HAL open science

Prosodic Cues of Speaker Changes in Speech Overlap Phenomena in French

Roxane Bertrand, Robert Espesser

To cite this version:

Roxane Bertrand, Robert Espesser. Prosodic Cues of Speaker Changes in Speech Overlap Phenomena in French. 2002. hal-03223931

HAL Id: hal-03223931

https://hal.science/hal-03223931v1

Preprint submitted on 11 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Prosodic Cues of Speaker Changes in Speech Overlap Phenomena in French

Roxane Bertrand & Robert Espesser
Laboratoire Parole et Langage, UMR 6057, CNRS
Université de Provence

This study aims at determining the role of prosodic parameters in speaker changes after speech overlap phenomena in spontaneous speech dialogues in French.

Acknowledgements
Adress for correspondence: Roxane Bertrand, Laboratoire Parole et Langage, Université de Provence, 29 avenue Robert Schuman, 13621, Aix-en-Provence, France, e-mail :roxane.bertrand@lpl.univ-aix.fr

Abstract

In this study, we statistically investigate in six French spontaneous speech dialogues the role of prosodic features in speaker changes which occur after speech overlap sequences. Specifically, we focus on duration and f0 parameters in the following phases: the phase preceding the overlap, the overlap itself and the phase just after it. Using of logistic regression, we shown that five explanatory variables have a significant effect on speaker changes: the duration of each of the three phases mentioned above and specific f0 values in this overlap sequence. Especially, a short duration of the phase preceding the overlap, low f0 values at the end of this phase, and high f0 values produced by the incoming speaker at the end of the overlap increase the probability of speaker change. Conversely, very short or very long durations of the overlap phase decrease this probability. Finally, we discuss some implications of our quantitative results, comparing them with previous studies which have characterized phenomenon of speech overlaps in turn-taking.

Key words

Speech overlap

Speaker change

Dialogue

Prosody

Logistic regression

1 INTRODUCTION

Conversation Analysis (CA) provides a general perspective of analysis in the study of speech overlaps. However, the first authors who worked on CA (Sacks et al., 1974) claimed that speakers tend to reduce the number of overlaps at alternating turn points.

1.1 Smooth transitions and TRPs

"Smooth speaker transitions", that is to say without overlaps and interruptions, should consequently be the most frequent structure.
According to Sacks and collaborators, there is a turn-taking system based on basic turn units called "turn-constructional units" (TCUs) ending up at potential points of completion called "transition relevance places" (TRPs).
Thus the major concern in the study of smooth transitions seems to be the characterization of linguistic and paralinguistic features that serve to identify TRPs. About this question of the nature of turn-taking cues, numerous studies have shown that in addition to syntax, gaze or semantico-pragmatic parameters, phonetic and prosodic features could serve as turn-delimitative or turn-ending cues (Meltzer et al. 1971; Duncan & Fiske, 1977; Local et al., 1985; Local et al., 1986; Cutler & Pearson, 1986; Ford & Thompson, 1996; Koiso et al., 1998). In other terms, these features, also called "contextualization cues"[i], appear as resources in signaling and contextualizing turn-transitions (Auer & Di Luzio, 1992; Couper-Kuhlen & Selting, 1996).

1.2 Why do overlap phenomena occur?

However, although the various features so far identified seem relevant in the turn-taking system, we can ask the same question as Wells and MacFarlane (1998, p. 266), i.e "if precise mechanisms are in fact available to participants that enable them to project turn-completion, why is it that overlap occurs at all?"
The authors give two types of explanations. On the one hand, overlaps could be seen as "a byproduct of some design imperfections or latitude in the mechanisms available for projecting turn completion", and on the other hand, they could appear as "an available resource for the next speaker (the incoming speaker) to achieve particular interactional ends, for example to gain the floor", without having to wait for the main previous speaker to complete his own turn. According to the authors, both assumptions are valid and some instances of overlaps can be interpreted by participants as the expression of a competition for taking the turn while others are not assigned such interactional implications. Among different patterns, Jefferson (1987) describes for example the *blind-spot onset* case which is a typical case of the transitional onset. The blind-spot onset is defined as a case "in which the recipient/next speaker's incoming talk starts just fractionally after the start of further talk by the previous

turn-occupant, following a TRP and a pause" (1987: 165-7). This case thus includes an overlap phase which is simply the result of a difference in delay between speakers to take taken the turn, given that the reaction time in humans is 200 ms (Ward , 1996).

1.3 Overlaps, interruptions, competitivity

Several studies have attempted to identify the prosodic features of competitive versus non competitive incomings (French & Local, 1983; Wells & MacFarlane, 1998; Yang, 1999).
In trying to extend and develop the systematic characterization of interruptive talk, French and Local (1986, p. 157) hypothesized that "participants make use of certain phonetic features, -and more precisely of prosodic features- of speech with a high degree of systematicity both in producing interruptive utterances and in identifying the specific interactional function that the interruptive speech of others is performing".
Those authors thus consider as turn-competitive an incoming talk with a relatively high pitch and a loud volume (henceforth <h+f>). Wells and MacFarlane claim there are differences in the properties of the incoming talk in terms of design (phonetic characteristics) and placement (with regard to the TRPs). They judge such a talk is competitive when produced with <h+f>, and appears, for instance, before the last main accented syllable, out a TRP. With regard to cooperative interruption, it often seems to occur at low or medium pitch levels (Lyang, 1999).

1.4 A necessary distinction : overlaps versus interruption

In studies pointing out the dichotomous categorization in competitive/non competitive talk, the difference in terminology sometimes renders comparison between the results confusing. Indeed, we can find such terms as "violative interruptions" (Levinson, 1983), "turn-competitive incomings" (French & Local, 1983) or "turn-disputed utterances" (Cutler & Pearson, 1986) which refer to points of speaker interruption, against the "non competitive" or "inadvertent" overlaps terms (Levinson, 1983). But this variability does not only concern the terms used but especially highlights the difference between the phenomena analyzed: so, even if overlaps and interruptions are very close, they are not synonyms. An interruption can occur in overlap or not and, conversely, an overlap can correspond to an interruption or not. What is sometimes unclear, is whether the concern is an interruption or an overlap.
Thus, we agree with Murata (1994, p. 386) who says: "although the demarcation between overlap and interruption is not always clear-cut, their differentiation is necessary since their roles in the turn-taking mechanism are different". Overlaps are consequently classified into two types: 'turn-competing' and 'response-oriented'. This latter is divided into two subcategories, 'misprojection' (the listener misinterprets a TRP) and 'back-channel overlaps' (the listener shows he is listening and encourages the speaker to continue). Murata next establishes another distinction between 'intrusive interruption'

(subdivided into three types as 'topic-changing', 'floor-taking' and 'disagreement interruption') which implies for the speaker a notion of threat to its territory, and 'cooperative interruption' which is used to "supply a word or a phrase for which the speaker is searching, or even completes it for him/her" (1994, p. 387).

1.5 Objectives

In this paper, we have chosen to analyze the phenomenon of speech overlap which often appears, with the exception of Wells and Macfarlane's study (1998), only as a criterion to define interruption. It is thus treated in an impressionistic and anecdotic manner. Our choice is also due to the formal nature of the phenomenon of overlap which makes it easily and automatically identifiable[ii] contrary to interruptions which require other levels of linguistic analysis.
To our knowledge, no study has attempted to account precisely and systematically for speaker change after speech overlaps, which actually appears, although not always explicitly, as a main and fundamental variable in turn-taking. Moreover, the role of prosodic features in turn-taking has been previously examined, but few studies have analyzed them in a quantitative way.
In this study, we focus on change versus no change of speaker after the overlap phase. We statistically investigate how duration and pitch are involved and can be used, in accordance with Local and French, with a high degree of systematicity at potential points of speaker change. Thus, we attempt to account for some basic mechanisms involved in alternating turns in a typical structure of speech overlap.

2 METHOD

2.1 Speakers

Six spontaneous speech dialogues in French (south French) were recorded. In each, the two were seated face to face in a sound-proof room. Despite such relative unnatural speaking conditions, interactions between speakers sounded natural and fluent. No discussion topic was provided and each speaker was free to speak or remain silent. This absence of constraint was motivated by our desire to create a situation as close as possible to those in everyday life where speakers have to manage their mutual presence and eventually interact and speak with one another.

2.2 Material

The type of experiment used in this study requires that each speaker be recorded on a separate channel and thus be equipped with laryngophone. The average duration of each dialogue being

approximately 20 minutes, the overall duration of the corpus is about 2 hours.

2.3 Unit of analysis

2.3.1. The interpausal unit (IPU)

Similarly to Koiso and collaborators (1998), we used the interpausal unit (IPU) as unit of analysis, defined as the production of speech between two silent pauses. In this study, pauses were identified as unvoiced segments longer than a threshold which is fixed at 200 ms (corresponding to the mean duration of brief silent pauses). The durations of IPUs beginning and/or ending with unvoiced consonants were consequently underestimated but we hypothesized that this did not introduce a bias in our results.
The choice of pauses as delimiters resolves difficulty in objectively and reliably defining the unit, in opposition to other types of units (grammatical, turn or prosodic units) also used in spontaneous speech. Moreover, as highlighted by Koiso et al., prosodic units such as "intermediate" and "intonational phrases" (Pierrehumbert & Hirschberg, 1990), for instance, cannot be extracted automatically and require hand-labeling by experts.
The reasons for the choice of the IPU were then two-fold:

(1) it is a formal criterion of identification
(2) it is easily detectable

2.3.2. IPUs segmentation

All speech analyses were performed with procedures available in the Mes software (Espesser, 1996). Pitch and voicing were first detected from the speech signals with a combination of three methods of pitch detection. The speech signals were then automatically divided into IPUs using the voice/unvoiced criterion.

2.4 Overlaps

2.4.1. Definition of Overlaps

Overlaps are considered as two simultaneous part of IPUs and were automatically detected by the presence of interleaved or embedded initial and final IPU labels.
Table 1 provides the basic data on number, mean duration and standard deviation (in ms) of IPUs and speech overlap phases for each dialogue.

Table 1. Basic data for IPUs and Speech Overlap Phases for Each Dialogue
unit duration is ms

	speakerA			*speakerB*					
	IPUs			**IPUs**			**Speech Overlaps**		
dial	N	Mean	SD	N	mean	SD	N	Mean	SD
1	418	479	466	591	790	789	137	252	268
2	696	914	844	492	957	921	323	327	298
3	626	975	840	532	758	933	318	365	378
4	493	622	614	688	676	645	203	251	234
5	489	774	890	624	1025	997	292	331	346
6	259	761	796	447	998	960	129	269	268
total	6355						1402		

2.4.2. The POC sequence

A pattern we named "*POC sequence*" was taken as the basic unit for our study.

P = *preceding phase*, i.e corresponds to the part of the first IPU until the overlap incoming

O = *overlap phase*, i.e corresponds to the phase containing a part of the two IPUs, -the two simultaneous voices of the main previous speaker (PS) and of the incoming speaker (IS)-

C = *continuation phase*, i.e corresponds to the part of the IPU following the overlap phase

A POC sequence consists of two IPUs, one for each speaker (see figure 1 and 2). POCs are obtained by automatically extracting the specific label patterns from the sequence of IPU labels.

Figures 1 and 2 show instances of POC sequences: in the first one, the same speaker intervenes before and after the O phase; in the second one, there is a change of speaker after the O phase.

```
              P              O                C
PS    |______________|______________|______________|      IPU1
IS                   |______________|                     IPU2
```

figure 1: no speaker change (C=P)

```
              P              O              C
PS    |______________|______________|                     IPU1
IS                   |______________|______________|      IPU2
```

figure 2: speaker change (C ≠ P)

A POC sequence is considered here as the typical structure accounting for the change/hold of a speaker and appears as the most frequent pattern in this corpus (about 2 per minute).

A total of 704 POCs were numbered in our corpus but only occurrences with all three phases displaying minimal durations of 130 ms each were chosen. This value corresponds to the mean duration of a typical unaccented French syllable in spontaneous speech. The

object of this selection was to keep the more stable and homogeneous linguistic units[iii]. This does not imply that cases below 130 ms are irrelevant in turn-taking transitions but may function differently.
277 POCs were remained after this selection.

3 STATISTICAL ANALYSIS

3.1 Choice of the model

A linear model appears particularly appropriate to analyze conversational speech data, which are often less controlled. A logistic model thus was chosen to account for the effects of prosodic parameters of turn-taking alternation following speech overlap phases. We mainly considered this model as an explanatory tool.
All statistical analyses were performed using the R software (Ihaka & Gentleman, 1996).

3.2 Dependent variable and explanatory variables

3.2.1. Dependent variable

The binary dependent variable is the presence/absence (coded 1/0) of speaker change after the speech overlap phase.

3.2.2. Explanatory variables

Parameters. Durations and fundamental frequency (f0) values in the POC sequence were considered. Speech energy values in the laryngographic signals were not accurate enough to be considered.
Durations. The three durations P, O, and C (in ms) were retained.
Fundamental Frequency. We first modeled the raw f0 curve by a continuous (i.e without voiced/unvoiced discontinuities) and smooth curve deprived of microprosodic variations, irrelevant for the analysis (Hirst et al., 2000).
In order to reduce interspeaker variability, especially the male/female variability, we used the speaker normalized f0 (SNf0), defined for each speaker as follow:

SNf0 = f0 (Hz)/mean_f0_speaker (Hz)

Mean_f0 speaker is computed over the whole duration of the dialogue.

Five specific points of the f0 curve were retained. Indeed, the relevant melodic information used by speakers to regulate their turns is located at specific points such as the final mora region (Koiso et al., 1998) or the last syllable of the turn (Wells & MacFarlane, 1998). It might also be found immediately before turn changes (Duncan & Fiske, 1977; Koiso et al., 1998). Accordingly, we chose a 130 ms window just before the speech overlap phase (concerning the previous speaker) and another 130 ms window at the end of the speech overlap phase of each speaker. We also chose a 130 ms window at the initial of the speech overlap phase of each speaker since the beginning of the incoming

talk is otherwise relevant to characterize competitive/non competitive talk (Wells & McFarlane, 1998) (see figure 3).

```
                 P                 O                    C

PS     |_____________Pe|POb_________POe|

IS                        |IOb__________IOe|_______________|
```

figure 3 : the five selected f0 localization points

The following eight explanatory variables were selected

- durP: P phase duration (ms)
- durO: O phase duration (ms)
- durC: C phase duration (ms)
- mf0Pe : mean SNf0 of the last 130 ms of P phase
- mf0POb: mean SNf0 of the first 130 ms of PS in O phase
- mf0POe : mean SNf0 of the last 130 ms of PS in O phase
- mf0IOb: mean SNf0 of the first 130 ms of IS in O phase
- mf0IOe: mean SNf0 of the last 130 ms of IS in O phase

P = Previous (speaker), which also refers to the phase preceding O (in which only the previous speaker appears)
I = Incoming (speaker)
b = beginning of the phase
e = end of the phase,
for example, POb means 'beginning of the Overlap of the Previous speaker'

3.2.3. Data screening

As we focused on the explanatory power of the model, we eliminated a priori redundant variables. The pair-wise correlations of the 8 variables revealed the following two major linear dependencies:

(1) High correlation of mf0Pe and mf0POb (R=0.71) probably due to the contiguity of the Pe and POb zones.

(2) Correlation of mf0POb and mf0POe (R=0.53)

The following two observations were made related to the localization of the relevant f0 points (see 3.2.2): since mf0Pe immediately precedes the speech overlap phase, it should be more relevant than mf0POb. In the same way, since mf0POe immediately precedes the change of speaker, we assume that it has a more important effect than mf0POb. These observations led us to eliminate the mf0POb variable, which additionally reduces the risk of collinearity in the model.
Table 2 presents basic statistics of the explanatory variables, in relation to the change/no change of speaker.

Table 2. Statistics of the Explanatory Variables
unit duration is ms, f0 values are speaker normalized f0

	durP	**durO**	**durC**	**mf0Pe**	**mf0IOe**	**mf0IOb**	**mf0POe**
				change (117)			
1st Qu.	220	240	410	0.8500	0.9604	0.8838	0.8312
Median	410	350	740	0.9402	1.1156	1.0130	0.9388
Mean	611	440	932	0.9772	1.1207	1.0610	1.0146
3rd Qu.	790	560	1270	1.0944	1.2656	1.1480	1.1286
				no-change (160)			
1st Qu.	330	190	345	0.8839	0.8611	0.8833	0.8836
Median	625	290	530	1.0038	0.9596	0.9886	0.9815
Mean	867	407	714	1.0768	0.9923	1.0310	1.0538
3rd Qu.	1170	473	913	1.1964	1.1031	1.1180	1.1719

It should be noted that 75% of the P durations are greater than 260 ms, (change and no-change pooled together), i.e at least two syllables, and that 75% of the C durations are greater than 370 ms, i.e about three syllables at least. Such long durations are a confirmation of speaker change whereas shorter durations could correspond to a mere time-shift between the two turn endings.

3.3 Model 1

277 cases and 7 variables (durP, durO, durC, mf0Pe, mf0Poe, mf0IOb and mf0IOe) were retained.
Table 3 summarizes the results of Model 1.

Table 3. Summary of the Logistic Regression Analysis for Variables Predicting Change (model 1)

	Estimate	Std. error	z value	Pr(>|z|)
(Intercept)	-0.67	0.8834	-0.758	0.448
durP	-8.415e-04	2.457e-04	-3.425	6.16e-04
durO	4.415e-04	4.193e-04	1.053	0.2923
durC	6.269e-04	2.203e-04	2.845	4.435e-03
mf0Pe	-1.915	0.6406	-2.989	2.797e-03
mf0POe	-0.2753	0.4766	-0.578	0.5635
mf0IOb	-0.3125	0.5479	-0.570	0.5683
mf0IOe	2.6889	0.632	4.255	2.09e-05

According to Menard (1995), the z-scores of the logistic regression analysis are not very reliable. A deviance analysis (table 4) was therefore run. Starting from the Null model (i.e with only the intercept), each parameter was sequentially added to the model.

Table 4. Deviance table. Terms added sequentially (first to last) (model 1)

	Deviance resid.	Df Resid.	Deviance	P (>|Chi|)
NULL		276	377.30	
durP	11.41	275	365.89	7.310e-04
durO	0.68	274	365.21	0.41
durC	7.91	273	357.30	4.929e-03
mf0Pe	10.90	272	346.40	9.599e-04
mf0POe	0.28	271	346.12	0.60
mf0IOb	1.89	270	344.23	0.17
mf0IOe	20.58	269	323.65	5.724e-06

DurP, durC, mf0Pe and mf0IOe demonstrate the following statistically significant effects on the dependent variable at the 5% level:

- the shorter durP, the higher the probability of change
- the longer durC, the higher the probability of change
- the lower mf0Pe, the higher the probability of change
- the higher mf0IOe, the higher the probability of change

DurO, mf0IOb and mf0POe have no significant effects.
Two-way interactions did not yield significant results.

3.3.1. Goodness of fit

The logistic Model 1 accounts for about 23% of the variability in the dependent variable (Nagelkerke R^2 = 0.23). This low percentage confirms that we cannot use this model as a prediction tool.
Other measures of goodness of fit are the correct classification rates (CCR). The CCR show (table 5) that Model 1 does not satisfactorily classify the change items but provides a correct classification of the no-change items.

Table 5. Observed and Predicted Group Memberships for the Dependent Variable (Model 1)

		Predicted		
		no change	change	% correct
Observed	no change	131	29	81
	change	55	62	52.9
Overall				69.67

3.3.2. Analysis of residuals

In logistic regression, the errors are assumed to have a binomial distribution which approximates a normal distribution for large samples (Menard, 1995). The distribution of the standardized deviance residuals in model 1 satisfied these assumptions: the standardized deviance residuals had a mean value of –0.04 (standard deviation = 1.10). 272 cases out of 277 (98.2 %) had a standardized residual value between –2 and + 2. The Q-Q plot revealed moderate deviations from the diagonal line. A gap for the residuals close to zero was observed, as commonly occurs when predictions are weak.

3.3.3. Influential cases

To detect possible weaknesses in the model, a second analysis was carried out without the influential cases, i.e the data which have a disproportionate influence on the estimates of the regression coefficients.
Let *n* be the number of observations (277) and *p* the number of estimated parameters (8). Influential cases can be determined with a Cook statistic value greater than 8/(n-2p) and/or a standardized leverage greater than 2p/(n-2p). 23 potential influential cases were founded among the 277 initial cases. Model 1 was then computed

without these 23 cases (model 1b). Both models 1 and 1b reveal similar results. The main difference concerns the deviance explained by mf0IOb which becomes marginally significant (p=0.02). The 23 influential cases do not influence the regression enough in model 1 to bias the estimates.

3.3.4. Improvement of model 1

Occurrences with durO above 1400 ms are among the influential removed cases. Examination of the smoothed partial residuals plots showed a non-linearity for durO. We conclude from these two observations that the longer durO are not well modelled; a quadratic transformation of durO defined by:

$$qdurO = (durO - 800)^2$$

may improve the model.

3.4 Model 2

277 cases and 6 variables (durP, qdurO, durC, mf0Pe, mf0IOb and mf0IOe) were retained for this model. The non-significant mf0POe variable was removed.
The dependent variable (change) is the same as in Model 1.

Table 6. Summary of the Logistic Regression Analysis for Variables Predicting Change (Model 2)

	Estimate	Std. error	z value	Pr(>|z|)
(Intercept)	-8.946e-02	9.109e-01	-0.098	0.9217
durP	-8.582e-04	2.492e-04	-3.443	5.75e-04
qdurO	-2.716e-06	9.474e-07	-2.867	4.144e-03
durC	6.462e-04	2.205e-04	2.931	3.378e-03
mf0Pe	-1.900	6.138e-01	-3.096	1.963e-03
mf0IOb	-5.307e-01	5.455e-01	-0.973	0.3306
mf0IOe	2.852	6.568e-01	4.343	1.41e-05

Table 7. Deviance Table. Terms Added Sequentially (first to last) (Model 2)

	Deviance resid.	Df Resid.	Deviance	P (>|Chi|)
NULL		276	377.30	
durP	11.41	275	365.89	7.310e-04
qdurO	9.25	274	356.65	2.361e-03
durC	8.45	273	348.19	3.642e-03
mf0Pe	10.62	272	337.58	1.121e-03
mf0IOb	0.92	271	336.65	0.34
mf0IOe	21.68	271	315.9	3.223e-06

As expected, table 7 shows that durP, durC, mf0Pe and mf0IOe are still significant, as in model 1. Mf0IOb is not significant.
A noteworthy difference concerns the qdurO variable, which is now markedly significant at the 5% level.

The negative coefficient associated with qdurO corresponds to a parabola with a maximum at 800 ms: the further from the 800 ms durO, the lower the probability of change.

3.4.1. Goodness of fit

Model 2 appears slightly better than Model 1 as shown by the Nagelkerke R^2 (0.27) and the CCR (table 8).

Table 8. Observed and Predicted Group Memberships for the Dependent Variable (Model 2)

		Predicted		
		no change	change	% correct
Observed	no change	125	35	78.1
	change	47	70	59.8
Overall				70.4

Overall CCR is almost the same as in Model 1 but scores for each group are more balanced. Compared to found values in Model 1, no change CCR slightly decreases from 81% to 78.1%, but change CCR clearly increases from 52.9 % to 59.8%.

3.4.2. Analysis of residuals

The distribution of the standardized deviance residuals is close to that found in Model 1. A second analysis was further carried on without the 22 potential influential cases found in Model 2. Results for Model 2 did not seem biased by these cases.
Mf0IOb remains unsignificant which ensures it was only marginally significant in model 1b.

3.4.3. Discussion

This time, the effect of durO has been taken into account since qdurO improves the modeling of the occurrences with very long durO values (i.e above 800 ms), which were mainly associated with a no-change of the speaker. Two other observations support this point: 1/ in a previous study (Bertrand & Espesser, 2001), we filtered out durO values greater than 1200 ms and found that the linear coefficient durO became significant; 2/ occurrences of durO above 1400 ms were considered as influential cases in Model 1, but not in Model 2. Nevertheless we cannot rely on the right side of the parabola (i.e durO above 800 ms) as much as on the left one, due to the scarcity of tokens on the right side.

3.4.4 Odd Ratios (OR)

The Odd Ratio (table 9) highlights the substantive significance of the effect of each variable on the dependent variable.

Table 9. Odd Ratio (Model 2, without mf0IOb)

variable	v1	v2	IQR Increase	Odd ratio
durP	260	1030	770	0.52 (1/1.93)
durC	370	1030	660	1.53
mf0Pe	0.8667	1.1583	0.2916	0.57 (1/1.76)
mf0IOe	0.8915	1.1778	0.2863	2.1

variable	v1	v2	variation	Odd ratio of qdurO
durO	800	400,1200	-400,+400	0.66 (1/1.52)

The OR is the multiplicative coefficient of the estimated probability when a variable switches from a value v1 to a value v2, all the other variables being fixed. For a continuous variable, typical ranges have been selected. Except for durO, we have chosen the interquartile range (IQR); for instance, when durP increases by one IQR (i.e 770 ms) the estimated probability of change is multiplied by 0.52 (i.e divided by 1.93, because the estimated coefficient durP is negative); and when durC increases by one IQR (660 ms), the estimated probability is multiplied by 1.53.
Concerning durO, starting from the maximum of the parabola at 800 ms, a variation of +/- 400 ms multiplies the estimated probability by 0.66.

4 GENERAL DISCUSSION

Based on a total of 277 cases of POC sequences, the logistic regression model accounts for a significant portion of the variability of the speaker change following speech overlap phases. Out of the 7 explanatory variables, the 5 variables durP, durO, mf0Pe and mf0IOe are significant while mf0IOb and mf0POe are not (model 2).
The implication of each variable on the dependent variable is examined here.

4.1 Durations variables

Shorter P durations imply a higher probability of speaker change. This suggests that it is easier for the incoming speaker to take the turn when the previous speaker is interrupted early and does not have sufficient time to get involved into his own speech. Relatively to the turn-taking system, we suggest that both speakers may consider the presence of a pause before the P phase as a TRP cue. If such was the case, the speech overlap phase could then result from a longer delay in taking the floor from IS. According to Ward (1996), the human reaction time is about 200 ms, which corresponds to a short P (the first quartile of durP with speaker change is 220 ms; see table 2). This type of configuration can refer to the “blind-spot case”, which belongs to the « transitional onset » category of Jefferson and precisely defined as a case “in which the recipient/next speaker’s incoming talk starts just fractionally after the start of further talk by the previous turn-occupant, following a TRP and a pause” (1987: 165-7). If the blind-spot belongs to the class of short P durations, the blind-spot situation would clearly increase the probability of speaker change, durP having

a strong effect on the probability of speaker change (odd ratio = 0.52, see table 9). Jefferson evokes this point when she says "the blind-spot onset can result from the incoming speaker's moving from a recipientship into a speakership mode at the TRP, (i.e after the original turn-occupant's previous utterance) and consequently, when s/he starts her/his own turn, not attending to the original turn-occupant's resumption of talk" (cited in Wells & Macfarlane, 1998, p. 277).

Shorter P durations can also refer to the "progressional onset" category. In this case, the incoming speaker would be the real main speaker -before the POC sequence- which had a difficulty (a lack of word for instance). The P phase would be then only a completion from the previous speaker in order to help the incoming speaker which takes his turn in O (legitimately his) and continues in C.

O durations exhibit a more complex behaviour. Firstly, shorter O durations imply a lower probability of speaker change. These can characterize back-channel signals, which often consist of 1 or 2 syllables[iv], such as "uh huh" ("mmummum") or "yeah" ("ouais"). They are used by listeners to signal sustained attention to the speaker (Roger, 1989) and also serve to extend floor-holding by the speaker. Their function in conversation thus appears to be the opposite of that found in interruptive simultaneous speech. Concerning this point, we mentioned above that in many studies the cooperative overlap category consists mainly in these back-channel signals.

As durO increases up to 800 ms, the probability of speaker change increases (cf. table 9). Since the shorter O durations were associated with the back-channel signals which correspond, by definition, to a minimal probability of speaker change, the lenghthening of durO necessarily increases the change probability. We interpret this result in terms of involvement. Due to his status of initiator of the overlap, we suppose that IS is more involved in his talk and that he has something to say. The longer duration of O, the more IS is involved in his talk. The speakers' alternation is favoured around 800 ms which appears as the optimal duration for a change in the talk.

On the other hand, above 800 ms, the probability of speaker change decreases. We suggest that such extended durations constitute a form of competition for the previous speaker, even in a case of a very cooperative overlap as in a long phase of agreement -indeed, whatever its form, a sort of competition occurs for who will take the next turn-. We suggest that this time zone around 800 ms could be a threshold[v] beyond which PS shows his willness to preserving his main speaker status, IS becoming then more and more illegitimate.

4.2 Pitch variables.

Lower mf0Pe values -which are clearly low on a perception scale since the first quartile of mf0Pe is 2.5 semitones lower than the speaker mean f0- imply a higher probability of speaker change. Such lower values which occur at the end of utterances immediately preceding the speech overlap phase could be perceived as turn-ending or less involvement cues. The P phase could be judged at this point informative enough by IS who begins his own turn while in fact the previous speaker has not yet relinquished his own speech. In Jefferson's onset categories, the "item-targeted-recognitional onset"

(in the interjacent onset category) refers to a point of semantic adequacy which is still not a TRP (syntactic point of completion). In such a case, the incoming speaker anticipates on the turn-ending which is the reason of the overlap phase. Wells and Macfarlane define the TRP as a transition space. Due to the character of terminality which is often associated with the lower values, these latter can be a cue among others of the presence of a (future) TRP. These lower values can refer to a lengthening or filled pause like "euh" which very often occurs with low pitch values (Guaïtella, 1996). "Euh" fulfills at least two roles in discourse as turn-ending cue when the speaker gives the turn up to the interlocutor, or as hesitation phenomenon when the speaker produces it in order to keep the floor during the conceptualization phase of the turn continuation. Whatever its function, IS can always support he perceived a turn-ending cue due to the lower values.

Higher mf0IOe values,–which are clearly high on a perception scale since the third quartile of mf0IOe is 2.8 semitone higher than the speaker mean f0- imply clearly a higher probability of speaker change, mf0IOe having the greatest odd ratio (2.1; see table 9). High values of f0 often illustrate an intention of continuation which then characterize a state of the turn in progress. This could be trivial and we could then expect the same effect for POe since IOe and POe zones are formally alike.

Yet, mf0POe is not significant (model 1, table 4), i.e there is no typical pitch value indicating whether PS continues to speak or not. The Goodwin's definition of TRP, which is in the same frame of mind of Wells and Macfarlane could provide an explanation. According to the author, a point of possible completion can appear after a pitch peak if it is associated with specific syntactic and pragmatic conditions. If we admit that lower values of Pe are a cue of the presence of a possible completion, the overlap phase consists of the "locus" of Goodwin (1986, in Wells & Macfarlane, 1998, p. 280) in the following definition which could provide a clue on the lack of significance of the POe variable:

> "Just after such a pitch peak is the locus for various orderly phenomena: it is where early-starting next turns regularly come in; it is where speakers initiate a 'rush-through' (Schegloff, 1982) if they mean to extend their talk through the transition space into a new turn-constructional unit; it is where continuers and other forms of interpolation into otherwise projectably extended spates of talk are placed if they overlap with the otherwise ongoing talk".

The overlap phase following lower values of Pe for instance (which have an higher probability of speaker change) can contain various phenomena such as "euh" (which occur in a low pitch as we mentionned above) but also phatic as "you know", which can be occurred with high pitch (Local et al., 1985), "you know what I mean" and so on. Accordingly, these different phenomena occurring with various pitch values lead to a great variability of mf0POe, which explains it is not significant. Therefore, there cannot exist a parallelism between mf0IOe and mf0POe. The both important points relating to the speaker change in the POC sequence are Pe for the previous speaker which manifests here his willness to stop to speak or not and in the same way IOe for the incoming speaker.

The last f0 parameter mf0IOb is easier to interpret. It has no effect on speaker change. We mentionned that this variable is a component of competitivity. As expected, our result illustrates that competitive overlap can not be reduced to speaker change.

4.3 Combination of variables : the back-channel pattern.

In many studies, the criterion of non interruption (i.e no speaker change in our work) is the main criterion for defining the cooperative overlaps category which mainly consists of back-channel signals.
We can then evaluate how the combination of the parameters presented here -as they occur in a real speech interaction- are in accordance with previous studies on the back-channel (BC).
We therefore defined a standard pattern by the following set of constraints on our parameters. We retained then a short O duration since a typical BC consist of 1 or 2 syllables (Roger, 1989); low mf0IOe because BC is often produced in a low pitch register (Müller, 1996) and low mf0Pe since BC is preferentially preceded by a low pitch point (region) (Ward, 1996). We finally retained a mid-long P duration since BC occur at points of completion (syntactic, semantic) or at utterance or clause endings (Koiso et al., 1998). Each of three -O duration, mf0IOe, mf0Pe- is smaller than its first respective quartile; P duration is greater than its median. There was no constraint on C duration. Whatever the values within these constraints, the model 2 (without mf0IOb) always computed a low probability of change, varying from 0.12 to 0.39 (first and last deciles) with a median at 0.23. The combined effects of durO, mf0IOe and durP counterbalance the effect of a low mf0Pe and ensure a low probability as expected for a BC. This result illustrates that the relative weights of the parameters in model 2 are correctly estimated for the BC pattern.

5 CONCLUSION

This study shows that prosodic features of duration and f0 play a role in speaker changes after speech overlap phases. Specifically, a short phase preceding the overlap, low f0 values produced by the previous speaker at the end of this same phase, and high f0 values produced by the incoming speaker at the end of the overlap phase, increase the probability of speaker change. The overlap phase duration also plays a role in speaker change: mid-durations increase this probability, whereas very short or very long durations decrease it. Otherwise, our results can be related to phenomena found in turn-taking studies, such as the blind-spot onset category or back-channel phenomena. In the blind-spot onset, the two speakers start to speak nearly simultaneously; in other words, the duration phase preceding the overlap is very short. According to our results, blind-spot cases increase the probability of speaker change. Back-channel phenomena consist of short productions from the incoming speaker, such as “hum”, “yeah”, which are used to signal he is listening to and he is not wanting to take the floor. According to our model, the prosodic features of back-channel phenomena actually decrease the probability of speaker change. Moreover, this result supports the hypothesis that

back-channels are systematically and mechanically produced with a particular characteristic pattern which transcends language boundaries. One study which mentions their important role in dialogues is Ward and Tsukahara's (2000). Following these authors who worked on English and Japanese, it appears important to assess the characteristic prosodic pattern of back-channels in French and, more specifically to describe those prosodic cues which condition its presence.

---

[i] See particularly Gumperz (1982), for the "contextualization approach".
[ii] Whenever possible, we have opted here for an automatic procedure in order to reduce especially the preliminary segmentation stages.
[iii] However, it is noteworthy that the syllable status is controversial in recent psycholinguistic studies (see Goslin and Frauenfelder, 2001).
[iv] In our corpus, the mean duration value of back-channel signals (manually labeled) is 285 ms.
[v] There is no sharp boundary since the negative curvature of the parabola is weak : for a variation of 130 ms around 800 ms, the probability is only modified by 4%.

REFERENCES

AUER, P., & DI LUZIO, A. (1992). *The contextualization of Language*. Amsterdam and Philadelphia: Benjamins.

BERTRAND, R., & ESPESSER, R. (2001). "About speech overlaps. : prosodic cues contribution in predicting a change of speaker". In S. Puppel & G. Demenko (Eds.), *Prosody 2000* Speech recognition and synthesis (pp. 29-35). Krakow.

COUPER-KUHLEN, E., & SELTING, M. (1996). *Prosody in Conversation*. Cambridge University Press.

CUTLER, A., & PEARSON, M. (1986). "On the analysis of prosodic turn-taking cues". In C. John-Lewis (Ed.), *Intonation in discourse* (pp. 139-155). Croom Helm, College-Hill Press.

DUNCAN, S., & FISKE, W.H. (1977). *Face to face interaction : research, methods and theory*. Hillsdale, NJ: Lawrence Erlbaum.

ESPESSER, R. (1996). "MES : Un environnement de traitement du signal". Actes des *XXIe Journées d'Etude sur la Parole*, Avignon, 447.
Online : http://www.lpl.univ-aix.fr/projects/mes_signaix

GUAÏTELLA , I. (1996). "Analyse prosodique des hésitations vocales : propositions pour un modèle rythmique". *Revue de Phonétique Appliquée*, **118/119**, 113-144.

GOSLIN, J., & FRAUENFELDER, U.H. (2001). "A comparison of theoretical and human syllabification". *Language and Speech*, **44 [4]**, 409-436.

FORD, C.E., & THOMPSON, S.A. (1996). "Interactional units in conversation: syntactic, intonational and pragmatic resources for the management of turns". In E. Ochs, E.A Schegloff & S.A. Thompson (Eds.), *Interaction and Grammar* (pp 134-184). Cambridge, UK: Cambridge University Press.

FRENCH, P. & LOCAL, J. (1983). "Turn-competitive incomings". *Journal of Pragmatics*, **7**, 17-38.

FRENCH, P., & LOCAL, J. (1986). "Prosodic features and the management of interruptions". In C. John-Lewis (Ed.), *Intonation in discourse* (pp. 157-180). Croom Helm, College-Hill Press.

GUMPERZ, J.J. (1982). *Discourse strategies*. Cambridge University Press.

HIRST, D.J., DI CRISTO, A., & ESPESSER R. (2000). "Levels of representation and levels of analysis for the description of intonation systems". In M. Horne (Ed.), *Prosody: Theory and Experiment* (pp. 51-87). Kluwer Academic Publishers.

IHAKA, R., & GENTLEMAN, R. (1996) "R: A language for data analysis and graphics". *Journal of Computational and Graphical Statistics*, **5 (3)**, 299-314.

JEFFERSON, G. (1987). "Notes on "latency" in overlap onset". In G. Button, P. Drew & J. Heritage (Eds.), Interaction and language use. Special issue of *Human Studies*, **9**, 153-183.

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 dialogues". *Language and Speech*, **41 (3-4)**, 295-321.

LEVINSON, S. (1983). *Pragmatics*. Cambridge University Press.

LOCAL, J.K., KELLY, J. & WELLS, W.H.G. (1986). "Towards a phonology of conversation: turn-taking in Tyneside English". *Journal of Linguistics*, **22**, 411-437.

LOCAL, J.K., WELLS, W.H.G., SEBBA, M. (1985). "Phonology for conversation: phonetic aspects of turn-delimitation in London Jamaican". *Journal of Pragmatics*, **9**, 309-330.

MELTZER, L., MORRIS, W.N., & HAYES, D.P. (1971). "Interruption outcomes and vocal amplitude: explorations in social psychophysics". *Journal of personality and Social Psychology*, **18**, 392-402.

MENARD, S. (1995). *Applied logistic regression analysis*. Thousand Oaks, London , New Delhi, Sage Publications.

MÜLLER, F.E. (1996)."Affiliating and disaffiliating with continuers". In E. Couper-Kuhlen & M. Selting (Eds.), *Prosody in Conversation* (pp. 131-176). Cambridge University Press.

MURATA, K. (1994). "Intrusive or cooperative? A cross cultural study of interruption". *Journal of Pragmatics*, **21**, 385-400.

PIERREHUMBERT, J.B., & HIRSCHBERG, J. (1990). "The meaning of intonational contours in the interpretation of discourse". In O.R. Cohen, J. Morgan, & M.E. Pollack (Eds.), *Intentions in communication* (pp.271-311). Cambridge, MA: MIT Press.

ROGER, D. (1989). “Experimental studies of dyadic turn-taking behaviour”. In D. Roger & P. Bull (Eds.) Peter Bull Editors, *Conversation* (pp. 75-95), , Multilingual Malters LTD.

SACKS, H., SCHEGLOFF, E.A., & JEFFERSON, G. (1974). “A simplest systematics for the organization of turn-taking for conversation”. *Language*, **50**, 696-735.

WARD N. (1996). “Using prosodic clues to decide when to produce back-channel utterances”, Proceedings *ICSLP 96*, Philadelphie, **3-4**, 1728-1731.

WARD, N. & TSUKAHARA, W. (2000). “Prosodic features which cue back-channel responses in english and japanese”, *Journal of Pragmatics*, **23**, 1177-1207.

WELLS, B. & MACFARLANE, S. (1998). “Prosody as an interactional resource: turn-projection and overlap”. *Language and Speech*, **41 (3-4)**, 265-294.

YANG, L. (1999). Prosodic correlates of interruptions in spoken dialogue, Proceedings ESCA, *Eurospeech 99*, Budapest, 239-242.