara, E., Shaw, P. (2006). Male and Female Witnesses' Speech in Swedish Criminal Trials. In *Journal of Language and Communication in Business*, (36), 129-156. https://doi.org/10.7146/HJLCB.V19I36.25842

Whiteside, S.P. (1995). Temporal-based speaker sex differences in read speech: A sociophonetic approach. *Proceedings of the International Congress of Phonetic Sciences*, Stockholm, Sweden, 14-19 August 1995.

Whiteside, S.P. (1996). Temporal-based acoustic-phonetic patterns in read speech: Some evidence for speaker sex differences. In *Journal of the International Phonetic Association*, 26(1): 23–40.

Zellers, M. (2016). Prosodic variation and segmental reduction and their roles in cuing turn transition in Swedish. In *Language and Speech* 60(3), 454-478. https://doi.org/10.1177/0023830916658680

Zellers, M. (2021). An overview of forms, functions and configurations of backchannels in Ruruuli/Lunyala. In *Journal of Pragmatics* 175(2021), 38-52. https://doi.org/10.1016/j.pragma.2021.01.012

Zellers, M., Gorisch, J., House, D. & Peters, B. (2019). Hand gestures and pitch contours and their distribution at possible speaker change locations: a first investigation. *Proceedings of Gesture and Speech in Interaction*, Paderborn, Germany, 11-13 September 2019.

Zellers, M., Schweitzer, A. (2017). An investigation of pitch matching across adjacent turns in a corpus of spontaneous German. *Proceedings of Interspeech,* Stockholm, Sweden, 20-24 August 2017. https://doi.org/10.21437/Interspeech.2017-811

Zimmerman, D., West, C. (1975). Sex roles, interruptions and silences in conversation. In Thorne, B., Henley, N. (Eds) *Language and Sex: Difference and Dominance*, 105–29. Rowley: Newbury House.