Modelling Turn-taking in a Simulation of Small Group Discussion

Emiliano Gomes Padilha

Doctor of Philosophy

Institute for Communicating and Collaborative Systems

School of Informatics

University of Edinburgh

2006

Abstract

The organization of taking turns at talk is an important part of any verbal interaction such as conversation, particularly in groups. Sociologists and psycholinguists have been studying turn-taking in conversation through empirical and statistical analysis, and identified some systematics in it. But to my knowledge no detailed computational modelling of verbal turn-taking has yet been attempted.

This thesis describes one such attempt, for a simulation of small group discussion—that is, engaged conversation in groups of up to seven participants, which researchers have found to be much like two-person dialogues with overhearers. The group discussion is simulated by a simple multi-agent framework with a blackboard architecture, where each agent represents a participant in the discussion and the blackboard is their channel of communication, or 'environment' of the discussion. Agents are modelled with just a set of probabilistic parameters that give their likelihood of doing the various turn-taking decisions in the simulation: when to talk, when to continue talking, when to interrupt, when to give feedback ("uh huh"), and so on. The simulation, therefore, consists of coordinating a one-at-a-time talk (symbolic talk) with speaker transitions, hesitation, yielding or keeping the *floor*, and managing simultaneous talk which occurs mostly around speaker transitions.

The turn-taking modelling considers whether participants are talking or not, and *when* they reach points of *possible completion* in their utterances that correspond to the places of transition-relevance, TRPs, where others could start to speak in attempts to take a new turn of talk. The agent behaviours (acts), their internal states and procedures are then described. The model is expanded with elaborate procedures for the resolution of simultaneous talk, for speaking hesitations and their potential interruption, and for the constraints of the different 'sorts' of utterance with respect to turn-taking: whether the TRP is free, or the speaker has selected someone to speak next, has encouraged anyone to speak, or has indicated the course of an extended multi-utterance turn at talk as in sentence beginnings like "first of all," or "let me tell you something:. . . ".

The model and extensions are then comprehensively analysed through a series of large quantitative evaluations computing various aggregate statistics such as: the total times of single talk, multiple talk and silences; total occurrences of utterances, silences, simultaneous talk, multiple starts, middle-of-utterance attempts at talking, false-starts, abandoned utterances (interrupted by others), and more.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Emiliano Gomes Padilha*)

Acknowledgements

My supervisor Jean Carletta, of course, was the main helper in conceiving and organizing the initial and later ideas behind this thesis. Without her orientation, and most importantly, concern, it probably might not have come to light. The basic design of the framework and the use of probabilities as a way to simulate speakers' decisions, for example, was her suggestion. Also, being new to the subject, many concepts of turn-taking were clarified by her. The structure of the thesis and its developing ideas were greatly improved by our discussions too.

My second supervisor, Mark Steedman, was also a good help in the initial kickstart process of the thesis: what to do, and how. My two examiners, John Lee and Richard Power, contributed to improve the thesis with their comments and revisions, especially regarding the evaluation chapter.

My officemates Alistair Gill, Timothy Willis, Carsten Brockmann, Jane Neumann, among others, certainly did not help me with my thesis (just kidding!), but made the time worthwhile with their friendship, wit, and technical assistance. To the people in the secretary of the department, David Dougal, Betty Hughes and others, for the friendliness and help with the submission and other course details.

To my teacher and friend Rosa Viccari, pela força, ajuda e confiança. E aos brasileiros e outros que proveram o calor humano que faltava para afastar minha solidão nesta terra de gente fria e distante ('polida' como gostam de pensar que são): ao Ricardo, Miguel, Constantino, Natasha, Kiko, Pedro, Gabriela, Zé Carlos, Bruno, Virna, e outros.

A minha mãe, a quem devo tudo: e acima de tudo, amor e apoio incondicionais.

A meu pai e a Severina (babá), que já se foram.

Me lembro bem dele naquelas manhãs de sol no terraço
no banco bem feito por ele; daqueles tão poucos cabelos
voltando às tardes do centro com a pasta embaixo do braço.
Me lembro bem dele arrumando à noite a coleção de selos,
batendo à máquina o livro das memórias de Jaboatão,
ou madrugando na televisão, insone e silente;
Das grandes mãos frias acolchoadas e do aperto de mão
caloroso dos últimos dias, já entanto bastante doente.
Me lembro bem dele na cama do hospital quando seu coração
havia falhado e o meu também indesculpavelmente.
Mas o tempo não se permite retornar para consertar enganos,
e agora só hei dele lembrar com um imenso vazio pesar
quando neste dez de janeiro já lá se vão quatro anos.
Por intermédio da internet do céu então eu queria mandar
uma afetuosa lembrança que vai por toda a família:
Que Deus abençoe meu pai, Givaldo Gomes Padilha.
(11/Jan/2004)

No human has ever learnt to speak except in a dialogic context.

— Jaffe & Feldstein, *Rhythms of Dialogue* (1970)

Moving from the study of sentences to the study of conversations is like moving from physics to biology: quite different analytical procedures and methods are appropriate

— Stephen C. Levinson, *Pragmatics* (1983)

The thing about speaking is to know when *not* to speak.

It's better to keep your mouth shut and let people think you are a fool than to open it and clear all doubts.

The real art of conversation is not only to say the right thing at the right moment, but to leave unsaid the wrong thing at the tempting one.

— various

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Turn taking is the systematic process by which people coordinate their talk in a verbal interaction such as conversation. It is the set of practices and techniques whereby conversants determine *when* to speak (once they want to) and when to listen in the interchange of talk, and what happens subsequently in various circumstances. In theory, it is all to keep the conversation intelligible, within cognitive limitations, since it is difficult to understand more than one talking at the same time. But, in practice, people conform to learned socio-cultural conventions that allow them nevertheless to assume various stances when needed. The result is what we observe roughly as *turns* of one-party-speaking-at-a-time most of the time (or some of the time), recurring throughout spoken exchanges.

The relevance of verbal turn-taking for human cognition, then, should not be underestimated. As part of a person's interactional procedures, these practices and techniques are ultimately at the *control* level of his or her behaviours, whence other cognitive processes are initiated and interrupted. They help organize one of the most crucial human activities: communication. And the basic, most frequent form of verbal communication is *conversation*, fundamental to cognition and where turn-taking is most readily employed. We converse with other people to learn, exchange information, accomplish goals and tasks, or simply to reaffirm emotional and social bonds; often more than one of these at once.

With slight variations, verbal turn-taking takes place in a wide range of conversational situations: in telephone talk, casual talk, discussions, informal meetings ('coffee-table talk', smalltalk), and a few others. Therefore it is an important process to understand

and reproduce if we want to have talking agents (avatars) that can interact naturally with a person or a group to realize operations or (say) facilitate communication. Current technology is still in a stage where intermediary components of artificial talk such as speech understanding are yet to be smooth and effortless, but once these hurdles are solved, fully natural, mixed-initiative conversation between man and machines will not be accomplished appropriately without reproducing verbal turn-taking as well.

And it is in conversation that turn-taking shows itself in full and is best characterized, because that is the freest, least constrained sort of verbal interaction. Other genres of interaction such as lectures, interviews, debates and chaired meetings differ from conversation in a number of parameters: restricted turn order, restricted turn length, more formality, explicit requests to speak, and so on. But conversations are frequently intermingled with other activities in a complex mesh of behaviours, specially when used as a means to achieve other tasks. In general then, people taking turns to talk are also doing other things simultaneously.

In this light, it is in *discussions* that pure conversation emerges more easily for extended periods; and in *group discussion* it is generalized to any number of participants. Imagine a group of people sitting around a table exclusively talking about something; it does not have to be a *heated* discussion: it can be slow, disinterested talk—but just talk. That is the best ground to reproduce general turn-taking, one where it is employed and displayed continuously (albeit fashioned to a focused talk). In other sorts of interaction, turn-taking is going to be limited, either because of external constraints or auxiliar ways to coordinate talk, or because people are doing other things too, and talk is not the only (or main) focus of behaviour.

1.1 Thesis

Hitherto, verbal turn-taking has been characterized only descriptively, based on empirical or 'anedoctal' data, or in statistical experiments which draw conclusions based on sets of recorded material. As far as I know, no computational account that operationalizes the process in a generated dialogue or conversation, or even in a simpler simulation, has been realized.

This work is an attempt to bridge this gap. The thesis here is that the systematics of conversational turn-taking can be modelled in a *simulation* of group discussion that

would focus on the relevant turn-taking-related issues, abstracting away all the other complexities of speech exchange. The generated discussion would thus have to be symbolic, without contents. No real talk or speech would be produced, just 'talk' at a turn-taking, *interaction-control* level. This would indicate only *when* conversants are talking or not talking, and when they reach places of *possible speaker-transition*, where listeners could start to speak to take new turns of talk.

Essentially, the simulation would be a working demonstration of the *issues* associated with the running of turn-taking, namely: backchannel feedback ("uh huh", "huh?" and the like), simultaneous talk and its resolution, sorts of utterances as turn-taking constraints, hesitation and interruption. It would also represent a first, basic operationalization of turn-taking in a distributed multi-agent simulation. As such, it could show the kinds of interactional states and contexts that artificial conversants would have to keep in the moment-to-moment management of a discussion, such as keeping track of the speakership state and the situation of the talk.

The discussing group would be best implemented in a *multi-agent* framework, where each conversant is modelled and behaves separately from the others. It is the turn-taking systematics embodied in each agent that would coordinate the individual behaviours in an organized interaction. This systematics would translate into the *operational procedure* of each agent in this framework: what they can do at each moment of the simulation. This procedure could be the same for all agents; they would only differ in their decision parameters (their profiles) and internal states recording the various contexts they are in. Hence the *collective* turn-taking behaviours would emerge from the coordinated behaviours of the individual agents in the group.

1.1.1 Probabilistic decisions

In this symbolic discussion without contents of talk, conversants can be modelled in a simple way that dispenses planning and the whole parafernalia of speech and dialogue generation. As such, they can have probabilistic parameters indicating how much they want to talk, continue talking, and do other decisions associated with turn-taking. Simulated agents can be thus a bunch of likelihood parameters only; their decisions during discussion would be random, but biased by these probabilities. In this way, these parameters could model various conversational *profiles* such as talkative, shy, insistent, polite, and their various combinations.

Of course, random stochastic decisions with abstract 'talk' would not produce a lot of meaningful discussion, or even coherent medium-term interaction. Agents would sometimes want to talk many at once, then persist simultaneously until some or all decide to stop, depending on their parameters. They could (unrealistically) decide to speak at every possible opportunity, notwithstanding what is being 'said' or what to 'say'. They might stop and restart immediately afterwards, or not decide to speak for long periods, resulting in gaps of silence of varying lengths.

In sum, there would be not much intelligible medium to long-term coherence, since agents do not model dialogue and discourse behaviours realistically for the contents of the talk. Instead, they would be generating only the short-term turn-taking interaction centered on the *current* and *next* turns. The aim of the simulation is therefore just to reproduce these micro-level behaviours of the moment-to-moment unfolding of talk and silence, and turn-taking-related behaviours. Medium- to long-term coherence, of course, would require dialogue planning, generation, comprehension, and so forth.

There is a justification for this simplification of reality. In the best known account of turn-taking, reviewed in the next chapter, Sacks et al. (1974) observe that the major aspects of the organization of turn-taking are 'context free', shaping the ways in which the particularities of context and content are exhibited. This is (hopefully) a simulation of those major aspects of turn-taking. So, a decision of *whether* one wants to talk in a given moment may be content-dependent; but *when* it occurs is not, particularly; or at least not *only* content-dependent. It depends rather more on the *surface* of the talk that displays this content, specifically the prosodic and other paralinguistic cues subjacent to speech, which can be abstracted into just the relevant elements: talk, silence, places of decision, and so forth.

1.1.2 Small group discussion

In recreating just an abstract form of focused talk in a continuous interaction, this simulation would produce a simplified version of 'group discussion'. One that is arguably not much distant from discussions that can be seen in informal meetings or in experiments where people are told to talk about a specific subject. However, I said earlier that the group could be generalized to any number of participants. That is only partially true, because the interaction generated by this simulation would correspond more directly to the patterns created in groups of a limited number: *small* groups.

Small groups of up to six or seven people tend to interact in discussions much like two-person dialogues with overhearers (Fay et al. 2000)—more than overhearers, in fact, but *side-participants* (Schober and Clark 1989). Turns and exchanges are still tight, fast and fluid as in two-person dialogues where conversants talk directly to, and *with*, the other party. It appears therefore that the patterns of interaction and turn-taking in small groups should not be much different from those of dyads. The difference is that there are side-participants now, which can take the initiative and 'mix in' with the current speaker and its interlocutor. Hopefully then, observations made in dyads can be assumed to hold in small groups too, with some degree of confidence.

Bigger groups of more than seven people, in contrast, either break down into smaller sub-groups of discussion, or tend to interact more formally, in a sequence of monologues (Fay et al. 2000). Talk is 'broadcast' to all participants, so the discussion tends to be slower, with less speaker change, and more formal in the patterns of turn-taking. This appears to be a sudden shift as the group size increases: either participants *feel* that the group is small enough to interact informally like they do when talking to just one other person, or the group is big enough so that a more formal approach is required and tacitly employed.

Therefore, the modelling of turn-taking and the patterns generated in this simulation will be assumed to represent only *small group* discussions.

1.2 Research on turn-taking

Aspects related to turn-taking first began to be investigated by people interested in the dynamics of small groups, their talk and interaction, such as Bales (1950) and Stephan and Mishler (1952). This emphasis in the *group process* has continued in psycholinguistic studies (Dabbs and Ruback 1987) and small-group dynamics (McGrath 1984). Then came the first ideas in simulating groups in discussion (Parker 1988, Stasser and Taylor 1991) which, however, reproduced only the speaking order ('turns') of their participants. The present simulation still has some of the spirit of those works in that it manipulates probabilities as a representation of participants' decisions.

In psychology (psycholinguistics properly), researchers would study individual behaviours involved in turn-taking often without linking them to any 'organization of talk' itself. Most initial observations used to be taken from actual but limited samples

of conversation in recorded materials. These studies were statistical or observational, and the relevance of their conclusions has to be taken carefully due to their limited data—with the possible exception of works like Goodwin (1981), Oreström (1983), Kalma (1992), which involved large corpora or data sets.

It was initially with Duncan (1972, 1974), Duncan and Niederehe (1974) complemented by Wiemann and Knapp (1975) and Oreström (1983) that the various *cues* of talking were framed as signals for 'speaking turns' (turn-yielding, turn-holding, within-unit), turn-taking (turn-claiming) and backchannels. The presence of a combination of cues in certain contexts would indicate a *signal* of a certain type. Nonverbal body behaviours such as gaze, posture, body motion, arm and head gestures (*kinesics*) were investigated not only as expressive actions (pointing, shrugs, etc), but in their turn-taking guises too (Birdwhistell 1952, Kendon 1967, 1972, Argyle and Cook 1976, Rosenfeld 1977, Beattie 1981, Goodwin 1981).

The research in the more *essential* cues of turn-taking—the intersection of syntax, semantics and prosodic features, mainly intonation—, may have used restricted data at first, but has continued until the present and therefore has become considerable, e.g. Schaffer (1983), Grosjean (1983), Cutler and Pearson (1986), Ford and Thompson (1996), Selting (1996), Wennerstrom and Siegel (2003). Simultaneous talk and interruption, however, appeared to receive much less attention: Meltzer et al. (1971), Ferguson (1977), Aleguire (1978), Schegloff (2000). This whole body of research together with nonverbal studies, mainly gestures and head behaviour, has informed much current work in computational conversants, such as e.g. Novick et al. (1996), Donaldson and Cohen (1996), Cassell et al. (1999, 2001), Edlund and Nordstrand (2002).

But it is in sociolinguistics that turn-taking was first characterized as an organized system in a descriptive account (Sacks et al. 1974). Sacks and his colleagues would draw conclusions from the aggregate of cases in transcribed conversations, in the ethnomethodological tradition now known as *conversation analysis* (Psathas 1995). Their 'simplest systematics' of turn-taking was a very generic, abstract account, not compromising itself with any concrete specifics of talk and language that could later be found to be incorrect. Further work in the same area would add to the understanding of specific parts of the system, such as e.g. Schegloff (1982, 2000), Selting (2000).

It is primarily on this literature that the work of the present thesis is based. The simulation of small group discussion developed here synthesizes that descriptive account by

simulating the coordination of turn-taking through symbolic talk. From thence comes the major organization for this modelling; specific definitions of lengths, timings and other choices are for the most part sensible assumptions taken from empirical observation of real data or indicated by studies such as Bull and Aylett (1998), among others already cited.

In the course of reviewing the literature for this thesis, I studied in detail a fifteen-minute five-person discussion, in which participants were told to talk about a hypothetical situation. It served as a learning experience on the general outline of group discussions that would inform the present simulation in ways difficult to acknowledge properly. The general shape of the utterances, hesitation, and feedback as modelled in this work (chapter 4) certainly owes something to that examination.

1.3 Contributions

One contribution is to demonstrate a simple design for a *simulation of multi-party conversation* that (here) reproduces the coordination of talk in small group discussions:

1. The simulation is a *distributed* system in a simple multi-agent architecture representing the group, whose agents inter-communicate through one *blackboard* channel: a scratchboard where they put their behaviours and read the others'. The interaction of individual entities, of agents behaving as conversants, realized through the behaviours exchanged there according to a model embedded in all the agents brings about the emergent properties of the process to be simulated: verbal turn-taking. Although this whole framework is not new (it is actually the simplest and most obvious one for the job), its application for a detailed multi-party conversation probably is.

2. The system is *synchronized* through a basic round-robin mechanism, so that agents work in *cycles* (much like CPU cycles). They give the simulation a resolution (or 'granularity': finer or coarser) according to what lapse of time each cycle is intended to represent: a parameter hereinafter called *cycle-time*. This framework is transparent to, and does not interfere with, the agent modelling, which can be programmed irrespective of the underlying framework (subject to minimal adaptations that could be made transparent as well).

   Probably even a more complex simulation of conversation—with, say, verbal and

nonverbal behaviours in various 'levels', talk contents, external events, artifacts, etc—could be realized in such a simple framework. It is a matter of defining a suitably representative set of small-sized behaviours, in various modalities, that the agents would identify and react to.

A second contribution is an agent model that operationalizes the multi-party simulation to be a *working representation of many central issues of turn-taking* in small group discussion, taken from a synthesis of the descriptive literature. This model is defined by four components:

1. the cycle-sized *behaviours* that only distinguish talk from silences and listener responses, reproducing the turn-taking-related or *interaction-control* level of talk that is abstracted away from higher linguistic levels (speech and prosody, grammar and contents of the talk, speech acts and dialogue moves), which could be later added on top of this structure;

2. *attributes* that model each agent individually with a probabilistic profile for turn-taking-related decisions, that could be taken (for example) from the statistics of real recorded conversations to reproduce or combine various conversant *profiles*: shy, talkative, insistent, polite, interruptive, and so forth;

3. interactional *contextual states* living in the agents' internal memory that have some general relevance: however they are to be implemented elsewhere, these are the *sorts* of states to be represented for turn-taking entities;

4. *procedures* that make agents recognize the behaviours of other parties at each moment, decide what to do and what to return in the next moment, based on their attributes and internal states. These procedures implement the following issues associated with turn-taking in small group discussion, which are reasonable assumptions from the literature:

    (a) Simple reproduction of *backchannel* feedback, such as "uh huh" and the like, and "huh?" and the like which ask the speaker to respond to the problem of hearing or understanding that was raised.

    (b) Realization of some general practices of *simultaneous talk* and its resolution: hitches (hesitations and perturbations), stepping-up to more competitive (e.g. louder) talk to indicate a stronger stance, and repeating part of the talk that was obscured by the simultaneity.

(c) Representation of the different sorts of turn-taking *constraints* that an utterance can convey: whether it selects a specific party to talk next, or encourages anyone to talk (e.g. "Has anybody..."), or entails that there is *more* the speaker is going to say, or leaves the floor open for anyone.

(d) Simulation of *hesitations*: that is, discontinuities in the middle of talk, either silent or voiced ('umhm', 'erm'), and how they relate to turn-taking. Listeners also distinguish hesitations from *pauses* in deciding to speak, and one can appropriately hesitate before starting to speak at certain times: when selected to speak, for example.

With this operationalization of turn-taking interaction and its various associated issues, another potential contribution of this work to psycholinguists and investigators of verbal interaction is to *point out* several specific details of the process that still need inquiry, and only came to light because a simulation spelling out many (perhaps previously unconsidered) details of this process was made.

Finally, a quantitative evaluation of the various sub-models in various small group sizes gives a host of aggregate measures that could serve as references for future comparisons (further simulations or real transcribed conversations): the total times of single talk, simultaneous (conflictive) talk, silences, overlapped talk, etc; total occurrences of complete utterances, continuing (same-turn) utterances, single starts, multiple starts, middle-of-talk starts, false-starts, 'collective' false-starts (all starting and stopping), incomplete utterances, and others.

1.3.1 What the thesis is NOT about

This thesis is *not* about:

- generating the *contents* of talk: planning, goals and intentions of the participants, dialogue and dialogue moves, grounding, all types of feedback, and the discussion itself with topics, socioemotional acts, reactions, etc;

- speech and natural language processing of utterances, intonation, overlap, etc;

- reproducing aspects of group discussion other than talk: nonverbal behaviours, spatial relations between participants, object manipulation, etc.

1.4   Definition of main concepts

These are the main concepts used in this thesis: some frequently, others not. You may not agree with some (in particular 'utterance', 'pause' and 'overlap'), but, right or wrong, *this* is what they will mean here:

backchannel vocalizations (or 'utterances') in the background of the talk-in-turn that refer to it, and are responses or reactions to the 'main' talk, but do not compete with it for the floor;

false-start an incipient beginning of talk that is soon abandoned (say, in a second?), almost always because of other simultaneous talk or starts of talk; in chapter 6, I will also refer to 'collective' false-starts: when all starting speakers stop because of each other;

feedback a subset of all the possible types of backchannels, that only indicates a yes-or-no uptake to the talk (which may still be an ongoing utterance); it corresponds to what others have called 'acknowledgement' or 'request-to-repair', and I will often refer to 'backchannel feedback' to be as unambiguous as possible;

gap the silence as referred to by Sacks et al. (1974): that is, between talk of *different* parties in speaker transitions, thus delimiting different *turns* of talk; I will often refer to 'silent gap' to be as unambiguous as possible;

group discussion a subset of multi-party conversations in which *talking* is at least the main activity of the group: that is, people engaged and focused in the talk, but an informal type of talk, not necessarily in a task-oriented, topic-directed or mediated way; the simulation developed here excludes actions other than talking, like drinking, looking around, manipulating artifacts, and so forth;

hesitation a disfluency in the normal articulation of talk, which may be silent (also termed 'unfilled'), or nonsilent such as "erm", "ahmm", "I mean", "like, you know, err" (fillers); also circumscribed as 'hesitation' here are self-interruptions, self-repairs and other disruptions ('hitches') in the talk, such as a word or syllable repetition, due tipically to simultaneous talk;

interruption either smooth or unsmooth (i.e. 'polite' or 'impolite'): smooth interruptions are starts of talk at TRPs (i.e. normal turn-taking) at times when the speaker was going (intended) to continue talking, as he or she was pausing or recently

finishing an utterance; whereas an unsmooth interruption is to start in the middle of someone's talk (a middle-start)—either in what one thought was a TRP or not—and insisting so as to make the current speaker quit his or her utterance or course of talk;

lapse an occasional lull in conversation when nobody is talking, as mentioned by Sacks et al. (1974): a silence that gets longer when the recently finished speaker does not resume talking, and nobody else does too; in case the same speaker restarts, it may be another 'turn' of talk by the same party (considering the intentionality in 'reengaging' the speakership), or just an extension of the same turn (in which case the lapse was an abnormally long, unintentional pause);

latch a speaker transition leaving no silent gap or overlap in the talk: that is, the starting speaker's utterance begins right after the previous one has finished;

listener the temporary non-speaking role in conversation—even if one is distracted, thinking away, and not truly *listening* to the talk—, which is elsewhere called 'auditor' or 'hearer'; in a group, a listener may either be an *addressee* of the current talk or just its *hearer*, while in two-person conversation the non-speaking party is (supposedly) always the addressee of the talk;

middle-start an attempt to talk in the middle of someone's talk, in a (supposed) mid-utterance TRP: a possible utterance completion in the middle of the current talk that was *recognised* as such by the starter (or the interrupter, if the intention was really to interrupt the speaker);

overlap here, only the overlapped talk in speaker transitions, when someone starts an utterance slightly before the current speaker finishes talking (and does not restart afterwards); this simultaneity is non-conflictive and in fact often not even *perceived* or cared about by the parties; other simultaneous talk such as multiple starts of talk or middle-starts (that is, non-backchannel talk intending for the same floor of attention), will be referred simply as *simultaneous* talk;

pause the silent interval between *same-speaker* utterances, extending a turn of talk; it differs from a silent hesitation when the just-finished talk is syntactically and prosodically complete (a finished utterance)—although it may not be 'discoursively' complete, in which case there is *more* the speaker is going to say (§2.4); I will often refer to it as 'TRP pause' to be as unambiguous as possible;

speaker the party (participant in the group) that generally holds the floor of attention, the 'turn' to talk, but may not necessarily be *speaking* in a given moment: he or she may be pausing or hesitating;

TRP the transition-relevance place as introduced by Sacks et al. (1974)—also referred to as a 'juncture' by others—, when one can start to speak to (smoothly) take a new turn of talk; it can either be a pause or lull (lapse) in the conversation, or a possible utterance completion in the current ongoing talk;

turn loosely equated to *floor* in one-at-a-time talk, and possibly formed of one or more utterances; Edelsky (1981) also called attention to *collectively-developed* floors, when more-than-one talk not only happens but is regarded as unproblematic by the parties, in no need of repetition, repair, or otherwise 'correction';

utterance the 'unit' of talk just as the sentence is the grammatical unit; an utterance can actually be just a growl, a word, a phrase, a clause (elliptic or not), or one or more sentences spoken together without any pause (which is what actually separates same-speaker utterances), but it may also be fragmented by any instances of hesitation; so, an utterance can be formed of one or more TCUs, the abstract turn-construction units of Sacks et al. (1974);

verbal interaction also talk-in-interaction, or what Sacks et al. (1974) termed speech-exchange system; its various genres would include types of conversation (casual, chat, smalltalk, discussions and meetings), interviews and inquiries, press conferences and question-and-answer sessions, lectures and tutorials, debates and panels (mediated or not), legal trials, ceremonies like the mass, and others.

1.5 Structure of the thesis

The thesis is divided in the following chapters:

1. this introduction: overview, literature overview, contributions, concepts;

2. review of the literature relating to turn-taking: the turn-taking systematics, criticisms to it, backchannel feedback, multi-utterance turns ('more-to-come'), simultaneous talk practices and resolution, and two other approaches to turn-taking that identify its cues and the collaborative reasons for interaction;

3. the framework of the simulation: description of the simple multi-agent system and blackboard architecture that runs in cycles of a parameterizable simulated time, and discussion of other aspects of group discussion;

4. the model of turn-taking, with description of its four components: behaviours, attributes, interactional states, and procedures;

5. examples of the simulation with different agent profiles and in the various extensions (with simultaneous talk procedures, sorts of TRPs, and hesitation);

6. a quantitative evaluation of the model and extensions, describing its design, the measures that were counted and which form a 'profile' of the generated discussions, and the examination of the various results in a range of parameters;

7. conclusion and discussion about possibilities for future work: an assynchronous parallel simulation with variable attention, the cues of turn-taking (syntax and prosody), nonverbal behaviours, 'fluctuating' attributes, and the possibility of integrating a system with dialogue and speech generation for audible evaluation of the model.

Readers who may wish to have a quick idea of what the simulation is about could head directly to chapter 5 to look at the examples and description of what behaviours the simulation is supposed to reproduce. The rest of the thesis is probably best read in order. Except that the turn-taking model which is the core of the thesis (chapter 4), can probably be understood without reading the previous chapters at full: the model is independent of (and does not make reference to) the framework of the simulation in chapter 3; and what is described therein is sufficiently intuitive with only a brief knowledge of the conversation analysis tradition of turn-taking, reviewed in chapter 2.

Throughout the text I refer to other sections or subsections with the paragraph symbol (§) as a shorthand (e.g. §2.1.2). The only other eccentricity to bear with me—besides the probable awkward style and rather direct tone, from *me* to *you*—is my taste for quaint or downright arcaic compound English adverbs (like 'thereupon').

The review chapter (chapter 2) is the only one that presents views, ideas and previous work that are not mine (except if indicated otherwise with 'my' or 'I'). Any simplification or misrepresentation in their exposition, however, is entirely my fault.

Chapter 2

Turn taking

One characteristic of group discussion and in fact many forms of verbal interaction is that participants in general take *turns* to talk. An organization of turn-taking seems to be fundamental to most joint interchanges of speech, and many other socially organized activities as well, such as: moves in games, customer attendance, traffic at intersections or through road narrowings, and others. Goffman (1964, p.135–136), as quoted by Sacks et al. (1974), characterised those activities as a "social organization of shared current orientation, [involving] an organized interplay of acts of some kind". In the case of verbal interactions, of turns at talking, about which he wrote:

> Talk is socially organized, not merely in terms of who speaks to whom in what language, but as a little system of mutually ratified and ritually governed face-to-face action, a social encounter. Once a state of talk has been ratified, cues must be available for requesting the floor and giving it up, and for informing the speaker as to the stability of the focus of attention he is receiving. Intimate collaboration must be sustained to ensure that one turn at talking neither overlaps the previous one too much, nor wants for inoffensive conversational supply, for someone's turn must always and exclusively be in progress.

Such an organization of verbal turn-taking will be described in section 2.1, henceforth called the *turn-taking systematics*. Section 2.2 will present some criticisms that followed this account, going mainly around the idea that one-at-a-time cannot be an enforcive, prescriptive system, but rather a common pattern. Section 2.3 shall summarize the concept of *backchannels*: listener responses like "uh huh" or "huh?" (and others) in the background of the talk-in-turn that inform the speaker "as to the stability of the focus of attention he is receiving". Section 2.4 will show how talk can also

indicate 'more-to-come' and thus constrain the subsequent turn-taking to guarantee an extended multi-utterance turn. And completing a review of the elements that informed the group discussion modelling of this work, section 2.5 will present the typical ways by which speakers deal with simultaneous talk: how they generally begin, persist, desist and resolve multiple attempts at talking.

In addition to this combined account of turn-taking focusing on the *structure* of turns more or less independently of their contents, two other complementary approaches are reviewed. Section 2.6 summarizes the classic psycholinguistic studies on the verbal and visual *cues* that encourage or discourage turn-taking. And section 2.7 presents the theory of *collaborative* dialogue focusing on the conversants' obligations in *grounding* their talk, which indirectly determines turn-taking, and has influenced much recent work on dialogue systems.

The differences among these are that, while the structural account developed in this thesis is devoted to characterize *what happens* once there is turn-taking, the psycholinguistic studies are focused on *when* it happens and what can trigger it, whereas the collaborative theory tries to establish *why* it happens: what leads to turn-taking, to the speaker changes. Only the structural approach is the subject of this thesis, and will be introduced next.

2.1 The turn-taking systematics

The best known account of verbal turn-taking was presented by Sacks, Schegloff, and Jefferson (1974) for *conversation*, revised slightly in Schenkein (1978). From a sociological perspective, the authors regarded turn-taking as "a prominent type of social organization, one whose instances are implicated in a wide range of other activities" (p.696)[1]. Conversation is seen as in one extreme of the range of *speech-exchange systems*, the genres of verbal interaction. It is the one with freest turn allocation, where participants can freely talk and take turns.

Other genres, such as interviews, meetings, debates, ceremonies, trials, conferences, lectures, etc, differ from conversation in various constraints on the turn-taking organization. Hence, "turn-taking systems can be workably built in various ways [s]ince

---

[1]Unadorned page numbers in this section will refer to Sacks et al. (1974). The whole section *is* a summary of that work (unless when referred otherwise), so there is some quoting and light paraphrasing from it in the following subsections.

they are used to organize sorts of activities that are quite different from one another"
(p.696).

The authors had for several years used audio recordings of naturally occurring con-
versations increasingly to characterize and describe the various types of sequential or-
ganization present in conversation, a methodology later termed *conversation analysis*.
The data was transcribed paying attention to the timing of pauses and interruptions,
though intonation was not annotated in detail, only suggested to some extent by punc-
tuation. The various points the authors make are illustrated by excerpts from these
transcriptions; no statistics is provided, so we have but to trust that the examples given
are typical (Power and Martello 1986).

This data made "the existence of organized turn-taking (. . .) increasingly plain. It
has become obvious that, overwhelmingly, one party talks at a time, though speakers
change, and though the size of turns and ordering of turns vary; that transitions are
finely coordinated; that techniques are used for allocating turns (. . .); and that there
are techniques for the construction of utterances relevant to their turn status, which
bear on the coordination of transfer and the allocation of speakership. (. . .) Focusing
on facts such as these, rather than on particular outcomes in particular settings, leads to
an investigation of the organization of turn-taking per se, rather than to its application
and consequences in particular contexts, although the more formal understanding of
turn-taking illuminates more particular findings" (p.699).

Thence an account of turn-taking was developed to be independent of parameters of
context—circumstances, topics of talk, the identities of participants in conversation—,
but capable of great context-sensitivity. That is so because conversation can accomo-
date a wide range of situations, interactions and changes amongst parties with any
potential identities and familiarity. So, the authors stated, an account should fit this
variability by design, yet in a manner that, requiring no reference to a particular of
context, would still capture the most general properties of conversation, in a *simplest
systematics*. Those properties are the "grossly apparent facts" about conversation sum-
marized as following (p.700–701):[2]

---

[2]Oreström added that conversations are casual, informal and spontaneous, not scripted or premedi-
tated, with a freedom to talk and to introduce new topics; they have backchannels, tag questions, inti-
macy signals ('you know', 'you see'), and frequent discontinuities such as hesitations, repetitions and
incomplete utterances. "In brief, conversation may be characterized as an informal speech event largely
guided by the spontaneousness and interests of the participants and may occur for no other reason than
social interaction. A debate, on the other hand, is a formal speech event, highly task-oriented and
organizationally efficient" (Oreström 1983, p.23).

- speaker-change recurs, or at least occurs;

- overwhelmingly, one party talks at a time;

- occurrences of more than one speaker at a time are common, but brief;

- transitions with no silent gap and no overlap between turns at talk are common, and together with those characterized by a slight gap or a slight overlap constitute the vast majority of transitions;

- turn order and turn size vary, being determined not in advance but locally, one-at-a-time;

- the length of conversation, the relative distribution of turns, and what parties say within them are also not specified in advance;

- number of parties can vary, even within the same conversation;

- talk can be continuous or discontinuous, within and between turns, because of hesitations, lapses in conversation, etc;

- techniques exist both for allocating turns and for the construction of utterances within turns;

- repair mechanisms exist for dealing with turn-taking errors and violations: e.g. if two parties find themselves talking at the same time, one of them generally stops speaking.

The *turn* is tacitly assumed to be—for it is never defined directly—the speaking space of one party up to the point when another takes over and the former has stopped. It is unclear whether overlapped utterances (or parts thereof) in speaker transitions are part of which 'turns'; this concept becomes less clear as simultaneous talk gets longer or frequent. Intervals of silence are supposedly part of a turn if the *same* speaker continues afterwards, making them *pauses* in the talk. But it is not clear when a (short) silent gap becomes a (long) *lapse* of silence, and whether subsequent talk by the *same* speaker is then taken as another 'turn' or not. As Power and Martello (1986) point out, the word *turn* is employed in two senses: as the right to speak or, loosely, the *floor* in "turn-allocation", and as the talk and pauses produced by the rightful speaker, as in "turn-construction".

Turn-taking is then the systematic realization of those units, accomodating the parties' interests and purposes. The serial character and local scope of the organization are thus

emphasized: "The turn-taking system is, in the first instance, a system for 'sequences of talk'" (p.710), for "it is built to organize but two turns at a time, current and next, and the transition from one to the other, without restriction on the number of such currents and nexts it can serially organize, so also it organizes but two speakers at a time" (p.712). Therefore, (what has been somewhat controversial) "the system allocates *single* turns to *single* speakers; any speaker gets, with the turn, exclusive rights to talk" (p.706, emphases from Schenkein (1978, p.15)).

The systematics is described in terms of two 'components', turn-construction and turn-allocation, and a set of ordered rules, later termed *practices* (Schegloff 2000) as 'rules' proved to be a source of misunderstanding. Turn-allocation can be further sub-divided into two techniques for allocating the next turn in sequence: current speaker selects the next to talk, and self-selection. The following subsections will address each of these parts, in turn (of course!).

2.1.1 Turn construction

In conversation, as the *length* of what is going to be said is not fixed in advance, the allocation of turn space is done by talk *in* the turn. In contrast, ceremonies may have speaking turns (of proclamations or announcements) fixed by tradition, in either the exact words to be spoken or in a length of time; the same for 'calling the register' in classroom. Even in some conventionalized exchanges in conversation itself, like greetings, one may allow just enough time for the other's expected response of "hello" or the like before continuing to talk. In such cases, a turn is not created by talk itself but allocated beforehand by one speaker, whether it is filled with talk or not.

The talk that *constructs* a turn is composed of instances from the various unit-types of a language that can be thus usable. In English, they comprise sentences, clauses, phrases and lexical items (such as 'thanks' and 'yes'). At the start of a turn, the speaker is initially entitled one such unit called a *turn-construction unit* (TCU), whatever the type it turns out to be. "The first possible completion of a first such unit constitutes an initial transition-relevance place. Transfer of speakership is coordinated by reference to such transition-relevance places, which any unit-type instance will reach" (p.703).

The *transition-relevance place* (TRP) is possibly the most visible concept of the sys-

tematics. It corresponds to the places[3] of *potential* speaker transition, recurring discretely in the course of a turn at the *possible completion* of turn-construction units. Therefore it is a most crucial feature of TCUs that they "project, from their beginnings, features of their construction, their direction, and what it will take to complete them" (p.719).

That is, TCUs can be identified while in their making, and suggest to attentive listeners their possible completions before their occurrence. Sentence beginnings, for instance, are capable of being analysed in the course of their production to project their possible directions and completion points. "In the course of its construction, any sentential unit will rapidly (in conversation) reveal projectable directions and conclusions, which its further course can modify, but will further define" (p.709). An initial *wh*-word, for example, powerfully constrains further development of an utterance to a 'question' type, with respective restriction of its further possibilities.

This *projectability* is a key aspect of the turn-taking systematics, and "will be compatible with a system of units which has this feature" (p.720). It explains the fine timing and coordination that are evident in many speaker transitions, with appropriate starts after turns composed of single-word, single-phrase, or single-clause constructions, without any gap: that is, without any waiting for a possible sentence completion. It would also explain the multiple simultaneous starts without gap occurring at some transitions, that testify to the independent but nearly identical projection of the TRPs.

Listeners have the capacity to start with precision in relation to the ongoing talk, selecting a place to speak so that their utterances sound as a continuation of the previous one (Jefferson 1973). Starting to speak so as to appear a continuation of prior talk, leaving no silent gap or overlapped talk, has been called *latching*; the speaker transition without any gap or overlap is then called a *latch*.

However, this projection is not always precise. "Variation in the articulation of the projected last part of a projectably last component of a turn's talk" (p.707) means that overlaps can occur, and they are "common but brief". According to Oreström (1983), there may be other reasons for simultaneous talk: the TRP may be misspotted, or someone objecting to what the speaker says may attempt to 'shut him down' by talking over, or one may try to retake the turn 'by the same measure' it was taken earlier; also, parties tend to 'clip redundancy' (things already surmised or going to

---

[3]Either points or intervals since both notions are possible: the TRP is extendable.

be repeated) wherewith the presumed or predicted redundant talk is overlapped; and finally, simultaneities also occur by sheer eagerness, by someone wanting to talk at *that* point in the conversation ('it is now or never').

Furthermore, although TCUs have been described in terms of syntactic categories, clearly aspects of speech production such as intonation, integrated to the semantic and pragmatic interpretation of the utterance, are essential in the projection of possible completions (Ford and Thompson 1996, Traum and Heeman 1996, Wennerstrom and Siegel 2003). It is what Kendon (1970) called the *running hypothesis*: the partial interpretation of the ongoing utterance; which by the way is never judged complete while a tone unit is in progress (Power and Martello 1986). Other related cues (§2.6) that may indicate utterance completion are the lowering of pitch and loudness, drawl in certain expressions like "you know", relaxation of body parts, and gazing back at an interlocutor (Kendon 1967, Duncan 1972, Beattie 1981).

2.1.2 Turn allocation: current speaker selects next

Four ways of *selecting* someone to speak next were described by Sacks et al. (1974):

1. The basic technique ("perhaps the central one") involves addressing a specific party whilst producing a *first pair-part* of one of various sorts of dialogue exchanges that were termed *adjacency pairs* (Schegloff and Sacks 1973): for example, question-answer, proposal-evaluation, offer-acceptance (or declining), greeting-greeting, farewell-farewell, etc. First pair-parts impart obligations on a cooperative party, and in so doing, make it relevant for its response to be in the next turn of talk. They set constraints on what should be done next: e.g. a question makes its answer the appropriate response. But they do not *by themselves* allocate the next turn to someone; it is by *addressing* someone that the next speaker is in fact selected, to do whatever the first-pair part requires. The two basic ways of addressing are:

   - gaze the addressee in face-to-face interaction (sometimes when underscoring second-person references in the utterance like 'you', 'your', 'sir');

   - attach a vocative, generally at the beginning or end: e.g. "John, is this right?", "It is up to you, mister", "Tell me, your honour, is it possible?". A vocative can be a first part by itself in summoning, e.g.: "chief!".

2. Addressed tag questions (e.g., in English, "you know?", "aren't you?") can also be attached to an utterance, whether it was constructed to be a first pair-part or not, turning it into one that selects a next speaker. By being the generally available 'exit technique' of a turn, they can be used to explicitly *exit* a turn that did not (clearly) select a next speaker. Such as when at the end of a turn nobody speaks: the recently finished speaker can then signal with a tag question that he or she *has* indeed finished. Tag questions thus provide a major source of indicating 'talk done' when the other parties do not seem intent on talking.

3. An elliptic, reduced question that follows or interrupts a turn is interpreted by reference to that turn's talk, thereby automatically addressing its speaker without any other affiliation needed. This is a variant on the use of a first pair-part, and "the only systematic mechanism available for next-speaker selection which can prefer, formally, a next speaker identified only in turn-taking terms (and thus context-free terms)" (p.717–718). There are at least two kinds of reduced questions that select the prior speaker (in English and other languages):

   - confirmations that echo part of the previous turn with a 'question' (i.e. rising) intonation: for instance, either "Anna?", "today?", or "*married* her?" could follow the utterance of "John married Anna today";

   - questions of interrogative pronouns: "where", "how much?", "who", "who me?", "whereabouts?", etc.

4. *Social identities* can also make someone immediately selectable without an explicit addressing. Sacks et al. (1974) give the example of two couples in conversation, so if someone says "You should go to the movies with us" there is no doubt as to who 'you' and 'us' refer, and consequently who is selected to speak.

In addition to these social identities, there may be ways that can address and select a next-speaker based on purely pragmatic reasons, in constrast to the use of explicit devices like vocatives and tag questions. Recognition of these associations may involve anything from situational or local knowledge, world or common-sense knowledge, the participants' goals and their mutual knowledge about it.

Indirect addressing, for example, may be one such practice not mentioned by Sacks et al. (1974): when a party feels compelled to speak by way of being talked about in the third person. For example: "John was there, he knows about it", where the person so referred is in the group. This may be accompanied by gaze, but needs not,

specially when teasing someone. Finally, a formal and explicit way of selecting the next speaker—such as "Let's hear what John has to say now."—can also occasionally take place in conversations, though it is perhaps a device of meetings, highly task- or topic-oriented, or of other more formal encounters than conversation.

2.1.3   Turn allocation: self-selection

The basic technique anyone has for selecting oneself to a turn at talk is just to 'start first': start quickly at a transition-relevance place so as to be the earliest, single next speaker. That is because, regularly, "first starter gets the turn", stated explicitly as one of the turn-taking rules in the next subsection. The motivation is that, if one had not started, and started fast, someone else would have. This emphasizes, most of all, that just a *single* speaker regularly starts and takes the next turn of talk, not that the first who usually starts amongst many 'gets' the turn.

This design creates a pressure at many TRPs which encourages turn-size minimization. At one side, parties are motivated to take the next earliest opportunity if they want to talk, otherwise they risk losing the opportunity in that context of conversation, to which their intended talk may be destined ('now or never'). At the other side, the speaker tries to "construct a turn's talk as to allow its intact formation in the fact of this pressure" (p.719). The result is that TRPs are often the main locus of overlapped talk, not only because self-selectors may misjudge the possible completion of the utterance, but because there may be 'post-completers' like tag-questions and redundant 'finishing talk', or articulatory extensions and variations in the last part of talk that lead up to the (perceived) possible completion. Should a self-selection occur too early, its beginning of talk would be overlapped.

In consequence, Sacks et al. (1974) observe that the need to speak as early as possible in a TRP is constrained by contingencies in planning and colocation of the turn's beginning, which, given its projectability, will have to reflect some degree of planning for the turn's talk. Given that turn beginnings are subject to these "multiple sources of overlap", careful timing is required, they say, for an overlap will impair the analyzability and impact of the utterance, particularly if it is a sentence.

In this regard, a class of constructions is of particular interest: appositional beginnings like "Well", "But", "And", "So", etc, which are "extraordinarily common" according

to Sacks et al. (1974). They do not need any planning by virtue of not being part of the turn's content; for the same reason their overlap would not impair its analyzability. So their use in beginning a turn's talk can be understood as a sort of self-selection technique. It then turns out that the basis for the use of these appositionals and tag questions, not evident linguistically, is that they have important turn-taking and turn-organizational functions.

While first starters are generally the *only* starters, multiple starts of talk at TRPs also do occur. Apart from simultaneous starts by self-selectors aiming to be the earliest single next-speaker, there are many instances of multiple starts where one clearly started first, and who subsequently did (or did not) take the turn. With no distinguishable first starter, it appears that the *loudest* usually ends up with the turn (Meltzer et al. 1971). When someone was clearly the first to speak, it is expected that subsequent starters realize their monitoring lapse and stop forthwith, giving way to the earlier starter.

But there are also cases in which, even though someone clearly started, a subsequent (closely-following) speaker was attended, or continued to speak nonetheless. There are then techniques or situations in which *second-starters* end up with the turn. The provision of 'first to get the turn' operates without respect to the type of utterance, independently of what the new beginning of turn may seem to be. In contrast: "Second starter techniques, and their efficacy in superseding the operation of the first-starter provision, are contingent on the type of utterance they can, from their starts, reveal themselves to be" (p.720). One case mentioned (which then seems to be the 'main case') is when a problem of understanding arises, since "addressing of problems of understanding in this way is a priority activity in conversation" (p.720). Probably because such a problem needs to be resolved in that context of conversation, whilst other talk that advances the discussion can wait.

Notwithstanding these considerations, individual decisions to talk at TRPs or to stop talking in simultaneous talk are related with eagerness to make a contribution and to the degree of involvement in a conversation, that affect the priority given to one's own turn over the others (Oreström 1983). Emotions often play their part in suppressing polite restraints: "a clash of opinions also means a clash of turn-taking" (Oreström 1983, p.159). All these may lead to more overlap and persistent simultaneous talking. Finally, participants tend to behave in terms of politeness according to the relative status, liking and acquaintance (or lack thereof) to each other.

2.1.4 Turn-taking rules

The following ordered rules (or 'practices') govern turn construction and provide the allocation of the next turn to one party, "coordinating transfer so as to minimize gap and overlap" (p.704); that is, localizing gap and overlap at TRPs and their immediate environment. The rules are ordered because there is a *priority* on their application: first rule first if possible, otherwise the second, or else, the third.

In any turn, at each TRP (identified, as we recall, as a possible completion of a turn-construction unit):[4]

1. if the turn-so-far was constructed to involve a 'current speaker selects next' technique, then "the party so selected has the right and is obliged to take next turn to speak; no others have such rights or obligations, and transfer occurs at that place" (p.704).

   If the turn-so-far does *not* involve a select-next technique, then two other provisions are possible (p.704):

2. "self-selection for next speakership may, but need not, be instituted; first starter acquires rights to a turn, and transfer occurs at that place.";

3. "current speaker may, but need not, continue, unless another self-selects".

Should the current speaker *not* continue to talk, rule 1 remains not in operation, and there is "further space (another round)" (p.715) available for self-selection. This appears to conform with (or result from) the observation that silent gaps in speaker transitions are shorter "on the average" than silence followed by continuation of the same speaker (Oreström 1983). As there is a pressure to speak at the earliest opportunity (if one wants to talk), there is pressure to minimize those gaps between turns, whereas when no one is intent on talking, the current speaker either takes longer to recognise it, or can more calmly pause before continuing.

Also, Wennerstrom and Siegel (2003) concluded in a study of syntax and intonation at places of speaker transition that the *probability* of turn-taking is highest at the onset of silence after utterance completion, falling slightly within the first 0.5s then increasing steadfastly again. This suggests that there is an optimal period during which an interlocutor is more likely to *latch* onto another's turn immediately, followed by a second

---

[4]The original numeration was 1a, 1b, 1c, with a rule 2 stating that the whole system should repeat.

period when, in the absence of a latch, the current speaker is more likely to continue. And thereafter the probability of a speaker change rises again as the pause gets longer, when it then becomes apparent that the current speaker is not going to continue.

Clearly, the options that the rules provide are not only constrained downwards by the explicit ordering of their formulation. They are also restrained upwards too, by the pressure the lower-priority rules exert on the higher-priority ones just by virtue of their presence in the set—regardless of whether they are actually used. So the option of selecting a next speaker can only be exercised as long as nobody has so far self-selected to speak, and this self-selection can only occur as long as the current speaker has not continued from a TRP which had no other self-selection. If the speaker intends to select someone to talk next, he or she is under constraint to accomplish the selection before first possible completion, lest an undesired other self-selects to speak there.

The disposition of rules 2 and 3, however, seems to suggest that it cannot possibly happen that both the prior speaker and some other start to speak simultaneously, or nearly so. Then it is not clear what should happen when that in fact happens. The rules and their ordering apparently imply that the current speaker must always give way to another starting speaker, but empirically this is not always true: there *are* cases where even when someone starts slightly before, the prior speaker resumes talking, insists and continues to speak, retaining the turn.

The lack of a *turn* definition wants two clarifications, given that the rules apply 'for any turn' and 'the turn-so-far'. First, the application of rules 1 or 2 marks the end of the current turn and the beginning of the next, though boundaries are not marked precisely, only 'transfer' is mentioned. Second, the use of rule 3 does not count as another 'turn' by the same speaker, but an expansion of the current one: "the system permits the use of that option to be treated as a within-turn event, counting not as an instance of a turn-allocation to a same speaker, but as an increment to turn size" (p.711).

The rule-set option cycle characterizes the turn-taking system for the most part as a *local management*, *party-administered* and *interactionally-controlled* system. That is, turn order and size together (the two features with which the system directly concerns itself) are determined one-by-one in an orientation of the *current* turn to the *next* one (locally managed); the organization and distribution of turn order and size is subjected to the parties of the conversation (party-administered); and the opposite contingencies of the rule hierarchy further make this intertwinned organization oriented to the con-

tribution, or lack, of other parties (interactionally-controlled). Turn size, as a case in hand, *is* multilaterally determined since a current speaker can expand his or her turn only so long as no other self-selects to speak at a TRP, and have his or her (intended) talk cut short by that very self-selection. "But however this particular model may be defective, (...) the appropriate model for turn-taking in conversation will be this *sort* of model" (p.725).

2.2 Criticisms

The presentation of the turn-taking systematics above has already incorporated some critical issues that have been pointed out, particularly by Power and Martello (1986). A few other criticisms are summarized in this section. The intent in their exposition is not to disqualify or invalidate the turn-taking systematics, but to shed light on under-specifications or controversial points which were (or were not) since then addressed by later work. In addition, some of the remaining sections of this chapter can be seen as adding up elements to that understanding.

Edelsky (1981) argued that one-at-a-time is not a conversational universal nor essential for communication. Instances of more-than-one at a time are certainly not always brief, repaired or degenerate. Earlier, Spelke et al. (1976), as mentioned by O'Connell et al. (1990), had showed in a laboratory study of language processing that it is not necessarily true that people can only process messages from one source at a time. Some instances of multiple talk *can* be grasped, or their basic gist, provided the utterances are not long or complex. There is even at least one speech community where naturally occurring simultaneous talk is frequent, expected, and processed (Reisman 1977).

Another criticism to the turn-taking systematics is that it supposedly is, or is intended to be, a convention prescribing 'rules' (i.e. one-at-a-time) to which conversants must adhere (procedurally perhaps), instead of being a general characterization of turn succession in conversation. The authors

> "assume too easily that a regularity in social behaviour must be due to a social convention specifically prescribing that regularity. Observing that the first person to speak up is usually allowed to continue, or that addressed questions are usually immediately answered by the addressee, they assume without discussion that these regularities are due to specific turn-taking conventions (...). In making this criticism we are assuming, of course,

> that the rules given by SSJ are supposed to represent a cultural convention
> and not merely a statistical regularity." (Power and Martello 1986, p.39)

Incidentally, the use of the word *rule* makes it seem as something to be followed. They are 'general rules' in a *descriptive*, not *prescriptive* way. Searle (1992) argues that the patterns thereby identified have no *causal* explanation, insofar as the *identification* of the patterns themselves does not explain anything. He makes slightly captious analogies with a 'rule for walking' (because it is tautological) and driving on the left or right of the road in different countries, wanting explanatory rules much *like* those of (his and others') pragmatic theories of speech-acts—a philosophical preoccupation.[5]

Here is *my* (slightly captious) analogy: the turn-taking systematics would be like a description of the morphology, or behaviour, of species in biology, showing what characteristics distinguish their individuals—something useful in and of itself. But then come people saying "that is no use anymore: we want to know *why* they are that way, what led them to be like that". In the case of the patterns or moves of conversational dialogue that realize turn-taking, possibly the only *sort* of explanation may be along the lines of Clark and Schaefer (1989), summarized in section 2.7.

In retrospect, it is understandable that the description in Sacks et al. (1974) may lead to this. Its terminology—apparently a tradition in Conversation Analysis (Psathas 1995)—, the use of 'components' and 'rules' qualified with words like 'machinery' and 'apparatus', seems to emphasize a sort of mental mechanics whereupon conversants would *follow* a one-at-a-time talk in spite of their goals and intentions. And consequently that simultaneous talk must needs be 'repaired' as soon as possible because it is an 'error and violation' of the one-at-a-time systematics.

Another criticism—rather more an open issue—is the question of cross-cultural validity. Whether the turn-taking systematics is 'valid' across cultures or even across different ages and social groups in the same culture, or how its details change in the different social groups, is a question to be settled with further studies. As one case in hand, the frequence and acceptance of *interruptions* is dependent on social norms and are thus but one variable reported to vary considerably in different cultures. For hindus, for example (and possibly other asian groups), interruption is not only common but *expected* as an indicative of cooperation and attention (Ervin-Tripp 1979). In

---

[5]"For when a good philosopher challenges a platitude, it usually turns out that the platitude was essentially right; but the philosopher has noticed trouble that one who did not think twice could not have met." (Lewis 1969, p.1).

western cultures, one extreme (anedoctal) example is given by Eco (1986), quoted by O'Connell et al. (1990):

> Italians interrupt one another. Everybody gets all excited and tries to make his views prevail by preventing the other from speaking (...). Americans speak in turns. (It is no accident that the pragmatic theory of 'conversation turns' originated in the United States. Italian researchers who write articles about this matter treat it as an excavation from Mars.)

The described 'italian' behaviour seems pretty much similar to those of many family gatherings such as in Christmas or New Year, probably all around the world. Nonetheless, Sacks et al. (1974, p.700, note 10) have remarked:

> We can report the validity of our assertions for the materials we have examined, and apparently for Thai materials examined by Moerman, New Guinea creole materials examined by G. Sankoff (personal communication), and for an undetermined number of languages (...). Furthermore, examination of cross-cultural conversation, where parties do not share a language of competence but a lingua franca in which all are only barely competent, is consistent with what follows (...).

There has been a substantial and ever growing number of studies of turn-taking-related aspects in a number of languages since then, in particular Japanese e.g. (Hayashi and Iwasaki 1998, Tanaka 1999, Ward and Tsukahara 2000, Tanaka 2001, Furo 2001), but also Thai (Iwasaki and Horie 1998), Korean (Kim 1999), German (Selting 1996, 2000), Swedish (Carlson et al. 2005), Dutch (Caspers 1998), Caribbean Creole (Sidnell 2001), and Mandarin (Clancy et al. 1996). The latter, and other tone languages in which intonation has lexical instead of prosodic function, probably use different mechanisms of projecting TRPs.

Lastly, two other criticisms are discussed at length in the following subsections.

2.2.1   Content and purpose

While utterances are formulated according to the content and context of the dialogue, they are realized in conversation within a system of turn exchange, and therefore, are within the constraints of this system. In its most abstract level, the system is independent of what fills the turns (Sacks et al. 1974, note 8):

> What we mean to note is that major aspects of the organization of turn-taking are insensitive to such parameters of context, and are, in that sense, 'context-free'; but it remains the case that examination of any particular

materials will display the context-free resources of the turn-taking sys-
tem to be employed, disposed in ways fitted to particulars of context. It is
the context-free structure which defines how and where context-sensitivity
can be displayed; the particularities of context are exhibited in systemati-
cally organized ways and places, and those are shaped by the context-free
organization.

There are constraints for what may be said in any one turn, but they are determined by
other *systems* orthogonal to the turn-taking system, that thus operates independently
of the various meanings displayed and put to use in the turns. However, this charac-
terization is rejected by researchers who argue that speakership organization cannot be
dissociated from considerations of content and purpose (O'Connell et al. 1990). The
turn-taking systematics is therefore seen as prescriptive in its 'idealistic aim' for "the
smooth interchange of speaking turns".

According to Power and Martello (1986, p.37), "the data on which the theory is based
can be explained by general principles of rationality and cooperativeness, without in-
voking turn-taking conventions at all. (. . .) once an utterance is under way it is not
usually interrupted without special reason; other participants who were planning to
speak usually withdraw. The general principle (. . .) is, we suggest, that one should
avoid wasted effort (. . .) [which] would be not just inefficient but also inconsiderate.
(. . .) The advantage of this line of explanation, apart from economy, is that it also
accounts for those cases in which interruption is acceptable. If the speaker's utterance
is unnecessary, or ineffective, or not directed to the goal of highest priority, principles
of rationality and cooperativeness may warrant or indeed oblige interruption."

Also, the turn-taking systematics "neglects the variety of purposes interlocutors have in
listening or speaking and the corresponding variety of turn-taking forms" (O'Connell
et al. 1990, p.346). For example, old fellows at a bar might have long intervening
pauses in between talk. In other situations, conversants might choose to harangue one
another for a time, or interrupt one another frequently; or they may chime in together
(chorally so to speak) in specific circumstances (O'Connell et al. 1990). Simultane-
ous speech, in particular, has a supportive role that is underplayed by a one-at-a-time
prescription (Coates 1989). These are not necessarily 'breakdowns' of the turn-taking
system, to which repair is necessary. Instead, they are deliberate ploys that carry mean-
ing on themselves, in the various ways turn-taking is effected (O'Connell et al. 1990).

Finally, again according to O'Connell et al. (1990), conversation is not a homogeneous
domain: different conversational *styles* cannot be considered a single speech-exchange

system. The interaction can be expected to differ systematically depending on intimate or formal relations, between strangers or friends, with complex or simple topics, urgent or not, and so on. It varies in accordance with all the finality contributed to variously by the designs and intentions of the participants, whose goals may be as diverse as information, deception, seduction, or the mere relishing of mutual presence. Therefore, the parameters relevant to turn-taking should be: politeness and cultural norms, probabilistic speaker and hearer cues, expectations, motivations, purposes and situational exigencies (O'Connell et al. 1990). That seems to be, in fact, the traditional psycholinguistic view.

2.2.2 Syntactic characterization

Until the middle to late 1970's, the variables studied in conversation were rarely above the level of syntax. Only up to the end of the 70's and 80's has there begun anything resembling a trend to investigate pragmatics and discourse in naturally occurring interactions (Edelsky 1981). For the turn-taking systematics from the early 70's to have any sort of generality and validity, it was designed—or described—in its most abstract ("simplest") level, without any reference to the specific mechanisms on how it is done, or attention, response times, etc.

Accordingly, it was defined—or described—in syntactic terms. But of course turn-taking decisions involving *when* to start speaking are not taken solely on those terms. Ellipsis, in particular, all the time renders utterances non-syntactical yet recognizable in context. The authors themselves consider "the partial character of the unit-types' description" (Sacks et al. 1974, p.722). There is no way in which the projection of possible completion can be reliably accomplished from purely syntactical terms, and even the 'real-time' analysis that a listener does while hearing an utterance unfold remains probabilistic throughout (Oreström 1983, O'Connell et al. 1990).

As we know, the rhythm, intonation, nonverbal behaviours and other paralinguistic factors can influence understanding, and in parallel the projection of possible completion. Discrimination of "what" either as a one-word question or as the start of a sentence is made intonationally in many languages; and any word can be made into a one-word question the same way. The projection of possible completion should be understood therefore as identifying cues of *possible* turn-yielding. The potential turn-taker must "calculate with probabilities" (Oreström 1983): the more assurance he or she may

wish, the more additional cues of turn-yielding are needed.

Regarding the techniques for selecting a next speaker, Power and Martello (1986) argue that the 'right' or 'obligation' of a selected speaker cannot always be ascribed on the basis of rules, or discerned in purely syntactic ways. Utterances that select a next speaker do not transfer the floor to another party until the speaker finally stops speaking: "What is it, Mary? I thought it was a stone or blah blah . . . ", "Sorry, go ahead, John. I thought you had finished and blah blah . . . ". Indeed, Power and Martello (1986, p.34–35) show that all four techniques for selecting a next speaker can have typical, perfectly acceptable counter-examples (slightly modified to a more condensed form):

- addressed questions: "Have you seen Mike, Jane? I want to show him the car. Ah, here you are. . . " (continues talking uninterrupted to Mike);

- tag questions addressed: "You don't mind helping me, do you? Just for a few minutes, because I want to blah blah. . . ";

- elliptical questions addressing the previous speaker: "Mary was the last person to see John.", "Where?", "Oh, in the garage." (this responded by Mary, not the first person, to whom 'Where?' was supposedly directed);

- social identities or other pragmatic inferences: "Come again next week?", "Yes, we'd love to have your company." (spouses talking to another couple).

It is likely that no technique for selecting the next speaker can really *guarantee* that the appointed party will talk next; so rule 1 also cannot enforce that much, at least not that "no others have such rights or obligations". Those techniques merely but powerfully constrain the subsequent course of the dialogue (Power and Martello 1986, p.35). The 'right or obligation' of a speaker to talk when selected cannot be discerned in a purely syntactic or superficial basis, but from the contents of the utterance, the perceived intentions of the speaker, and ultimately the listeners' decisions.

2.3 Backchannels

Conversation is an interactional achievement, incrementally accomplished. It involves collaboration with the other parties present, collaboration which is interlaced throughout the interaction (Schegloff 1982). Even when only one does the talking, the other participants who are silent *are* relevant to the talk. After all, talk is designed towards

their recipients, who then may nod, laugh, smile, express surprise, say "mhm" or "uh huh" and a host of other vocalizations. It is part of the listener's role to show to be an attentive and interested partner in the conversation (Oreström 1983).

Listeners' reactions can affect the course of the talk, as the speaker is constantly monitoring them to remain recipient-oriented. Exclamations of surprise and doubt ("gosh", "really?", "I don't believe!") or requests for repetition ("huh?", "sorry?", "what?") are common cases. Completing the list of short listener reactions to the talk are questions of clarification and confirmation such as "where", "why not?" and "John!?", mentioned by Sacks et al. (1974) as ways of selecting the recent speaker to talk next (§2.1.2); that is, to continue talking.

On the other side, explicit ways for the speaker to obtain listener responses involve *elicits*, by which the speaker can get confirmation or just attention, even without relinquishing the floor. The following are some kinds of elicits (Oreström 1983): declarative questions with a 'question' intonation ("You got home safely then?"), check-ups ("and you do feel that you're—", "yes"), conclusions ("oh so you know this area is—", "sure"), uncertainty ("January, I suppose...", "yeah"), and lack of knowledge ("I don't know if you would—", "well, I blah blah...").

Tag questions can also invite interaction in the same manner. Besides being used as a 'turn-exit' technique (§2.1.2), they can be just *fillers* and increase 'social contact' ("it's beautiful, isn't it, I always..."), or for quick confirmation ("I see you would go back, wouldn't you?", "mhm", "I thought so and blah blah...").

The listener responses emitted in the *background* of the talk-in-turn, which are not 'turns' and are not *meant* to grab the floor—the attention of the others—, have generally been called *backchannels* or backchannel actions after Yngve (1970). Duncan and Niederehe (1974) classified as such an even broader range of vocalizations, including even longer utterances. They are, for English:

1. signals of continued attention: nods, 'mm', 'uh huh', 'hm mhm', 'yeah', 'yes', 'ok', 'right', 'I know', 'I see', 'thats right', etc;

2. exclamations: 'oh my', 'gosh', 'good God!', 'bloody hell!', 'wow!', 'really?', 'no way!';

3. questions: 'huh?', 'what?', 'where', 'whereabouts?', 'why not', 'did he?', 'was it?', 'who me?';

4. sentence completions: "I think she'll be calmer—", "as she gets comfortable, hm mhm" (completing the other's utterance);

5. brief restatements: "having to pick up the pieces—", "the broken dishes, yeah";

6. clarification requests: "You mean these anxieties, concern with it?" (this one probably is talk-in-turn already, even if it is overlapped to other talk).

2.3.1 Backchannel feedback

From these I shall focus attention only on the simpler forms hereafter called *backchannel feedback*. Utterances of the first group ("uh huh", etc) will be *positive feedback* in contraposition to the short 'question' utterances like "huh?", "sorry?", "what?" which, together with partial repeats of prior talk, will be called *negative feedback*. Hopefully this definition is not that much different from the usual in dialogue systems (Derriks and Willems 1998, Bell and Gustafson 2000).

These responses give the speaker a yes-or-no indication as to how the talk was so far received. They have little content but much interactional value, acting on the relationship-level of the talk (Oreström 1983); they relate to the good of the communication rather than to the talk itself. According to Oreström (1983), those vocalizations "help sustain the flow of interaction; without them, the speaker would sooner or later start wondering whether he is being listened".

In a study of a large corpus of conversation (Svartvik and Quirk 1980), Oreström (1983) made some interesting findings. A significant number of speaking turns (14%) started with "mm", "yes", "yeah" or "ah" as turn initiators—meaning that it is only sequentially that one can identify them as just backchannels. He also equated "mm" as one of the 10 most frequent 'words' in conversation. Further, 77% of all backchannels came at syntactic and/or intonation boundaries (roughly, TRPs), and 54% were uttered *without any overlap*, not in simultaneity with talk. From the one-fourth backchannels that did not come precisely at his syntactic-intonational boundaries, 20% were after the first 1–5s of it, 43% after 6–10s, 17% after 11–15s, and the other 20% came even later than that. All listeners in his data were seldom silent for more than 15 seconds, so there was a high degree of periodicity in the backchannels; it seems a contributional rule of interaction: not to be inactive for too long.

Positive vocalizations ("uh huh" and the like) mean that the talk was adequately re-

ceived, and that there are no problems with it so far, so the speaker can proceed with his or her discourse. These responses were first described as 'signals of continued attention' by Fries (1952): they indeed claim continued attention, understanding, and possibly agreement to some extent. By occurring alone at points at which "the turn might have been claimed" (Duncan and Niederehe 1974) (that is, at TRPs), they also imply that its producer does not want to talk at the moment, exactly because he or she is *passing an opportunity to talk* (Schegloff 1982). They mean 'I am listening, you talk'; whereas talk-in-turn means 'I am talking, you listen' (Oreström 1983).

Negative feedback also conveys no intention to talk (unless its producer continues immediately with a long utterance), but it accuses a problem of hearing or understanding, or possibly doubt or non-agreement. Utterances of this type serve to *initiate the remedying* of any such problems in the current or just-finished talk, and have been termed elsewhere 'other-initiated repair' or 'next-turn repair initiation' (Schegloff et al. 1977).

According to Schegloff (1982), any talk can be a source of trouble, therefore 'after any talk' is a place for its repair to be initiated. Indeed, speakers who are continuing to speak can leave a moment of non-talk for any potential repair. Negative feedback is thus potentially relevant on the possible completion of any unit of talk by another; even when this talk is just suspected: e.g. (the current speaker gets silent), "Huh?", "I didn't say anything".

Schegloff (1982) argues that positive feedback ("uh huh" and the like), in passing the opportunity to do a full turn at talk, can be seen as passing an opportunity to accuse a problem of hearing or understanding on the preceding talk as well. In this sense, it is specifically the *converse* of negative feedback. It is not (he says) that there is a direct semantic convention that equates such utterances as "uh huh" and the like to a claim of understanding or agreement: it is rather that *by passing an opportunity to initiate repair* that they are taken as signalling *the absence* of such problems.

Anyway, he also suggests that the turn status of "uh huh" and the like must "be assessed on a case-by-case basis, by reference to the local sequential environment, and by reference to the sequential and interactional issues which animate that environment" (Schegloff 1982, p.92). Because those vocalizations can also be used as meaning 'yes' after yes-or-no questions, in which case they are *not* backchannels: "Do you think I can do this?", "uh huh", "Right, ...".

Another use of backchannel feedback is in response to an extended gaze by the speaker

which appears to solicit a sign of understanding (Sacks and Schegloff 1979). Such instances occur when the speaker refers to someone who he or she is uncertain whether the listeners know, marking the name of the person with a rising tone to elicit a recognition, what later Clark and Schaefer (1989) called a *trial marker* (§2.7).

2.3.2 Continuers

"Perhaps the most common use" of "uh huh" and the like is to exhibit an understanding that an *extended unit of talk* is under way, and that it is not yet, or may not yet be complete (Schegloff 1982, p.81). It takes the stance that the speaker of that extended unit should continue talking, and continue *that* extended unit—as if saying 'I am following you, do continue'. An 'extended unit of talk' can be a multiple-utterance discourse as exemplified in the next section.

In this context, positive feedback are *continuers*. It is structurally relevant for parties to display their understanding of the ongoing talk at the points where they could have taken turns to talk instead. Continuers display an understanding of the current *state* of the talk, not an understanding of *the talk itself*. A typical use of continuers is in showing that an extended 'unit of talk' is in progress, as in the following example from Schegloff (1982, p.82). There, the extended unit is proposed in the second line below (probably still first utterance) by a preliminary to a preliminary, in a variant of 'Can I ask you a question?' (Schegloff 1980):

    B: I've listen' to all the things that chu've said, an' I agree with you <u>so</u> much.
    B: Now I wanna ask you something,
    B: I wrote a letter
         ((pause))
    A: Mh hm,
    B: T'the governor,
    A: Mh hm::,
    B: –telling 'him what I thought about i(hh)m!
    A: (Sh::::!)
    B: Will I get an answer d'you think,
    A: Ye:s,

Here, some display of understanding is made relevant by the speaker withholding further talk until one is produced, as shown by the intentional pause after "I wrote a letter".

The 'Ye:s' at the end of the example is a full turn as the answer to the last question, rather than passing a turn as in the continuers "Mh hm" and "Sh::::".

Again according to Schegloff (1982), the use of *different* acknowledgements each time, by reflecting a range of different reactions to the talk, may be a signal of interest in it. In contrast, the use of the *same* feedback at four or five consecutive slots may hint at an incipient disinterest. Because of the availability of various options—exclamations of surprise, special interest, assessment—, their non-use would underscore that the listener is not finding anything interesting, newsworthy, or assessable in the talk. It should be noted also that "uh huh" and the like can be spoken in a quasi-infinite extendable range of ways, variously conveying surprise, appreciation, assessment, etc.

2.4 Multi-utterance turns

The turn-taking systematics established that transition-relevance places (TRPs) would occur at the 'possible completion' of turn-construction units (TCUs). But certainly not all TCU completions correspond to equally acceptable places of turn-taking. When the utterance so far projects (variously, as we will see) that *more* is to come, and thus that the speaker is engaged in an extended multi-unit or multi-utterance turn, turn-taking is likely to be discouraged. It remains for the listeners to honour this projection and withhold talk at places at which it would otherwise be appropriate (Schegloff 1982).

More-to-come can be indicated at various linguistic levels: syntactic, semantic, pragmatic, discourse, and dialogue level. An utterance can project a possible completion prosodically, but not be complete syntactically, or semantically, or pragmatically. For example, it may be like the following coordinated sentences (Selting 2000), the brace indicating simultaneity:

> Ida: . . . either (0.7) live at home and work for a car
> Nat: ⌈mhm.
> Ida: ⌊or live here and work for a flat.

In this case there is no pause after "a car" (only a hesitation earlier), so "mhm" comes in the middle of a single utterance. But even if the speaker Ida had paused after "a car", a sense of more-to-come would have made others wait for the rest, in a second utterance then. Similarly, a subordinated construction can be broken in two utterances

separated by a pause, filled or not with listener responses (as in this case):

> A: Although I agree that the process should continue
>
> B: yeah.
>
> A: (0.5) I think that . . .

Semantically and pragmatically there are many possibilities of formulating utterances that are not (and cannot be) wholly understood without the further talk that will follow. This usually makes listeners just 'wait and see' whether there will be more talk to explain what was said, instead of (impatiently) initiating repair or beginning other talk. But the speaker can also announce or propose explicitly some form of extended talk by devices such as 'first of all,' (Schegloff 1982) or other list-initiating methods such as:

> A: There are three things you have to know. (0.5)
>
> B: mm.
>
> A: One is . . .

There might otherwise be no particular need to say that there are 'a number of' or 'a first' thing to say, *other than* to inform listeners exactly that 'more is to come', so that they allow it to be realized.

Another form of multi-utterance turn that can be potentially much longer, and is held together at a higher level of discourse or dialogue organization, is the 'story' or 'big package' of talk (Selting 2000). A whole utterance or turn can be devoted to proposing and negotiating such an extended talk, like in story prefaces (Sacks 1974), such as:

> A: You won't believe what happened to me yesterday!
>
> B: What.
>
> A: I was walking in the park as always. ⎡ Then . . .
>
> B: ⎣ mm hm.

Or in 'preliminaries to preliminaries' of the form 'Can I ask you a question?' (Schegloff 1980), such as in the example of the previous section where the speaker first asks whether it can ask a question, then tells a whole story (the preliminary) before finally making the announced question. So the protocolar "Can I ask you a question?" is not intended to verify the obvious: whether a question can be asked. It is intended to orient listeners to the action therein announced, that is assumed to require, or in any case is preceded by, a preface or explanatory multi-utterance or multi-turn talk.

One last practice of indicating the continuation of talk and hopefully prevent an interruption by potential turn-takers is what I noted frequently in an examined discussion: avoiding TRP pauses in favor of hesitations. That is, finishing an utterance then immediately tagging "and" or "but" or "like" or some other beginning that does not need thinking, *and then* stop, to formulate what more one is going to say. In a sense, it is shifting from tidy pausing *between* utterances to what appears to be hesitation in the *middle* of one, by incipiently beginning it with some appropriate liasion word or phrase: a conjunction, most frequently. Some examples, slightly simplified from a real recorded discussion (commas indicate the end of tone units or intonational phrases):

> . . . some I don't like at all, so (2.0) if- I know if . . .
>
> This ah gives tha- university a bad name as well, because (0.8) yeah
>
> . . . you were ina– a tutorial group, and (1.5) this one person . . .
>
> . . . personal facts don't matter, but (1.8) the other facts . . .

Note that this is different from actually indicating more-to-come *in* the utterance: the TRP thereafter is still a free one, but the speaker manages to stop (to 'pause') in a non-grammatical place so as to *appear* to hesitate instead of strictly pausing. How much this really makes a difference in preventing others from starting still has to be investigated, but it appears to leave those silences in an intermediary position in terms of restricting turn-taking, between a real pause (after nicely finishing an utterance) and a more-to-come as in the examples earlier in this section.

It is likely that in some circumstances this may become as frequent or *more* than making tidy utterances followed by TRP pauses, which might be like *asking* to be interrupted, to be talked over, depending on the group's talkativeness or interest in the topic at the moment. It can be considered a *turn-holding* technique that resorts to a device similar to the appositional beginnings like "well. . . " described by Sacks et al. (1974), which do not require thinking and seem to mark a claim to the 'turn'. As other similar indications of holding the turn, Duncan (1972) found that gesticulation was the main cue of turn-holding in face-to-face conversation (§2.6), and Schegloff (2000) observed a practice of 'rush through' a TRP, possibly changing the intonational contour, to prevent others from barging in (§2.5.3).

2.5   Simultaneous talk: Schegloff's account

By its generality, there were underspecified points in the turn-taking systematics of Sacks et al. (1974). For example, rule 2 (§2.1.4) provides for the possibility that more than one self-selector start to speak at a TRP. Who gets the turn then?

The characterization proposed was a *second-order ordering* of practices in which, foremost, 'first starter goes'. Not only this does not always hold—there *are* cases of second-starters getting the turn—, but it does not explain what happens when more than one party starts to speak at the same time, more or less.

Another underspecification lies in rule 3, whereof a current speaker only continues at a TRP if no one else self-selects to speak. But instances in which both the current speaker and a new one start in the pause (or one starting while another continues to talk) surely exist, even where they did not start simultaneously: one of them started earlier, but none subsequently quit talking.

Schegloff (2000) tried to cover these gaps in his account of simultaneous ('overlapping') talk, encompassing instances of interruption too. It is given as an organization of practices and resources by which the parties involved can arrive at a resolution in a fashion which "allows all parties to incorporate and display the stance they mean to take in view of that moment in the interaction—its content, its issues, its engaged participants, its context, its priorities, etc.—and allows them to adjust that stance moment by moment, beat by beat, as the other's stance is revealed as well" (p.45).[6] This is described in this section.

Just as the turn-taking system operates independently of the context of conversation, this organization is taken to be independent, insensitive, to the *mode of onset* of simultaneous talk. The account is based on this premise, although it is acknowledged that the terms of analysis that are its product may in turn be employed to reexamine the premise.

2.5.1   Definition

Various instances of simultaneous talk are non-problematic, in that their producers are not contesting or claiming the turn space. These include: the common but brief

---

[6]Unadorned page numbers in this section shall refer to Schegloff (2000).

overlaps in speaker transitions, most backchannel feedback ("uh huh", "huh?") and collaborative co-construction (e.g. another's completion of an utterance, word-search suggestions), and also choral or convergent vocalization as in laughter, collective greetings and congratulations, leave-takings, and so forth.

However, there are other sorts of simultaneous talk in which parties do appear to be claiming the turn space and clearly are intent on being heard. These represent 'violations' of the one-at-a-time normal practice of turn-taking to which the systematics is not oriented to, yet it provides the possibility to occur. The most obvious way to resolve such situations then is just stop talking. But who should stop? To the organization it is indifferent, though the individual parties may care much, or not at all.

Simultaneous talk means, overwhelmingly, just *two* talking at the same time. More than two at a time is reduced to two even more effectively than two is reduced to one. The basic configurations of multiple talk orientation are therefore these three (p.8):

$$1) A \leftrightarrow B \qquad 2) A \rightarrow B \qquad 3) A \rightarrow B$$
$$\qquad\qquad\qquad\qquad \uparrow \qquad\qquad \downarrow$$
$$C \qquad\qquad\quad C \qquad\qquad\quad C$$
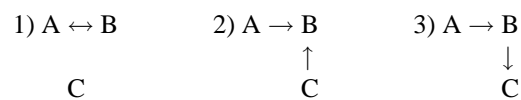
The first is the most common case, involving two people talking to each other, with others listening. In the second case, where two parties are talking to a third one (B), the gaze of this third party is going to figure centrally as to whom he or she is seen to be listening, which may indicate who will continue talking. In particular, B would pay attention to one party (say, C) and subsequently respond to it, in which case the orientation (assuming the other party's simultaneity persists) changes to the third configuration. This in turn can change back to the second configuration when the recipient of the attention responds, and so on. In this light, 2 and 3 are natural alternators.

As with the "grossly observable facts" that were the premises of the turn-taking system, Schegloff starts up by laying out other grossly apparent observations that any account of this topic should come to terms with, explain, and foster (p.10–11):

- most occurrences of simultaneous talk are over very quickly (by the second or third 'beat' whereat the parties involved recognize the simultaneity);

- some persist to considerable length, although no specific stipulation of a limit can, or even should, be made;

- many such occurrences are the locus of hitches and perturbations;

- the management and resolution of simultaneous talk should accomodate other

non-turn-taking interactional interests, such as the participants' stances and reasons for persisting to talk simultaneously;

- and it should be compatible and systematically related to the organization of turn-taking.

The organization of practices for resolving simultaneous talk, like the turn-taking system, is described as composed of: a set of resources of turn production; a set of places, or phases, at which those resources are 'deployed'; and an interactive *logic* by which the application of those resources in those places constitute moves of a describable sort. The 'logic' amounts to the specification of the parties' alternatives beat-by-beat (moment-by-moment) in the course of the simultaneous talk. These components will be explained in the following subsections in turn (and not simultaneously).

2.5.2 Resources: hitches and perturbations

One of the observations above is that many occurrences of simultaneous talk are the locus of *hitches* and *perturbations*. Hitches are discontinuities in the course of the talk's production, and perturbations are marked departures from the normal prosodic articulation of talk. Few of these (if any) are exclusive to simultaneous talk. It is their 'especially dense' occurrence and 'strategic import' that are worth of note here.

The ongoing talk can get louder suddenly, higher in pitch, and faster or slower in pace. It can also be self-interrupted (cut-off), the last word or syllable be repeated, or some next phoneme may be prolonged or stretched out, and any these combined. Most typical is a sequence of cut-offs and repetitions of the last word or syllable.

These hitches or perturbations reflect the speaker registering or responding to the fact that another party is speaking simultaneously. Some may be taken as strategic manoeuvres in a competitive undertaking, in the *fight for the floor*, or as casualties of the process when their participants eventually accept and concede. Phoneme stretchings are common examples of the latter, that can be explained as cognitive interference of the simultaneous talk in one's own articulation. For example (p.13):

> Deb: How ⌈come you get thiz:: <u>thi:s</u> v::::::ersion of jovial
> Anne: ⌊W'd you please concentrate on driving the ca:r.

So, these are some resources applied in managing the course and resolution of simultaneous talk. A simultaneous speaker may make use of them to restart his or her intended utterance *in the clear* once the other speaker arrives at a projected possible completion. It is a strategy of 'getting' the conflict behind and override the other's talk.

In this context, the projected possible completion of turn-construction units are the places towards which parties orient themselves to be the first to emerge in the clear and have the advantage of precedence ('first starter goes') to take the next turn. Hitches and perturbations are deployed at the *prospect* of imminent resolution of the simultaneous talk to absorb the remaining overlap. And more than one party, or all, may attempt to do the same thing in the same episode, leading to interleaving sequences of false-starts and restarts as they try to get over the other and come out alone in the clear.

This is but *one* strategy or practice of the occurrence of those hitches and perturbations. Not much more is said about why and which of those resources get used in what cases, except with general statements followed by examples such as (p.14)[7]:

Deb:	I don' re⌈member, it wz SUCH A MU::DDLE I w'z GRA:DUATING that–
Anne:	⌊I HAVE NO WAY OF PRO:VING IT, YOU GOTTA-

2.5.3 Phases of simultaneous talk

Responses or changes of stance due to simultaneous talk can begin even before its onset, when the parties detect its potential, and can extend further after its resolution, when they readjust themselves back to the normal (one-at-a-time) talk. In this respect, the places or *phases* relevant to simultaneous talk are:

- the pre-onset phase;

- the post-onset phase;

- a possible post-post-onset phase, in cases of simultaneous talk that extend to substantial length *in the middle of it*, becoming the arena of exchanges of move and countermove in the dispute for the turn's space (the floor);

- the pre-resolution phase;

---

[7]As in other Conversation Analysis examples, transcriptions use a set of conventional symbols: colons for phone stretching, punctuation for intonation and slight pauses, underlining for emphasis, dashes for cut-offs, braces for the beginning of simultaneous talk, and capitals for relatively louder talk.

- the post-resolution phase;

- and a possible post-post-resolution phase, in the turns subsequent to the one following the resolution, where repairs of various sorts are attempted to 'heal' possible *casualties* of the episode.

In the pre-onset phase, an already speaking party can detect the possibility of an incipient start by another, via *turn-claiming behaviours* like body posture displays or common pre-turn-beginning practices such as audible inhalation. The current speaker can address the prospect of an imminent simultaneity and try to interdict it before its onset using the previously presented resources. The speaker may suddenly raise the intensity (loudness) or pitch, or change the rhythm or tempo (pace, speed) of his ongoing utterance, or all of these combined.

Another case is when a projected possible completion in the speaker's utterance is the event to which one other or more parties are orienting themselves to start to speak. In this situation, the speaker may *rush through* it, barring prospecting self-selectors of ever starting simultaneous talk. This practice of rushing through a potential utterance completion (a TRP) is done by 'deploying' changes in the talk: accelerating its tempo or rhythm, levelling the intonation and raising loudness. These actions are then real *turn-holding signals*, or strategies for keeping the floor.

In the post-onset phase, after simultaneous talk has already begun, hitches and perturbations are used to register this fact, and can be reactions to the other simultaneous talkers in the form of an upgrade, or 'stepping-up', to *competitive talk*. This process is mentioned in the next subsection. Upgrades to more competitive talk can include tempo changes (e.g. slowing down), sharp loudness increase, phoneme stretching, cut-offs and repetitions, and restarts of utterance beginnings. However, no account of their differential application (which resources are applied when) is attempted, nor is any statistics available.

As was previously commented, the pre-resolution phase is a common site of hitches and perturbations because of the prospect of natural resolution by the upcoming possible turn-unit completion. Speakers can then *sacrifice* (interrupt) their utterances-in-progress at that point in order to restart 'in the clear' at the projected utterance completion of the other's talk.

In the post-resolution phase, adjustments may be needed in gearing down the remaining speaker's talk to solo production, after it may have been upgraded to a competitive

mode during the simultaneity. The practices that are warrantable in such conflicts are otherwise problematic in the normal one-at-a-time condition. In the face of simultaneous talk, speakers may have raised the intensity or pitch of their talk up to the point of almost shouting to one another. Hence they may find themselves accountable in quite a different way once the simultaneous talk is over. It is in this post-resolution phase that the sole speaker has to bring his or her talk back to normal one-at-a-time condition.

2.5.4 The interactive 'logic' of simultaneous talk

In the normal process of one-at-a-time talk, the *position* towards which parties orient themselves is always 'next turn'. In simultaneous talk, this cannot supply the grounds for interactivity, since it is exactly the turn space that is at issue, that is being contested. The relevant level of granularity in which decisions are based, then, appears to be the *beat* ('moment'), which Schegloff has assumed for the while to be "substantially equivalent" to the syllable (or foot, prosodically).

"I am using the term 'beat' not in the technical sense of the literature in linguistics, but as a simpler, perhaps even vernacular, term for the syllable-like increments of production by which talk-in-its-course (and its silences) is produced" (p.51, note 22). This seems to bear an association with the timing and rhythm of talk, albeit the connection is not made explicit. Edelsky (1981, p.194) has emphasized timing and rhythm as basic interactional components in her observations on *collaboratively-developed* 'floors' of talk in a study of informal committee meetings.

With the onset of simultaneous talk, the parties involved may decide at each beat in its developmental course one of the following: whether to withdraw, to continue, or upgrade the talk to a more competitive mode. The specification of the parties' alternatives beat-by-beat (moment by moment) in the course of the simultaneous talk constitutes its "topography of sequential and interactive organization" (p.20):[8]

- first beat: at the onset of simultaneous talk, parties are busy producing their utterances; any reaction can only take place later;

- second beat: speakers have then heard the simultaneous talk and (in case they

---

[8]In the simulation of discussion described later in the thesis, the 'beat' or moment can be conveniently equated to the *cycle* of simulation to which the agents' behaviours and decisions are synchronized. The *response-times* of people in conversation still have to be determined: how long it takes for one to recognize changes in the interaction and react to them.

recognise it) have to decide whether to stop or continue, or upgrade to a 'competitive' talk by applying some of the resources (hitches and perturbations) previously described: e.g. raise loudness, pitch (or both), stretch a phoneme, begin a cut-off and restart cycle, etc;

- third beat: if the speakers continue by the third beat, they have now heard initially what stance the other or others have taken regarding the situation, and then face the decision of how to respond to it: if one has upgraded talk (to a more competitive mode), the others must decide whether to withdraw, to continue speaking as if in solo talk (paying no attention to the simultaneity), or 'take the challenge' and raise the tone as well;

- fourth beat: if a second speaker has upgraded the tone of talk too, the first one faces the decision of whether to finally drop out, to continue in the same 'level', or to raise the ante even further: getting louder still, higher in pitch, or recycling the turn-so-far (cutting-off and restarting).

It is by the third beat that the majority of episodes of simultaneous talk is resolved (according to Schegloff), as was earlier observed: "most cases are over very quickly". By the fourth beat, if speakers proceed in the course of alternately raising the ante, we have the sort of extended simultaneous talk that grows long enough to include a post-post-onset phase in which the contending speakers move and countermove trying at the same time to evade the overlapping talk and to deliver their utterances in a sequentially implicative way, so they get *heard*. These cases are rare—at least in the sorts of data the author has examined.

2.5.5 Resolutions

Many episodes of simultaneous talk are resolved after the first beat by the withdrawal of one or both parties involved (leaving behind false-starts then, if they had just started). This is common in speaker transitions when two self-selectors start to speak simultaneously (or nearly so) but only one continues:

> B: ⌈Please—
> A: ⌊Wha' was that again ma'am?

Both speakers withdrawing is also usual, sometimes engendering an alternating succession of overlaps and gaps of silence not unlike two pedestrians walking towards

each other and resolving the potential encounter by choosing the same bypass direction, facing each other again at a stand-off, then repeating this until someone gives in to the other. Having all the simultaneous starters stop because of each other is what I will later call a 'collective' false-start (chapter 6).

Of the remaining cases of simultaneous talk, a great many are resolved within one beat of someone upgrading talk to a more competitive mode:

A: I⌈say c' d–
D: ⌊Her <u>name</u> is Kellerman, si:r.

These possibilities account for the "vast majority of cases" resolved to a single speaker by the third or sometimes fourth beat.

Persistence in talking simultaneously makes relevant the speakers' outward turn-taking interests, in their need to produce the ongoing utterances only then, and just there. It is the need of *that* turn in particular for a responsive action that requires *that* turn in that context, and cannot be delayed ('now or never'). This is common in debates or interviews with politicians where the 'windows of opportunity' to talk about specific topics are short and fast moving.

Or it may involve status issues for which any sort of deferral may be consequential. For example, a pun that would require that very turn-position for its recognition as such, or a credibility issue demanding immediate reaction, to which an ommission would have implicational consequences. The contenders' persistence in speaking simultaneously is thus one way of displaying that some interests are being pursued, and that their identification by the others is possible at the moment.

However, by no means the practices of simultaneous talk resolution involve only conscious decisions driven by specific interests in the moment of the interaction. The identity of the parties in the process may weight significantly in the decisions to continue or withdraw at the prospect of simultaneous talk: e.g. employees talking to their boss. Identities and relationships of the parties in conversation (such as their acquaintance and liking, too) are thus significant variables in this context. Conversants who deal with each other on a routine basis may well have developed recurrent, if not routinized, trajectories of resolution of simultaneous talk.

But to 'win' the floor is not the only goal to which participants in simultaneous talk orient themselves. Oftentimes a speaker just wants to complete his utterance, and then

may pay no attention to simultaneous talk. Strangely, this winds up being the strongest stance that one can possibly display, precisely by the non-recognition of the simultaneity itself. If completing an ongoing utterance is not feasible or otherwise desirable anymore by the fact of the simultaneity, speakers at such a situation may persist just to the point the utterance's *thrust* or *upshot* has been projected and is recognizable, so as to make the effort worth it.

Finally, they may attempt to obtain some sequential implicativeness or consequentiality: i.e. by getting oneself's talk, and not the *other's* talk, to be addressed later. This can be achieved by the usual strategy of withdrawing from the simultaneous talk to restart the utterance *immediately* after the other's talk has ended in order to get over it and have one's *own* talk subsequently addressed ('losing the battle to win the war').

A similar strategy is when the speaker senses another person coming in to anticipatorily and collaboratively complete the turn he or she has begun. In such a case, the ongoing speaker simply lets the other talk by withdrawing from the simultaneity. That participant then delays his or her own completion of the utterance until after the interloper's contribution has finished: in a sense, just passing over and overlooking (or ignoring) the other's attempt at collaboration.

2.5.6 The aftermath: degrees of taking notice

After the resolution of simultaneous talk, adjustments need to be made, specially if the remaining speaker's talk was upgraded to a competitive tone during the episode. How it may have been taken by the remaining speaker is shown by the reactions (hitches and perturbations) after its resolution.

The remaining speaker may have taken notice of it or not, registered it or not. If it was registered, it can be taken as problematic or non-problematic, in which case it may have passed simply as an unnoticed blip or a positively sought co-construction. It may require a response at the level of turn-position occupancy, or it may be attended to for its bearing on larger units of interaction, such as shaping the immediately ensuing direction of talk. Or alternatively it may be attended for its consequences on the overlapped talk's hearability or understandability, or its efficacy.

The forms and degrees of taking notice and registering the consequences of the simultaneous talk are thus:

1. no notice was taken;

2. just a post-resolution hitch in the aftermath (with no perturbation been displayed during the episode itself), which is possibly the subtlest way of registering the occurrence;

3. recycling (the whole or part of) the overlapped talk by speaking it again 'in the clear';

4. addressing the other's overlapped talk via negative feedback, at times showing it to have been robustly grasped, at others treating it as in need of repair ("huh?");

5. addressing the other's overlapped talk, showing it to have been grasped, then immediately restarting (or restating) one's own;

6. initiating repair on the overlapped talk of the other speaker, via a repeat or partial repeat, or by "huh?" or "what" (or other initiators), which indicates that nothing was understood.

The last three are all negative feedback that vary from a single backchannel-length utterance to a whole (longer) turn. In these cases, again, there's the possibility of synchronized alternation of gaps or overlaps wherein more than one contending speaker tries to do the same thing. After one party's overlapped talk is repaired, the other's may be requested to be, too.

2.6   The psycholinguistic approach

Until the late 1980's there have been two broad approaches to describe turn-taking. One was the ethnomethodology approach of Conversation Analysis examined in previous sections. The other was the psycholinguistic approach which tried to identify in analytic studies the various cues and signals that would facilitate turn-taking. I will describe results of this line of research here for completeness' sake, although it bears little import on the simulation described in this thesis. Nevertheless, it shall demonstrate an obvious road for its expansion that is outlined as future work in chapter 7.

2.6.1 Cues and signals of turn-taking

There are several cues in talking that can inform the listeners as to whether the speaker is finishing a turn-unit (an utterance) and whether he or she is intending to continue or may yield the floor. These range from the syntax and semantics of the talk, the intonation and a few other paralinguistic cues, and nonverbal (visual) behaviours in face-to-face conversations. Likewise, there are signals given off by the listeners that indicate whether they are accompanying the talk (continuers) or intending to talk.

One advantage in looking for such signals is that they are relatively independent of the *contents* of the talk, and thus make it easier to identify beginnings and ends of turns without too much complex natural-language analysis. One evidence of their importance for turn-taking is that, for example, syntactic and semantic completion seem to be overruled by certain nonverbal behaviours (Beattie 1981).

Duncan (1972) was one of the earliest to identify such cues, proposing a turn-taking mechanism mediated through a series of "signals composed of clear-cut behavioural cues, considered to be perceived as discrete" (p.283–4). He videotaped, transcribed and analysed a pair of behaviour-intensive, 19-minute face-to-face conversations between two people (dyads). One was a routine intake interview at a Counselling and Psychotherapy Research center between a 20-year-old female regular applicant for therapy and an experienced 40-year-old male therapist interviewer, both previously unacquainted. The second dyad was a relaxed and lively conversation between two 40-year-old male therapists who were friends and had known each other for about 10 years. Both conversations would have taken place regardless of the recording.

Based on their analysis, Duncan hypothesized three turn-taking *signals* with associated rules. The first was the *turn-yielding signal* which indicates to auditors (listeners) that the speaker may yield the floor: "Under the proper operation of the turn-taking mechanism, if the auditor acts to take his turn in response to a yielding signal by the speaker, the speaker will immediately yield his turn" (p.286). The *signal* would be indicated by one or more of six turn-yielding cues identified as:

1. syntax: completion of grammatical units ('junctures' in the talk);

2. intonation: rising or falling pitch at the juncture of grammatical units;

3. body motion: termination of a hand gesture away from the body, or relaxation of a tensed hand position such as a fist;

4. drawl: lengthening in the final or stressed syllable of a terminal clause;

5. paralanguage: any drop in pitch and/or loudness in conjunction with stereotyped expressions; and

6. stereotyped expressions like (for American English) "you know", "but uh", "or something", often coupled with a marked paralinguistic *trailing-off* effect.

Duncan found that the probability of (auditor) speaking attempts appeared to increase in a linear fashion as more of these cues were conjointly displayed, to a total of 92% of those speaking attempts (with one or more cues) resulting in smooth transitions. However, probability to speak was less than 50% generally[9]: that means the auditor still retained considerable discretion over *whether* to talk or not. Furthermore, the chance of occurring simultaneous talk (excluding backchannels) was sharply decreased after the display of turn-yielding signals. On the other hand, each time an auditor attempted to speak in the absence of any cue, simultaneous talking ensued. Those attempts could be straightforwardly interpreted as 'interruptions' (unsmooth ones).

The second hypothesized signal was an attempt-suppressing, or *turn-holding signal*, displayed by the speaker. It consisted of one or both of the speaker's hands being engaged in gesticulation[10], suppressing any auditor attempts regardless of the number of turn-yielding cues concurrently being displayed. Such attempts were practically zero when the turn-holding signal was displayed together with one or more cues of turn-yielding. Curiously, *more* auditor attempts took place when the turn-holding signal was displayed with *none* of the turn-yielding cues. It is difficult to conclude anything from this, except that such instances can also be considered interruptions.

The third signal Duncan hypothesized was the backchannel vocalization. "[I]t appears that, when a speaker is displaying a turn-yielding signal, the back channel is often used by the auditor to avoid taking his speaking turn. In this sense, taking a turn and communicating in the back channel may be considered to be contrasting tacks" (p.288). He only identified the types of backchannels, as already listed in a previous section (§2.3); he did not present any quantitative or qualitative (contextual) analysis, though Oreström (1983) did (see §2.3.1).

---

[9] Frequency of auditor speaking attempts was 10% with one turn-yielding cue (12 attempts), 17% in the most frequent case, the display of two cues (25 attempts), and 33% with three cues (29 attempts). It reached 50% only with all six cues: 1 attempt in 2 cases: hardly evidence of any pattern here.

[10] All hand and arm movements performed at some distance from the body ('gesticulations') are linked with speaking, strongly with the content and rhythm of speech. "They may punctuate, qualify, illustrate or concretize what is being said" (Freedman and Hoffman 1967).

However, backchannels were investigated in more detail in two later studies. Duncan and Niederehe (1974) noted that "for some of the longer back channels, particularly the brief restatements, the boundary between back channels and speaking turns became uncertain. On an intuitive basis, some of these longer back channels appeared to take on the quality of a turn" (p.237). This suggested that more than the description of backchannel forms would be needed to differentiate them from attempts to speak in turn. They therefore decided to search for a signal that would unambiguously differentiate speaking turns from the beginnings of backchannels.

Analysing the same dyadic conversations of the previous study, they noticed some behaviours occurring regularly near the beginning of speaking turns. This led to the identification of four cues characterizing a *speaker-state*, or auditor *turn-claiming signal*. These cues would be displayed typically at the vicinity of turn beginnings, but not with backchannels, marking those points at which an auditor would shift to a speaking state. The cues identified are: shifting head direction away from towards the other, initiating gesticulation, audible inhalation (sharp in-breathing), and overloudness at the beginning of the talk (which is supposedly not a backchannel).

In the first conversation (the interview between unacquainted different-age and -sex persons), one or more turn-claiming cues were displayed at 72% of 61 turn beginnings and at 9% of 32 backchannels. The body behaviours (head turning away and gesticulation) were the most frequent, any or both occurring in 62% of turn beginnings, whereas audible inhalation and overloudness occurred only in 25% of them. In the second conversation (the informal talk between two male friends), one or more cues appeared at 95% of 20 turn beginnings and 19% of 85 backchannels. Inhalation and/or overloudness occurred in 55% of the 20 turn beginnings. Furthermore, 18 of 19 instances of simultaneous talk in the data—though a small number for the results to be more than suggestive—could have had their resolution predicted by the display of the turn-claiming and turn-yielding cues: whoever displayed more of the former and less of the latter retained the floor.

Beattie (1981) criticized these studies on the basis of quantity of evidence and interobserver reliability on identifying cues. He proceeded to his own study of 6 natural dyadic conversations between university supervisors and supervisees, involving 12 different people and more than 3 hours in total. There were 214 smooth speaker-switches and 16 non-smooth ones. He found that 13.5% of all smooth transitions occurred *without* any of the hypothesized turn-yielding cues; Duncan did not observe any such cases—

there were 5 transitions without cues that were considered unsmooth because they led to simultaneous talking.

In Beattie's study, there were a lot more speaker transitions with a finishing speaker displaying 3 turn-yielding cues, and none with 5 or all the 6 cues described by Duncan. Syntactic clause completion was the most frequent turn-yielding cue, being observed in 61% of transitions involving any cue. It tended to be most often accompanied by a change in pitch level (rising or falling intonation), and less often by drawl on the final syllable. In contrast, gesture relaxation occurred in only 8.7% of smooth transitions (80% of them after clause completion), with pronounced differences between speakers. Like Oreström (1983), Beattie then concluded that "syntactic and accompanying paralinguistic cues play the dominant roles in the regulation of turn-taking in conversation, and that visual and other nonverbal cues are much less significant" (p.63). He attributed the proportional differences in drawl and intonation to the different accent (British) of his study.

2.6.2 Speaker-auditor interaction

Duncan (1974) also further analysed his same transcribed data for evidence of interaction between the speaker and auditor during talk. He noticed relationships between auditor backchannel behaviours and the speaker's verbal and nonverbal behaviours during his or her turn. This led to the hypothesis of a speaker *within-turn signal* marked by the completion of grammatical clause and/or the turning of the speaker's head towards the auditor. The former is a turn-yielding cue, but not the latter "because it failed to differentiate smooth exchanges of speaking turns from instances of simultaneous turns" (p.167)—more on this in the next subsection. These cues occurred at the ends of the units of analysis: tone units or phonemic clauses, seemingly equating to the TCUs of Sacks et al. (1974).

So, the within-turn signals occurred when the speaker ending a 'unit' of talk (a TCU) turned his head toward the listener for uptake. Either cue were directly associated with the subsequent displays of auditor backchannels, both verbal and nonverbal (nods or other head or face responses), and by *speaker-continuation* signals: the speaker turning his head away from the auditor. This would mark the beginning of new 'units' of talk much in the same way as speaker-state signals beginning a new turn of talk. This is in line with the observation that the turning away of the speaker's head occurred not only

at the beginning of turns, as Kendon (1967) had reported, but also in their middle.

Duncan interpreted those findings as suggesting the existence of 'units' segmenting the talk—again much like the TCUs of Sacks et al. (1974). Regarding their relationship to auditor backchannels, he noted that *early* backchannels (within a 'unit') significantly increased the probability of the speaker displaying the continuation signal (i.e. turning of the head away), whereas a 'between-unit' backchannel (i.e. at a TRP) did not. Between-unit backchannels were also not followed by speaker continuation signals when *not* preceded by the speaker within-turn signals (not at the ends of grammatical clauses or prompted by the turning of the speaker's head toward the auditor). It then appears that

> "both the display of an auditor back channel, and its location, may play a part in speaker-auditor interaction. That is, an early back channel may not be merely misplaced, but rather it may carry significant information for the interaction. (...) an early auditor back channel may indicate, not only that the auditor is following the speaker's message, but also that the auditor is actually ahead of it. (...) In contrast, a between-unit auditor back channel would indicate that the auditor is following the speaker's message as it is developing. (...) By the same logic, a late auditor back channel would indicate some auditor acknowledgement, but also that he is not quite following the speaker's message." (Duncan 1974, p.179)

Finally, Duncan found out that speaker turn-holding signals (gesticulation) do not appear to affect the display of backchannels, which were widely distributed throughout the turn. Visual backchannels (head nods or shakes) tended to occur more frequently than expected between units of talk, whereas vocal backchannels less frequently. All backchannels tended to follow the display of the speaker within-turn signal (syntactic completion and/or turning away of the head): 89% of them did. "An almost perfect linear relationship was found between the number of speaker cues displayed and the probability of an auditor back channel, either vocal or visual. Similar relationships were not found for the remaining speaker turn cues" (p.172). All the while, none of the findings with respect to speaker cue display suggested that vocal backchannels should be considered distinct from visual backchannels.

2.6.3 The role of gaze

Apart from Duncan, a few other studies mention kinesic (body movement) behaviours at all in regards to regulating the flow of talk in face-to-face encounters. Kendon (1972)

speculated that body movement would precede and thus indicate the introduction of any speech unit: the larger the unit, the more extensive the movement. Wiemann and Knapp (1975) indicated that certain body behaviours like gesticulation and leaning forward would be effective as turn-claiming signals, though they seemed to be the only ones to study this. They also identified leaning backwards as a further indication of turn-yielding by the speaker.

Kendon (1967) was the first to study *gaze*—which Duncan identified only by 'head direction'—as having any function in regulating turn-taking. Analyzing 5 to 9-minute samples from 7 dyadic conversations, he identified the speaker gaze as an important signal of both yielding and holding the turn. The speaker would typically gaze away when beginning an utterance and gaze back at the auditor when finishing it, much what Duncan (1974), Argyle and Cook (1976) later observed. He noticed that more than 70% of the utterances terminating with speaker gaze were followed immediately by talk from the auditor, in contrast to only 29% terminating *without* gaze. But 38% of all smooth transitions occurred this way, without speaker gaze.

Beattie (1978), in contrast, did not find that gaze at the end of utterances influenced either speaker transitions or their length, based on the proportions of immediate and short-latency intervals in his data. Neither did it in longer utterances of 30 seconds or more (Beattie 1979). Instead, he found more immediate transitions when speakers were *not* looking at auditors at the conclusion of a turn than otherwise. Reviewing other studies, he concluded that speaker gaze is clearly not an essential cue (though a prominent one) in regulating turn-taking. It may facilitate turn-taking in contexts where overall gaze is low, such as between strangers as in Kendon's data, or in 'difficult' topics, such as intimate topics spoken to non-intimate persons (Beattie 1981).

Rutter et al. (1978) were also unable to confirm Kendon's pattern of gaze away at the beginning of utterances, but found instead that speakers *were* generally gazing at auditors at the beginning of new utterances. Moreover, they concluded that *auditor* gaze was also not essential to turn-taking, since almost 33% of speaker-switches they had examined occured without it. Kendon (1967) and Argyle (1972) described the auditor behaviour of customarily looking away and looking back at the speaker.

Another non-speaker behaviour that was absent from dyads (and so unmentioned by any of the previous researchers) is the distraction, the temporary 'withdrawal' of attention from a discussion, presumably to think away for a few seconds, which Kalma

(1992) found in triads (3-person groups). In dyads a person is always an interlocutor, a partner in the verbal interaction, whereas in bigger groups one or more of the participants can afford to be inattentive from time to time.

The same conclusion reached Harrigan and Steffen (1982) in analyzing a 5-person group of acquainted mixed-sex young adults videotaped during a heated discussion of a local teachers' strike. Unlike speakers in dyads, speakers in this group tended to gaze toward an auditor at the beginning of 79% of 250 speaking turns, more so in interruptions, either successful (90%) or not (83%). Even in 63% of overlapped beginnings (with a mean length of 0.4s), new speakers were gazing someone.

"A dyadic speaker need only assess the listener's intention of taking a turn and so can afford to look away from the auditor when beginning one since there are no turn competitors. In a group interaction, however, one may be required to gaze at an auditor in an attempt to win the auditor's attention and thus have an acknowledged speaking turn" (Harrigan and Steffen 1982, p.168). Gazing an auditor serves not only as a cue for taking a turn, but also as a way of engaging someone's attention, which is not as essential in dyads as it is in group conversations: "a group speaker cannot assume an auditor but must engage one" (p.168), which sums up the fulcrum of most differences between dyads and small groups.

On the other hand, gaze orientation at the conclusion of speaking turns was consistent with previous findings: 69% of speakers gazed toward an auditor when finishing a turn. And auditors were looking at the speaker when emitting 65% of 93 backchannel responses.

Kalma (1992) further investigated gaze patterns in a series of three experiments involving 3-person group discussions (triads), identifying a special gaze pattern at the end of a speaker's turn. The first experiment examining the nonverbal behaviour of 120 males in triads revealed a distinctive pattern of extended gaze at the end of turns, as if inviting someone to speak. In 95% of cases, the person looked at was the next to speak; when this did not happen, the current speaker simply continued after a pause.

A second experiment confirmed the prolonged gaze by *appointed* leaders in a task-oriented conversation, showing that the behaviour would not be just an individual characteristic. Prolonged gaze was defined more precisely as only those cases where the speaker began gazing at someone shortly before the end of his utterance and continued for at least one second afterwards while no one spoke. In 83% of the cases in

which this gaze was displayed, the person looked at took the next turn. The situational demands of the task, however (including the need to get information about the other conversants to assign speaking turns), may have stimulated the use of these prolonged gazes. The author's interpretation was similar to Harrigan and Steffen (1982): in 3-person groups the problem of who will speak next is not as trivial as in dyads, so gazing is one mechanism that indicates from whom the next contribution is expected.

The third experiment registered a total of three hours of free discussion in 23 same-sex triad groups, 13 of females and 10 of males: there were no gender differences, so the results were combined. They showed that the person who displayed a prolonged gaze would yield the floor, with the receiver of the gaze being the most likely to take the floor. Prolonged gaze here occurred in only 2% of the total speaker switches, once every 2.2 minutes. Only in one instance the person displaying the prolonged gaze continued speaking, whereas the receiver of the gaze took over 70% of the time—significantly more than expected by chance. They also found out, in relation to the distinction between gaze and head direction, that few instances of gaze shifts occurred *without* head movement, excluding *expressive acts* in head nods or shakes, poising or cocking of the head, and shrugs.

It can be concluded then that gaze is a prominent but not an essential component of face-to-face verbal turn-taking. It was even argued (Beattie 1978) that, when gaze is eliminated as in telephone conversations, verbal cues like intonation and grammatical junctures would 'take over' the function of turn-yielding, indicating reasonably well when someone has finished speaking. Kalma (1992) argued that this may be true in dyads where there can be no mistake about the next speaker, but not in the triads he studied, where gaze can be used as an additional signal to indicate from whom a response is expected. Contrary to Kendon (1967), he found out that utterances ending with an extended gaze of more than one second had significantly longer transition intervals (silent gaps) than utterances ending without gaze—and not shorter, as in the hypothesis attributing to gaze a floor-apportionment function.

2.7 The collaborative dialogue approach

From the 1980s onwards, various models of discourse and dialogue have been developed, such as Kamp (1981), Polanyi and Scha (1985), Grosz and Sidner (1986), Litman

and Allen (1987), Clark and Schaefer (1989), that provide systematic elements on *what* to say, and how to reply to what the other says—dialogues were always assumed to be between two people only. In terms of turn-taking, these theories should provide the means for *why* speaker transitions would occur as they do. From these models and theories, work until the present day has tried to extend the capabilities of dialogue systems with richer discourse planning and linguistic resources that can interact more naturally with people.

Models of discourse and dialogue, therefore, should help in the production of conversational turns by informing *why* and then *whether* to talk at a given moment or to a reply in a given context of the discourse, thus realizing turn-taking in the process. But they say nothing about the management between listening and speaking, when to continue listening for more, and exactly *when* to begin a reply or a new contribution (except as: when the other has stopped speaking). What these models lack to reproduce human conversation more naturally is (among other things) the moment-to-moment behaviours that the simulation in this thesis tries to reproduce symbolically, which allow parties to negotiate the floor, yield it to others, and start to speak at appropriate junctures, as people do in group discussions—not necessarily ever successfully and smoothly. The way forward then is to couple those models with turn-taking behaviours like those simulated here to obtain more natural conversants (avatars).

One common suposition of the predominant theories of discourse and dialogue is the accumulation of some form of *common ground*: the mutual knowledge that participants build as a result of the talk, knowledge held by them all. It does not include the beliefs, assumptions and other information they may have individually, which in general is one cause for talking so that (some of) this information is communicated to others—turned into common ground—when it is convenient for the parties to do so.

To Clark and Schaefer (1987, 1989), conversants contribute information to their common ground in an orderly way. They have a *mutual responsibility* (Clark and Wilkes-Gibbs 1986) to ensure that what is meant by a speaker each time is understood well enough for the current purposes. Hence speakers make sure they are being attended and heard (Goodwin 1981), and they suit the contribution to their specific listeners, or *addressees* (Schober and Clark 1989). Listeners in turn give feedback as to whether the utterance was heard and understood, and how it was received. This feedback can come in the backchannel of talk, or when the speaker pauses to catch breath or monitor the uptake of the utterance.

So speaker and listeners engage collaboratively in a process of presenting a contribution, accepting it, accepting the acceptance, and so on, via feedback and subsequent contributions. Positive feedback such as "uh-huh", "yeah" and many other displays of understanding and evaluation indicate acceptance. Negative feedback such as "huh?", "sorry?" and various other queries inform that something in the contribution was not successfully communicated, or requires confirmation, repetition or amendment. Speakers can exploit the precise timing of their utterances with brief interruptions to manage this process. And if a listener interrupts the speaker at some point, it is generally because what is going to be said is relevant at *that* moment.

Clark and Schaefer (1987) calls *participatory acts* the actions of presentation and coordinated acceptance in this process, whereof speaker and listeners are engaged in building their common ground. It is very much like shaking hands or singing in unison: the collective result depends on the coordinate acts of the parties. The process can be seen in this simple exchange (Clark 1992, p.146–7) from the London-Lund corpus (Svartvik and Quirk 1980)[11]:

> A.   is it . how much does Norman get off – –
> B.   pardon
> A.   how much does Norman get off
> B.   oh, only Friday and Monday
> A.   m

Listener B indicated with negative feedback ("pardon") that she did not heard or understand the question. The questioner (A) then repeats his utterance, with B now showing understanding ("oh") and proceeding to answer it, which by itself gives more evidence of understanding. A might have rejected B's answer (and informed why) if it showed the answerer did not understand the question, but he just says it is satisfactory ("m"), and the dialogue proceeds.

The first B-A exchange above is a *repair* side-sequence, begun by the listener. Repairs can also be initiated by the speaker himself (Schegloff et al. 1977) as in A's self-correction in the first utterance ("is it . how much..."). That is an instance of what I circumscribe as *hesitation* in the simulation of chapter 4. Usually the speaker tries to correct his or her utterance as soon as he detects a problem. But he may also change or expand the course of the talk in the middle of it because of a visual reaction

---

[11]In this notation, dots indicate short pauses of a light syllable (between 0.1 and 0.2 seconds), and dashes, not-so-short ones of a stress unit or foot (up to 0.4 or 0.5 seconds).

from a listener: a puzzled look, a surprised or disdainful one, etc.

In general, then, conversants try to minimize their efforts, a trend Clark and Wilkes-Gibbs (1986) called the principle of *least collaborative effort*. It is assumed to underlie the whole process of contribution and acceptance: people try to formulate the shortest utterances enabling comprehension, with a minimum of repairs as to minimize effort in understanding them. In general, the more effort put in the formulation of a contribution, the less is spent in accepting it (Clark and Schaefer 1989).

The principle also underlies turn-taking: listeners refrain from interrupting the speaker every time they do not understand something, allowing for uncertainty to creep up in the comprehension of the ongoing utterance, but trusting it will be cleared up later when the speaker finishes (Clark and Brennan 1991). Or in the minimum effort of clarificational questions (negative feedback) like "where?", "who?", in which only the relevant information is asked, and not repeated unnecessarily. The overarching priority is always to *get on* with the interaction and the topic talked about; effort is then directed to minimize clutter and 'noise' (disfluencies, etc)—including in the process of contribution and acceptance that underpins the building of common ground.

In accepting a contribution, there are various degrees to which a listener can show understanding with positive feedback. The *strongest* displays are repetion or paraphrasing of all or part of a contribution (§2.3), which shows one's appropriate understanding of what was said. Not as strong are acknowledgements like "uh huh" or "yeah". *Less* strong yet (more subtle display) are just initiation of a relevant next contribution (at as high a level as the current one), or only a continued attention without speaking, which is the weakest, least evident indication of understanding—if anything, indication of the *unawareness* of any problem (Clark and Schaefer 1987). Of course it may always mean plain disinterest or distraction instead.

Any expected or appropriate degree of acceptance (that a contribution was understood as meant) depends in general on the purpose and importance of the dialogues. Task-oriented discussions, for example, seem to require stronger evidence of understanding than casual conversations (Clark and Wilkes-Gibbs 1986). Acquaintance and relationship amongst the conversants may also influence this.

The acceptance process is recursive: each acceptance is itself a contribution, so it must be accepted as well. What prevents an endless cycle of nested acceptances is a principle stipulating that the next nested acceptance (the acceptance of an acceptance) shall be

weaker than the previous. Thus every acceptance cycle ends in a new contribution or continued attention, the weakest displays. In this way, recursion rarely goes beyond two or three acceptances, such as the following (Clark 1992, p.154), with two:

    A. F . six two
    B. F six two
    A. yes
    B. thanks very much

We can see that every subsequent acceptance is weaker than the previous: B's repetition ("F six two") is followed by an acknowledgement ("yes"), which is followed by the initiation of a new contribution ("thanks very much"). In this case, B finished the acceptance process of the first utterance, but it might also have happened that she did not talk, just kept a continued attention, with A proceeding to begin the next contribution instead (e.g. A. "yes . well, that's what I've got anyway...").

"Almost every time a speaker starts a new turn, he or she either (a) accepts what the last speaker has just said or (b) initiates a repair of the problem they ran into accepting it" (Clark 1992, p.156–7). So the utterance of a second pair-part of an adjacency-pair (Schegloff and Sacks 1973) such as a question-answer *is* an implicit acceptance of the first part, as well as fulfillment of the obligation thereby made relevant. The second part itself must then be accepted subsequently; the most common way this is done is simply to begin the next contribution on the same level. In the case of a question-answer, it is expectable that the questioner is going to make use of the information conveyed in the answer.

This leads to an important generalization: "A new contribution is initiated with every cooperative change in turns" (Clark 1992, p.164). Therefore, the *turn-taking process itself* (actually the parties' contributions) bears a *structuring function* to the discourse, other than just alternate speakers.

But full-sentence utterances in a turn are not the only type of contribution, and may not even be the most common one. Utterances may contain parts of sentences, usually single words or phrases, contributing just a piece of information. The reason may be: the least-effort tendency to satisfy just what is needed at the moment, or the speaker is uncertain and needs help from the others, or he is dividing the presentation in smaller pieces (installments) that are easier to understand (Clark and Schaefer 1987). "Generally, the more difficult it is anticipated a unit will be to understand well enough for the

current purposes, the more contributions it will be divided into" (Clark 1992, p.175). One example of this was reproduced in §2.3.2; a similar case is when communicating an address over the telephone (Clark 1992, p.167–8):

B. Banque Nationale de Liban – – –
A. yes
B. nine to thirteen.
A. sorry
B. nine . to . thirteen
A. yeah .
B. King Edward Street – –
A. yeah –
B. London .
A. ah yes

Speaker B breaks up a contribution in various small utterances intercalated with pauses as in a list, placing a rising or fall-rise intonation (a 'question' or 'exclamation' tone) on all but the last item, which gets a falling intonation. This compels the listener to give feedback by explicitly acknowledging or otherwise indicating any problem of hearing or understanding in the installments as they arise. Other instances in ordinary conversation are just the telling of a 'story' (§2.4) in installments (Clark 1992, p.169):

B. how how was the wedding –
A. oh it was it was really good, it was uh it was a lovely day
B. yes
A. and . it was a super place, . to have it . of course
B. yes

Another form of interaction is when the speaker requests confirmation with a *trial* marker, conveyed typically by means of a rising intonation (Sacks and Schegloff 1979) because of uncertainty on the information, or whether (say) a reference is expected to be relatable or understandable. The speaker would present the troublesome part with a rising intonation followed by a slight pause, so that the listener can confirm or correct it before the talk proceeds, such as (Clark 1992, p.170):

A. . . . disappeared by this time, certainly, a man called Annegra? –
B. yeah, Allegra
A. Allegra, uh replied, . . .

The question mark above indicates a rising intonation in the first utterance, where speaker A is uncertain about the name of the man he is referring. He then makes a slight pause ('–') to check if the listener recognises the name, which she does, and further corrects the reference. A then accepts the correction by repeating it and going on with the talk.

Lastly, utterances may be completed by a listener just as Duncan (1974) indicated (§2.3); for example (Clark 1992, p.171):

> A.   . . . you've got to get planning consent –
>
> B.   before you start –
>
> A.   before you start on that part, yes . . .

In this case, speaker A pauses after a syntactically-complete (but not discoursively-complete) utterance, perhaps searching for a way to better formulate what she wants to say next. Then B offers a completion, with A both repeating and completing the completion, further confirming she accepts it ("yes"). Usually the speaker presents a sentence fragment and may indicate he or she is having trouble with it. An interlocutor may then offer a completion, often with a rising intonation to indicate it is a *trial* one. The original speaker may reject, accept, or redisplay it in some other way, and the conversation continues (Clark and Wilkes-Gibbs 1986). Another case is when the conversants search for a name:

> A.   I was talking to . uh . whats-his-name –
>
> B.   John.
>
> A.   No it's . uh, Paul, I think.
>
> B.   Yeah, Paul
>
> A.   Well I was . . .

Turn-taking can thus be viewed as emerging from the participatory acts of the conversants to guarantee each contribution is added properly to their common ground. Although it does not specifically inform on a minute, moment-by-moment turn-taking process, this collaborative model may well explain the occurrence of typical conversational interchanges and subdialogues (side-sequences).

Chapter 3

The framework of simulation

To model the turn-taking systematics of small group discussion, we first need to have a simulation of such discussions that operates in the moment-to-moment basis required by process. So the basic framework wherein turn-taking can be modelled in this simulation has to be defined. And in order to do it, I will look into some of the characteristics of the process to reproduce: not just turn-taking, but the group in discussion.

Although the modelling of turn-taking was restricted to *small* group discussion, since the conversational behaviour of participants in groups bigger than seven ends up being distinct, the requirements of this framework would not be different with any limitation in size. Hence, with an eye on generality and extensibility, I will examine some of the characteristics of the general process now for the foundation of the simulation, and take the restriction on group size only for the modelling of turn-taking. The characteristics are examined in the following two sections. The framework of the simulation proper is described in section 3.3.

3.1 Group Discussion

What is group discussion then? What does it involve? What are the basic aspects that need to be reproduced by the framework of the simulation?

Group discussion consists of a group of people, more than two, that are together (e.g. sitting around a table) *engaged* in more or less informal talk about one or more topics. How informal is the talk or the interaction, and how engaged are the participants in it is

difficult to delimit. Compare a coffee-table talk among friends with a group discussion in a meeting or in a 'laboratory' recording with strangers who were told to talk about an arbitrary subject. These can all be discussions but they are likely to have very different dynamics of interaction and participant engagement.

Fundamentally, however, talk must be the only activity. In constrast, more informal or casual conversation may not have talk as the only or even the *main* activity. People in a coffee-room, or sitting around a bonfire in a camping, or at sofas in a lounge drinking tea may enter in discussions. But they may occasionally or frequently drift away from the talk to focus at nothing (i.e. just thinking and looking around) or on other activities, such as flipping pencils and making annotations, throwing wood into the bonfire or filling and passing cups of tea around while watching television.

Therefore, let us take "group discussion" to be just the verbal interaction itself, a sub-genre of conversation, stripped off of other activities, artifacts and the various settings in which it might occur that are secondary to the conversation itself. There should be no change of participants and external events either, such as someone coming in to talk to the group then going out—albeit these would not pose any difficulty or require extra provisions to be implemented *in the framework itself*. The simulation will thus only concern itself with the static group and its verbal interaction; not leisurely conversation, but engaged, focused discussion.

Another thing that will be overlooked here is the different spatial relationships of the parties in discussion: whether they are to the side or in front of each other, standing or lying down, and so on. These distinctions may affect turn-taking performance in the discussion because of 'lines of sight' and other psychological effects (Steinzor 1950, Lobb 1982). Reproducing them would require first that the simulated conversants behave accordingly to the positions of the others; but the framework of the simulation outwith the conversants would also have to *represent* spatial information somehow, so that this information should affect what the parties perceive orally and visually.

3.2 Multi-party interaction

Group discussion will therefore correspond here to a *multi-party interaction* exclusively of conversation. Parties would act and react to the talk individually and independently of each other. They do it by speaking or not speaking, and (in the future)

they could do other non-speech behaviours if these are to be represented: changes of posture, gestures, gazing around and to others, taking and dropping things, etc. They speak in utterances ranging from short responses to long and complex sentences; they can pause, hesitate, and stop speaking in the middle of an utterance when interrupted. Behaviours of each party can be simultaneous with those of the others, for a short or long time, since talk can be simultaneous in several ways, as in overlapped speaker transitions and in disputes for the floor.

The ensuing interaction should then range from situations in which everybody is silent, others in which only one party is talking, to more 'complicated' occasions of several participants speaking at the same time. This all means that the actions of the participants must be fragmented in small 'bits' of behaviour, from one moment to the next. The *collective* behaviours of the group would then emerge from these moment-to-moment *individual* 'behaviours' (acts) of the parties. Though individual, they are frequently interrelated since they can be reactions to others' behaviours and talk.

The verbal interaction would be thus coordinated not in a centralized way or by any external means as a mediator, but by the talk itself in a *distributed*, party-administered manner that is a characteristic of conversation (§2.1). It is the set of practices and constraints of turn-taking followed (variably) by each party that should ensure some sort of coordination to keep the interaction 'intelligible', within cognitive and attentive limitations (supposedly).

One way to simulate this sort of organization of individual conversants is with a *multi-agent system*. In such a system, the agents representing the conversants are individually modelled and are independently operating entities. The simplest way this can be implemented is with a synchronizing loop that would *activate* (or call) every agent in sequence collecting the behaviours they decide to do each time. Hence a complete loop would represent *one moment* of discussion. Talk is thus divided in *moment-sized* behaviours that can be simultaneous, and would form the utterances, silences and other actions in the discussion. This most simple and intuitive way to simulate parallelism is the organization implemented in this simulation, described in §3.3.

3.2.1 The communication

The predominant configuration of a group discussion is not one-to-one communication but *one-to-all*, and many-to-all some of the time. So we can generalize and assume that all talk is immediately accessible to all parties all the time. This is a simplification that leaves out the possibility of occasional one-to-one messages such as someone speaking in the ear of another person. In such a case, some spatial representation would be needed to simulate parties that are closer to others and thus would be able (or not) to hear talk that is whispered to them, for example.

But with such a close-by environment that is group discussion, specially in a *small* group, we can here assume that the behaviours of all parties are automatically accessible to all the others. The open and transparent *channel* that underlies this communication, and is its medium, will be called the *environment* of the interaction. It is through it that all the behaviours are realized at every moment and *perceived* by the others. Any *destination* (addressee) intended in the messages broadcast this way is solely a matter of individual interpretation of their contents. And whether they are recognized by each party and/or would invoke a reaction is another story.

This environment can be implemented in the multi-agent framework as a simple *blackboard architecture*. This blackboard gathers the behaviours each agent decides at every one moment and gives them to be read by all in the next moment. It also provides the *results* of the simulation: everything that happens is in the blackboard at each moment. These can be shown while the simulation runs. The other thing that is relevant for accompanying the whole process and is not there is the agents' internal states, such as to whom they are paying attention (who they think is the speaker), etc. [1].

3.2.2 Multiple modalities

As group discussion is a face-to-face activity, nonverbal behaviours such as gaze and gesture might need to be represented as well (although not in the present modelling). There might be various different nonverbal *modalities*: gaze, arm and head gestures, facial expressions, body posture, and possibly others. These behaviours involve visual

---

[1]Some internal agent states *are* indicated in a subtle way in the simulation results (cf. examples in chapter 5): for example, slightly different 'silences' for abandoned talk, pausing between utterances while the agent who has just talked still considers itself 'the speaker', and other ordinary silence.

constraints: not seeing someone means one cannot receive (perceive) his or her non-verbal behaviours as well. This would also require some sort of spatial representation for the group.

Theoretically, it should be the domain of the environment to filter out visual behaviours accessible to each party. However, it is more practical if these would be just ignored by the conversant agents *themselves* at the level of their program of operation. Visual behaviours could be *disconsidered* in certain conditions, when the gaze of the receptor is not in their general direction.

So the environment can still be a simple all-to-all channel, with individual agents having various 'filters'. For example, a *physical* visual one on top of the presumed *cognitive* filter (§4.4.1): respectively, when not gazing at a party and when not paying attention to it. This arrangement has the advantage that the 'degrees' with which simulated conversants are able to perceive others' behaviours could be more easily managed individually: for example, either fully perceiving visual behaviours when looking directly to them, or recognizing only barely but not in detail when at the 'corner of the eye'. This makes behaviour *perception* similar operationally to behaviour *interpretation*: the recognition of intention, that is, whether the behaviour is intended to a specific party and to who.

3.2.3   The environment

This concept that I called the *environment*, i.e. the blackboard containing the behaviours of each moment, warrants some consideration. We take it for granted in reality—hence what follows may seem stating the obvious—, but it is useful to spell out in the clear some of its properties now.

First, as already described, it is a channel of one-to-all and many-to-all communication. This does *not* make it appropriate to be implemented as a system that sends every behaviour individually to every participant, like a mailing list where each one receives a separate copy of the same behaviour. Rather, already established, it seems more like a news board that can be implemented with a blackboard architecture: behaviours put in the blackboard one moment are read by all the agents in the next. Therefore they need to be tagged with origin information: which agent produced which behaviour (or 'packet' of behaviours if there are various modalities).

Second, the contents of the environment are *transient* and fast moving: they only remain there for one moment. Behaviours are produced by the parties at every moment, appearing in the environment and being read by the others. *Then* they disappear, making room for new ones, or their absence, at the next moment. Examples of discrete, signal-like behaviours that could be one-moment short are listener responses such as "uh-huh" and "I see", or head nods and shakes.

But longer *continuous* actions like speaking, being silent, gesturing and gazing, may also have to be represented. These extend themselves from one moment to the next, subject to their producer's *sustained* behaviour. In this transient frame they need to be streamlined in a sequence of continuous moment-sized behaviours that are maintained actively from one moment to the next in order to be seen as forming a long act.

Talk is thus a sequence of 'talk' behaviours, deconstructed as such and reconstructed by the receivers as a continuous act. The *contents* can change but the *type* of behaviour is the same. Indeed, this sort of 'clocking' (dividing time in small equal 'slices') is possibly the only way to emulate human *analogic* behaviour and its cognition in a *digital* device like a computer (or a robot).

As in human cognition, the agents then require an internal temporal model of the continuous acts to be able to deconstruct them for transmission. Likewise, it takes the same or a similar model to reconstruct the sequence to interpret it into a longer act. In summary, the environment will *not* keep a history of the interaction for the agents to consult; the agents *themselves* must therefore keep their own records in their internal ('mental') models.

Accordingly, *inactions* that are inertial—continue naturally without any effort, such as being silent as when listening, pausing or hesitating, gazing at the same direction, and remaining in an unchanged posture, for example—would also have to be broken down into moment-sized behaviours, and generated by the agents at every moment of the simulation too. The *lack* of any action is therefore a behaviour too. Inactions in each modality can be generated as the *default* (like being silent) when an agent does not decide to do anything.

3.3 The multi-agent framework

So the group in discussion is best implemented as a multi-agent system. Conversants are *agents*, modelled and operated separately. Each agent has its own internal memory and is modelled with individual parameters relevant to the simulation. They all 'behave' simultaneously, by getting input in the behaviours of the previous moment (from the environment), deciding what to do in accordance with their own parameters and internal states, then outputting the behaviour (or behaviours in multiple modalities) chosen at that moment of simulation. Agent communication occurs only through these behaviours.

Of course, such multiple individual entities performing simultaneous actions would be best implemented in a parallel way: each agent as a separate (operating-system) process, or a thread, running simultaneously with others. This is a common conception of agents and, with a simple design, their operation would be a continuous iteration of the mentioned steps—input, decision and output—synchronized in some way with the other agents. A more complex model could aim at reproducing, for instance, cascading cognitive processes such as the conceptualization and formulation of speech (Levelt 1989), in which agents would act as interconnected sub-processes or sub-threads. More on this possibility is discussed in chapter 7.

But for the current purposes—simulating simple group discussion turn-taking—this is not really necessary. A simpler framework is sufficient, one which does not require parallel processing or multiple subprocesses, and can be implemented in a sequential manner: a 'round-robin' loop. As mentioned earlier, this is the intuitive way of simulating a parallel, distributed process: run a central loop by calling all the agents in sequence, giving an equal share of execution to each one, then take all behaviours in the full loop as being 'simultaneous'. This is what we will see next.

3.3.1 Sequential simulation

In this multi-agent framework, agents are not *active* processes running simultaneously; they are *passive*, activated by the simulation's central loop each time. This loop, a *cycle* of the simulation, corresponds to the 'clocking' iteration that the agents themselves would have to maintain internally were they running in parallel. Instead, with the sequential framework this is taken out of them, simplifying their operation a little.

Figure 3.1: Procedures for one cycle of simulation.

for each agent **A** in the group, do:

    slot **A** of `environment` ← execute agent **A**'s `cycle( blackboard )`

show (output) contents of `environment` (the behaviours of the current cycle)

for the next cycle, make `blackboard` ← `environment`

then clear `environment`.

However, because the agents will not be permanently active as in a parallel process, they have to reconsult their internal states at *every* activation to restore the various contexts they may be in at each time.

Each cycle of the simulation will then correspond to a *moment* of the discussion. How much this moment would represent in terms of 'simulated time' may vary, as seen in the next subsection. Agents are *activated* in sequence to execute just one iteration of their *program of operation*. This program of operation can be the same for all agents, which would then differ only in their internal states and their modelling in the form of a set of parameters that would characterize them differently from each other.

Behaviours collected in one cycle are considered to be simultaneous. Hence, the order in which agents are activated does not matter, because they do not depend on the behaviours of the other agents already produced in the *same* cycle, only on the behaviours of the *previous* cycle. Therefore, there must be at least two blackboards: one containing the behaviours of the previous cycle that the agents are reading, and another collecting the behaviours of the *current* cycle, which will become the 'previous' behaviours of the *next* cycle.

A blackboard can simply be an array of the behaviours (or 'packets' of behaviours) with the length of the group size. One slot for each agent, with an indication of *whose* agents are each behaviours. The framework's blackboards are then the `environment`, for collecting the behaviours of the current cycle, and `blackboard`, for holding the behaviours of the previous cycle. If the program of operation of each agent is called `cycle`, receiving as argument one blackboard containing the behaviours of a previuous cycle, then the simulation of a group discussion would correspond to an iteration of the procedures of figure 3.1.

These procedures should be repeated for as long as the simulation runs, which is decided elsewhere. In the first iteration, `blackboard` is empty. In all the others, it is

passed as input to the agents' procedures (`cycle`) with the behaviours of the previous cycle, copied from `environment` at the end of each one.

3.3.2 Cycles

In the framework, then, simulated time is divided in equal slices (the cycles) to which all behaviours are adjusted, like e.g. a (pixelated) image in a computer screen. Obviously this is a discrete, limited reproduction of the continuous flow of conversation. Therefore it is relevant to define what each cycle represents in terms of time. I called this *cycle-time* throughout the thesis, which is a parameter of the simulation and can be adjusted to different values. In the case of the present turn-taking modelling, the reasonable range it could vary is between 0.1s to 1s.

This cycle-time parameter thus determines the granularity or *resolution* of the simulation. If it is not small enough, it will not represent relevant phenomena adequately, regarding their timing. For turn-taking this means it cannot represent simultaneous starts of talk, for example, if the cycles are, say, 1 second each, because people take decisions about whether to talk or stop talking in much less of that length of time. Likewise, decisions to continue or to stop when there is multiple talk at speaker transitions may be down to tenths of a second. Hence, a value closer to 0.1 of a second would be more appropriate for this representation.

In terms of the modelling of turn-taking, a smaller or greater cycle-time is a tradeoff between simplifying behaviours and simplifying procedures. The smaller the cycle the finer and more precisely that interactions such as in speaker transitions can be represented and simulated. However, the more complex will be the procedures, having to keep more intermediary contextual states across cycles. That is because agents operate in *activations* of their program of operation: so at the beginning of each cycle they must consult some or all of their internal states to restore the contexts they were in last time around.

The smaller the cycle-time the more states that probably have to be represented. For example, if the cycle-time is small enough so that a typical nod or a feedback vocalization like "uh-huh" now spans two or more cycles, then the acting agent has to represent whether it is in a state of 'doing it' or not. If they can be in one cycle only, then no such states are needed.

Another issue is that cognitive *response-times* have to be emulated too: humans take at least a certain amount of time to perceive and react to events, and simply defining each cycle as, say, 0.1 of a second shall produce simulated discussions with conversants behaving faster than would be humanly possible ('robotically'). With a greater cycle-time value, on the other hand, the interaction would be more coarsely represented, increasing the occurrence of behaviours that cannot be distinguished in time, as in the case of simultaneous starts of talk.

For example, the first participant to speak in a TRP typically gets the turn. With a coarse granularity of simulation, simultaneous starts would have to be more finely distinguished, for example requiring 'start-of-talk' behaviours with a timestamp so that the first to have actually started could be distinguished (Padilha and Carletta 2002). In such a case, depending on the timing difference, the other simultaneous starters might be taken as (say) having had an 'intention' to speak that was preempted. If the cycle-time is small enough, however, none of this is necessary.

3.4 Focus of the simulation

The ostensive aim of this work, as pointed out in §1.3, was to create a simple simulation demonstrating some central issues of turn-taking (timings, speaker transitions, simultaneous talk, hesitation, and others). 'Simple' in that important concepts, like the practices and the states that are maintained in the moment-to-moment verbal interaction, could be easily apprehended and, thus, could be replicated elsewhere.

Another objective was to create a *distributed* simulation of interaction, tackling a problem of pure coordination by independent entities that has often been only referred to in game theory. The most direct way of creating these entities presently is with agents; indeed turn-taking seems to be a representative case of agent-based programming (Hulstijn and Vreeswijk 2003).

In that sense, this work is an improvement of previous centralized simulations such as Stasser and Taylor (1991), who also used probabilistic parameters, but only generated the speaking turn order of participants in a group discussion: i.e. A–B–A–C–A–D–E–B–C, etc. They used stable probabilities like the agent attributes in the present simulation (§4.2), and transitory ones, such as how recently someone has spoken, to determine who would speak next, but without any more fine-grained detail on how, for

instance, transitions are realized.

Regarding this characteristic: *fine-grained*—perhaps better called *interactivity*—, it is relevant to note that most current dialogue and otherwise speech-production systems have little of it. Usually we can only interact with them (and vice-versa) by pressing a key or waiting for a length of silence (Ferrer et al. 2002)—and only one interlocutor at a time (Kirchhoff and Ostendorf 2003)—, as they operate in atomic events ('utterances' or 'turns') that cannot be interrupted.[2]

They lack a level of interaction control over deciding whether to continue generating speech or to stop, whether to continue recognising the input (and emit signals of understanding or otherwise) or to barge in at an appropriate place for a smooth turn-taking. This need for a moment-to-moment operation, instead of turn-by-turn or utterance-by-utterance, led to the design of this simulation that operates in units of time (the cycles) smaller than utterances or changes of speaker (turns). Preferably, intervals as small as humans are capable of distinguishing and reacting in turn-taking, which seems to be between 0.1–0.5s (Bull and Aylett 1998, Wennerstrom and Siegel 2003).

Dabbs and Ruback (1987), who investigated the patterns of talk and silence in small groups with their 'Grouptalk model', defined a set of classes that could be measured in recordings of group discussion: individual talk, individual pause (i.e. silence), group talk, group pause and speaker-switching pauses (i.e. silent gaps). Classifying and computing the total time of just talk distinguished from silence, and whether either of these was individual or from the group, could be done automatically by their hardware, and was used in studies of group dynamics.

This pointed out that generating only the talk and silences of a discussion, apart from the turn-taking dynamics per se, could be useful by itself in particular if coupled with a way of calculating all the statistics of large simulations (as is done in chapter 6). This would not be just a simple reproduction of the surface of turn-taking phenomena, but the modelling of a higher level control (though in an abstract way) of interactive conversant agents that would represent many of human conversants' turn-taking decisions in talk.

---

[2]Providing these systems with more interactivity is the object of much current ongoing work, e.g. (Donaldson and Cohen 1996, Derriks and Willems 1998, Bell and Gustafson 2000, Cassell et al. 2001, Edlund and Nordstrand 2002).

3.4.1 Restricting the focus

Although in reviewing the characteristics of group discussion I have mentioned non-verbal behaviours and other actions of participants in conversation, the turn-taking modelling developed in this thesis was narrowed down to focus only on its structural aspects. Initial work in my simulation (Padilha and Carletta 2002, 2003) have tended to bundle things together without distinction, mixing *conceptual* behaviours like 'pTRP' to actual cues of turn-taking like gestures and raising intonation. But here I had to limit the scope of the modelling only to the structural concepts, more or less the scope of the Conversation Analysis literature of Sacks et al. (1974) and others.

Nonverbal behaviours like gestures and gaze, together with the syntax, semantics, prosody and other paralinguistic features of the talk, are *cues* that identify transition-relevance places (TRPs), and the participants' intentions to take, yield or keep the floor. This is the focus of the Psycholinguistic literature of Duncan (1972) and others (§2.6), which in the present has been abstracted away in a single 'pre-TRP' behaviour. Generating and interpreting all these cues, the syntax-semantics, intonation, and paralinguistic signals like loudness and gestures would require a more sophisticated simulation, and were left to future work extending this one. It is already the subject of much research currently (Ward and Tsukahara 2000, Ferrer et al. 2002, 2003, Edlund et al. 2005, Carlson et al. 2005).

What was also left to future work as well is how the *contents* of talk and the goals of the participants in conversation affect turn-taking: not only to convey messages, but to satisfy the obligations of the dialogue. Not only how they might affect the interaction and dialogue, but how they might interfere with the normal turn-taking that would have occurred had the contents of talk or participant goals been different. Would there be more interruption, or would intending speakers start earlier or later, perhaps indicating so with more backchannels, and so forth?

One underlying motivation (or justification) in using at first only abstract behaviours without contents was stated in an earlier paper: that "the contents [of the discussion] do not directly affect the turn-taking behaviour" (Padilha and Carletta 2002). This may not appear true when we note that what is being talked about partly determines whether a participant wants to speak at a given moment and take a turn at talk. So the contents of the talk may activate turn-taking by informing the possible attempts to talk, and thus turn *order*. But these are *not* ultimately the focus of this simulation, in modelling

participants by a set of probabilistic values.

The focus in this thesis is on the mechanics of turn-taking, through a simple simulation of conversation; not *whether* a participant attempts to talk, but what happens if he does. Not the turn order or a specific turn order, but how speaking turns (whatsoever) in real discussions are achieved and coordinated. The contents of discussion does not affect this dynamics directly. Not to mention that *what* participants are talking is not the only reason people may take turns to talk: they also do so simply to reinforce self-esteem (theirs and others), and other emotional bonds.

So the abstract behaviours without contents of talk coupled with probabilistic parameters are a simple way of simulating the process of talk that would allow modelling the turn-taking dynamics. These abstracted components can later be replaced (instantiated) with a more and more realistic emulation of talk: the cues of turn-taking (§2.6), nonverbal behaviours, the practices of taking and holding the floor, the contents of the talk, changing and keeping the topic, phrasing, and the reasons for interchanges such as in §2.7.

But in order to model turn-taking, we obviously need a conversation that provides its opportunities. The idea of using a *simulation* of group discussion, besides simplicity, has the advantage of *controlling* the conditions of the verbal interaction whereon turn-taking would occur, yet entailing a simplification of the phenomena that has to be dealt with. So the turn-taking modelling of next chapter implicitly has two parts, though they are interdependent: the turn-taking and the turn-*making* sides of the simulation. The latter is the simulation of group discussion, generating utterances in a simplified way: it is just a means to provide the model for the former.

So the important thing in this simulation is not its results, the discussions it generates, but the *ideas* behind it, the concepts underlying the procedures, internal states and the make-up of the agents that create the simulation. It is easy to see that the simple simulation here is just a beginning. There are obvious roads for extension and improvement in this symbolic system, even before one would consider manipulating actual speech production and recognition. One possible destination of these improvements could be a small group discussion simulation with conversant agents visually represented, generating real-time speech for their talk and animation for their visual behaviours. More immediate improvements are outlined in chapter 7.

3.5 Summary

Summing up, the simulation of group discussion is thus restricted:

- only the interaction of a small group is considered; the setting, events, objects and other activities are ignored;

- no distinctions of spatial position and relation in the group are represented; and

- the discussion is further simplified to talk in an interaction-control level: no contents of talk except for what is directly associated to turn-taking, and no nonverbal behaviours. So there must be explicit behaviours signalling the TRPs (in their different 'sorts' of turn-taking constraints).

The characteristics of group discussion relevant for the framework are these:

- multi-party interaction: more than two participants behaving individually and independently of each other in a moment-to-moment basis;

- engaged interaction, with exchanges ranging from short responses to long utterances and sequences thereof, with pauses and hesitations;

- coordinated by a turn-taking systematics that relies on signals present in the contents of the interaction itself, in the conversation, and not on outside means like a mediator;

- nonverbal behaviours could be represented as well, in various modalities: gaze, head and arm gestures, facial expression, body posture; these would involve restricted accessibility according to gaze;

- communication is all the time through a one-to-all channel: the environment; hence, no one-to-one messages: intended 'destinations' of messages, if any, are a matter of interpretation of the contents by individual parties.

The environment of interaction has the following characteristics:

- it is more appropriately like a news board: behaviours realized there are perceived by all parties in the next moment; they must have origin information: who produced them;

- continuous talk and other actions are streamlined into a series of congruent moment-sized behaviours put in the environment that must be reconstructed and

interpreted by the receiving parties;

- state-like inactions are also streamlined into the same 'no-action' behaviours taken as the default when the party does not decide on anything;

- behaviours are transient and fast moving: the environment does not keep a history of the interaction for the benefit of the participants;

- therefore, they must have internal temporal models of the various actions and represent the various contexts they are in with their interactional 'mental' states.

These basic aspects of group discussion and the environment make this sort of framework as more appropriate and practical for the simulation:

- the group can be a multi-agent system: each agent behaving independently and with its own set of internal states that form its temporal interactional model;

- visual restrictions of nonverbal behaviours can be more simply operationalized in the agents themselves as a kind of (involuntary) 'interpretation': in this case, blocking their perception;

- the environment is a simple blackboard architecture which holds the behaviours the agents decide at each moment to become accessible to all agents at the next moment;

- there need to be at least two blackboards then: one containing the behaviours of the previous moment which the agents are reading, and one collecting the behaviours of the current moment which will be 'previous behaviours' of the next moment; lastly,

- the contents of the blackboard at each moment correspond to the results of the simulation, which can be shown one way or another while the simulation runs.

Chapter 4

The modelling of turn-taking

In the previous chapter I described the framework for this simulation of group discussion: a simple multi-agent system for the participants of the group and a blackboard channel for their communication. It operates in *cycles* that synchronize agent action, each cycle corresponding to a fixed time indicated by a parameter, varying between 0.1s and 1s. This is the (minimal) unit of simulated time, determining the *resolution* of the simulation. It also determines the *response-time* of the agents, because the model described here has no 'buffering' of input to emulate varying delays of attention and understanding: so agents react immediately (in the next cycle) to each input.

In this chapter I will describe the turn-taking model that generates the coordination in small group discussion. This is an abstract simulation, only representing an *interactional* level of conversation on top of which actual conversation could take place in various linguistic levels (syntax-semantics, speech acts, dialogue moves, etc).

'Behaviours' are the units occurring in the cycles, and just indicate whether participants are talking or not and whether they come to the points of *transition-relevance* that lead to turn-taking and other interaction like backchannel feedback (e.g. "uh huh" or "huh?"). The content of talk is only minimally represented in the different pre-TRPs indicating sorts of turn-taking constraints; and also in the distinction between backchannel feedback from discussion talk, (other) utterances.

In the simulation, agents make decisions (i.e. when to speak, to stop, or to give feedback) based on the input from the blackboard at each cycle, their own internal states that provide them with basic interactional contexts, and their *attributes* that give likelihoods of doing various actions in the discussion, thus allowing some modelling of

different 'types of conversants'. As no talk content is produced, agents take random decisions, biased by these likelihoods. The point is not to simulate *what* they decide at each time, as that is context- and content-dependent which are all external to the turn-taking system. The point is what happens when they do, according to a system that coordinates the interaction and demonstrates actual turn-taking patterns.

The turn-taking model is defined in four components in the subsequent sections: behaviours, agent attributes, contextual states, and procedures. After them is a summary of the assumptions made in the model, and conclusions.

4.1 Behaviours

Behaviours are what the agents do at each cycle, each moment of the simulation, and what they recognise the others are doing. They are one cycle long and represent one cycle-time of discussion, for whatever value this parameter is set. They form the utterances, feedback and silences of the discussion, in an interactional level, of whether there is silence or talk and other behaviours relevant to turn-taking.

The behaviours are named with mnemonic codes used in the display of the simulation. First is the one for silence; everything else is talk of some sort:

- `-` : the agent is silent, either listening to others or making pauses when speaking. This is the default behaviour when not deciding anything.

- `talk`: the agent is talking as part of the progression and unfolding of an utterance. As one of the aims of the turn-taking system is to achieve one-at-a-time most of the time, agents are oriented to listen to those who are talking and resolve simultaneous talk as soon as possible.

- `ta-` : a disfluency that breaks the flow of talk, such as a self-interruption or repetition of one or more words; also, a voiced hesitation such as "erm", "umm" or some other non-silent interruption of talk. The agent is 'speaking' but is not (momentarily) contributing towards the progression of an utterance. Together with silence, this is *hesitation* when occurring in the middle of an utterance, and it can be a reaction to simultaneous talk.

- `TALK`: a step up into 'competitive' talk in case of simultaneous talk, corresponding to one or both of: sharp increase in loudness or change in tempo (either faster,

or slower and clearer talk, stretching syllables). Once simultaneity is over, the remaining speaker (if any) must soon readjust back to normal talk.

An utterance being constituted from the above talk behaviours in sequence is nonetheless as yet incomplete (in its syntax, semantics and/or intonation). As no contents or speech information is conveyed (and so, no syntax, semantics and intonation), utterance uptake or *possible completion* has to be indicated some other way.

Therefore one important assumption is made in this model: that the points in the talk recognised as at-the-moment possible completions of the utterance are signalled by the speaker as explicit behaviours. These should also be considered 'talk' behaviours, but they further indicate a partial or final possible completion of the utterance, to which listeners might want to react. From all they know these points *project* TRPs (transition-relevance places), hence the symbolic behaviours will be called *pre-TRPs*. The following pre-TRPs represent four sorts of turn-taking constraints from the utterance so far:

- `pTRP`: projects a *free* TRP, in which anybody is free to start to speak and take the floor. It means that the utterance so far does not pose any restrictions on turn-taking, but it is also not clear (so far) whether the speaker is continuing to talk or not. This is the most general and most common case, projecting what is generally referred as 'TRP'. The following pre-TRPs are specific cases.

- `SELECT(X)`[1]: rather than a free TRP, the utterance so far *selects* a specific party to speak next, as in e.g. an addressed question. When the utterance ends—which may not be *now*, as we will see later—, the floor is transferred to that selected party, who is then obliged to speak. Any silent interval until it speaks is heard as hesitation in his or her already-granted 'turn'. This behaviour then is not so much turn-taking, but turn-selecting.

- `Any?`: this is a variation of selecting-next in that not a specific party is selected, but in fact *anyone*, as in questions of the type "Anybody ... ?". In reality, it is actually more a variation of 'pTRP' additionally *encouraging* anyone to speak, and further implying that its producer is *not* continuing after the utterance— which may not necessarily be *now*, too, as we will see later.

---

[1]Behaviour names are defined as 4-character 'codes' for neat displaying in the simulation, as they represent short pieces of time. In line with this, `SELECT(X)` is actually *shown* in the simulation just as the 4-character name of the agent being selected: say, 'AgtA', 'AgtB', 'Anna', 'Paul', etc. Some examples are in chapter 5 and §4.1.7.

- `More`: the utterance so far conveys that *more* is to come and that the speaker is continuing to speak to realize it. The (free) TRP is 'postponed' in what is to become a multi-utterance turn, discouraging any interaction but for backchannels like "uh huh" (feedback or continuers). The notion of 'incompleteness' thus can come from the syntax, semantics and/or pragmatics of the utterance so far.

These four behaviours is what characterises this model of turn-taking: the assumption of *explicit signals* of possible utterance completion given off unambiguously *by the speaker*. As such, listeners do not identify the places of transition-relevance (TRPs) independently or variably in accordance with their own individual interests, knowledge and attention, which is a significant simplification from reality.

Lastly there is backchannel feedback. These are short utterances given *in the background* and in response to the main talk that generally function as feedback, conveying yes-or-no acknowledgement and possibly some measure of *agreement*. If there is more than that to the utterance, then it is probably agreement followed by talk in turn that carries on the discussion. Backchannels, on the other hand, do not 'carry on' the discussion, relating only to the good of the communication.

The following two behaviours are feedback as such. Since they can be of variable length, one or more of the *same* behaviour will form the whole vocalization. They can be positive or negative, with the latter characterised by an obligation on the current or recently finished speaker to address the problem raised (that is, to restart or interrupt the current talk):

- `uhuh`: positive feedback, like "uh huh", "mm", "I see", "yeah", "that's right", and so forth, acknowledging at least understanding and possibly agreement to some extent. In a multi-utterance turn, they can represent *continuer* signals that encourage the speaker to proceed with his or her talk. But the above listed responses are clearly not only or always used as feedback. They are not, for example, when answering confirmatory or yes-or-no questions: then they are normal turn talk that 'carries on' the discussion.

- `huh?`: negative feedback such as "huh?", "sorry?", "who!", "really?", calling upon the (recent) speaker to continue talking but to address the problem of hearing, understanding or non-agreement (doubt) that was raised. In a turn-taking sense one could think of them as 'selecting' the current speaker to continue talking. As such are also other reactions that feed back to the speaker and potentially

change his or her course of talk, such as exclamations like "my god!" and "no way!" that, yet apparently agreeing with the speaker, may encourage a further confirmation or elaboration on what was said. But these various other sorts of feedback will not be distinguished here.

The length of time each behaviour represents is determined by a global parameter of the simulation. Thus, all lengths in the model are defined in seconds, so that the change in that parameter affects the *number* of behaviours that would form each utterance, backchannel and silence.

This global parameter (cycle-time) determines the granularity of the simulation. The smaller the value the lengthier the simulation, and vice-versa. A good default is a setting of 0.2s, because the smallest response times that have been measured in taking turns and in perceiving interval significance are around that length. Smaller values would provide even finer (albeit lengthier) simulations, whereas greater ones would result in shorter but 'coarser' representations of discussion.

4.1.1 Mid-utterance talk

When an agent decides to speak, it begins an utterance with 'talk', having determined its length and the turn-taking *sort* which will come as the pre-TRP at the end. In a simulation without contents, this replaces the actual 'impulse' or initial concept of what a real conversant would have to say when beginning an utterance formulation. If there is no other talk, the starting agent is taken as the speaker having the floor. It continues outputting 'talk' for the length of the utterance, possibly leaving silent hesitations (-) or voiced hesitations and speech disfluencies (ta-) like self-repairs midway through (Carletta et al. 1993). For example:

$$\text{utterance}$$
$$\overbrace{\text{Agent A: talk talk ta- talk - - talk talk ...}}$$

An example of a hesitant utterance comes from Sacks et al. (1974, p.702)[2]:

J: <u>Oh</u> you know, Mittie- <u>Gor</u>don, eh- <u>Gor</u>don, Mittie's <u>hus</u>band died.

---

[2]Examples are reproduced here with the original transcription; in this case, from Sacks et al. (1974), where: punctuation indicates intonation, not grammatical function (',' for partial falling tone, '.' for final falling tone, '?' for rising tone, and '!' for rising-falling); underlining is emphasis in pitch or loudness; colons prolong syllables; dashes are self-interruption; silences are indicated with lengths in parentheses; and brackets mark simultaneous talk.

Hesitations are relevant for turn-taking because they can lead to listener interaction. And in conversation where listeners may be distracted by external events, hesitations can function to gather attention (Goodwin 1981). Backchannels like "uh huh" may be reponded as well, as a sort of automatic 'go on' incentive (continuer). And if the hesitation persists for too long, listeners may decide to speak (to *interrupt*), such as in Sacks et al. (1974, p.704):

> Claire: So then we were worse o- 'n she an' she went down four,
>
>   (0.5) but uhm (1.5)
>
> Claire: ⌈ Uh-
>
> Chloe: ⌊ Well then it was <u>her</u> fault Claire,

Although one may think intuitively that silent hesitations would lead to interruption more often than voiced hesitations or speech disfluencies (after all, the speaker is silent), this model does not assume any such distinctions yet. The main assumption it makes is that the decision to *interrupt* at hesitations (silent or voiced) is taken differently from the decision to talk at TRPs, at normal turn-taking. The interruption is made to be less likely to occur, and only after some repeated hesitation.

4.1.2 Mid-utterance TRPs

The simulated utterance given above by agent A, however, is as yet incomplete in syntax, semantics, and intonation: it is still an *unfolding* utterance. It is only complete or partially complete when a pre-TRP representing a possible utterace completion is produced. Listener agents thus wait for this signal to make any decision.

But a possible completion is no guarantee that it is indeed the end of the utterance. It is a *possible* completion: listeners cannot know yet for sure just at that moment. For all they know the speaker could be finishing there, so (in case one wants to speak) one can either decide to speak immediately in the hope of making a smooth turn transition, or wait (politely) to see if the speaker is going to stop talking. Generally, participants intending to speak attempt to take the turn as soon as possible (depending on the level of interest in the discussion) because 'first-starter gets the turn' (Sacks et al. 1974). As in this example from Sacks et al. (1974, p.721) in which Janet tries to speak at possible completions in Penny's utterance:

Penny: An' the fact is I- is- I jus' thought it was so kind of stupid
        ⌈I didn' even say anything⌈when I came ho:me.
Janet: ⌊Y-   ⌊Eh-

So this model assumes that utterances can have, or listeners recognise, *intermediary* points of possible completion besides completion at their ends. Hence pre-TRPs are generated in the middle of the utterance as well, projecting TRPs not only at the silent interval when the speaker stops after an utterance, but also in the middle of talk when he or she is *not* stopping. With the simple behaviours here, the only distinction therefore comes from what happens *afterwards*: intermediary TRPs are followed immediately by `talk` whereas the last pre-TRP is followed by silence or 'finishing talk' (§4.1.4).

Listeners can then give feedback or attempt to take a turn midway through the utterance if they do not wait for the speaker to finish. An example with intermediary feedback (the more common case) is below. A listener response to an intermediary TRP may represent reaction to a partial uptake or a *guess* of 'where the speaker is going' with it. In starting to speak in the middle of an utterance (a middle-start) and successfully overtaking the speaker (another form of interruption), one may be 'clipping redundancy', or wanting to 'shut down' the speaker (Oreström 1983), or is being plainly impatient and impolite.

```
Agent A: talk talk ta-  talk pTRP talk talk talk talk ...
Agent B: -    -    -     -    -    uhuh uhuh -    -
```

4.1.3   Simultaneous talk

If one or more participants start to speak simultaneously, or at an intermediary possible completion in the ongoing utterace, there will be simultaneous talk. It needs to be resolved by those involved: either it is terminated quickly by one or more speakers stopping soon after, or it goes longer resulting in wholly overlapped utterances.

In the first case, either a starting speaker stops in a *false-start* or the already established speaker stops short of finishing the utterance: he or she was then *cut-off*. In the following excerpt from a 5-person discussion (Fay 2000, p.192)[3], there are two episodes of simultaneous talk:

---

[3] The original transcription here is very sparse, but dashes still mean self-interruption (cut-off), and the asterisks (∗ ∗) are unrecognized syllables.

5:      What about the fact that other students probably plagiarise and
        don't get caught though
3:      ⌈Yeah it is quite important
4:      ⌊Yeah that is quite, quite high up as well because I mean-
3:       He's just been unlucky ⌈to ∗ ∗, he was unlucky that-
4:                              ⌊It was he just got unlucky there to get...

In the first, participants 3 and 4 start together, and 4 is interrupted without overlap when 3 starts to speak at an intermediary point in her utterance. Soon after, 4 restarts again at the first intermediary point in 3's utterance and cuts her off, regaining the turn by the same measure—a common pattern (Oreström 1983, Schegloff 2000).

In the case of a more protracted simultaneity leading to more overlapped utterances, usually one or more contenders try to outspeak the other by imposing a more competitive tone, 'stepping-up' their talk: speaking louder and/or faster or slower, stretching out syllables. This 'step-up' is represented in the following transcripted conversation (Schegloff 2000, p.9) with capital letters:

James: But dis [person thet <u>D</u>ID IT,
Vic: [If I see the person,
James: -IS GOT TUH BE:: hh taken care of. <u>You</u> know
         what [I mean,
Vic: [Well Ja:mes, [if I see duh person=
James: [Yeh right. e(hh) !e(hh)!

In the simulation, this is represented with '`TALK`', that expresses no intention to give up the turn, unless of course the other or others also 'step-up' their talk too. To keep a simple set of behaviours, there is no '`TA-` ' equivalent to '`ta-` '. The latter is used as well for all cases of hitches and disfluencies that occur in simultaneous talk too (such as cyclic self-interruption and repetition). A possible simulation of extended simultaneous talk could be:

```
Agent A: ...talk pTRP talk ta-  TALK TALK ta-  TALK talk ...
Agent B:    -    -   talk talk talk TALK ta-   -    -
```

Apart from backchannel behaviours, there is no other distinction between 'main' talk and 'background' talk such as *asides* or other feedback that just repeats or rephrases

part of what was said (Oreström 1983). So these are not recognised here. Any other background talk is then considered simultaneous talk and resolved in the same way. More examples are given in chapter 5.

4.1.4 End of utterance and speaker transitions

When completing an utterance, the agent outputs the pre-TRP that it has decided at the beginning. It may then stop speaking, *or* still talk for a short while before stopping. This 'tail' of *finishing talk* represents the "addition of optional elements which can specifically go after first possible completion, without intending continuation" (Sacks et al. 1974, p.707), such as terms of address and etiquette or the completion of syntactic requirements superfluous to utterance uptake. These are sources of overlap at turn transitions, for example (Sacks et al. 1974, slightly modified from p.702):
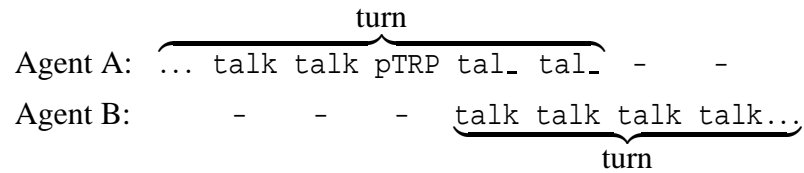
$$
\begin{array}{ll}
\text{Desk:} & \text{What is your last name} \left[\text{Lorraine.} \right. \\
\text{Caller:} & \left. \right[\text{Dinnis.(\quad)It's Dinnis.}
\end{array}
$$

Another case resulting from variation in the articulation of the projected end of the utterance is exemplified by this almost smooth transition, which is in reality overlapped by the stretching in "me" (Sacks et al. 1974, p.707):
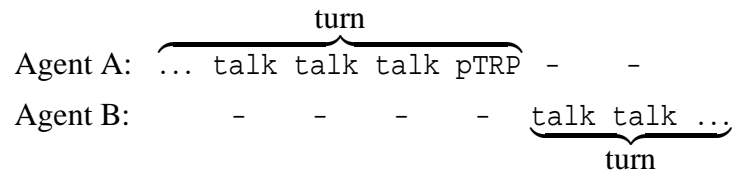
$$
\begin{array}{ll}
\text{B:} & \text{Well it wasn't me} \left[\text{::} \right. \\
\text{A:} & \left. \right[\text{No, but you know who it was.}
\end{array}
$$

Finishing talk is identifiable by the semantics, intonation and its reduced pitch and/or loudness. It is evident that listeners most of the time do not take it as 'continuing talk', so it cannot be represented with the simple behaviours thus far defined: having the same 'talk' to follow the last pre-TRP as finishing talk would make it indistinguishable from continuing talk. Then a different representaion is needed, lest a starting agent take it as such and stop its utterance beginning just in overlapping a finishing utterance.
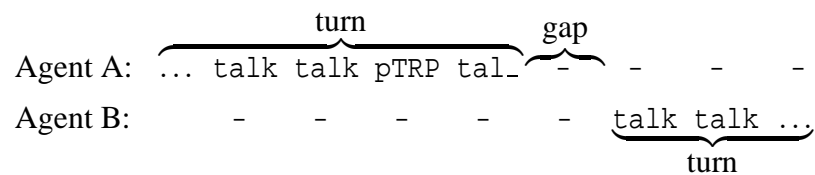
So a different 'talk' (tal_) represents the different sort of talk that is 'finishing talk', preventing any confusion. Agents thus distinguish middle and finishing talk and react appropriately: taking the former as simultaneous talk and ignoring the latter. This 'tail' of finishing talk at the end of utterances, additional talk coming after last projected completion, is assumed to vary in length between 0.1 and 0.6 of a second. An overlapped end-of-utterance transition would then look something like this:

```
                           turn
Agent A:  ... talk talk pTRP tal_ tal_  -    -
Agent B:     -    -    -  talk talk talk talk...
                                    turn
```

But not all utterances have a 'tail' of finishing talk. Many end right at the projectable completion; in the simulation, at the pre-TRP. These would not allow for turn-taking overlap, but rather for smooth transitions (*latches*) if an agent decides to speak immediately after the pre-TRP, such as:

```
                        turn
Agent A:  ... talk talk talk pTRP  -    -
Agent B:     -    -    -    -  talk talk ...
                                    turn
```

On the other hand, if the agent wanting to talk waits for the speaker to stop in order to start, there will be a one-cycle silent gap[4] (whether there is finishing talk or not):

```
                    turn            gap
Agent A:  ... talk talk pTRP tal_   -    -    -    -
Agent B:     -    -    -    -    -  talk talk ...
                                         turn
```

A similar transition occurs when the agent decides *not* to talk just after the pre-TRP, but later when the speaker has stopped (leaving a free TRP) and nobody has yet started. In this case the gap will be of two or more cycles of silence, not just one. The assumption is that such conversants would always start *later* necessarily than those who decide to speak *at* the pre-TRP but wait for the speaker to stop. They have not decided to speak on the uptake of the utterance, so they do not react as readily to its end. Actually, they do not *react* to the end of talk, but to the ensuing silence ('nobody is talking') after coming up with something to say, so it is reasonable that they must be later if starting.

It is assumed that this can engender a temporary monitoring lapse at the split-second of thinking up what to say before actually saying it. So those who decide to start because 'nobody is talking' may find themselves therein as *second starters*, starting to speak one cycle after someone else has already started. What happens afterwards is simultaneous talk, to be resolved accordingly; for example:

---

[4]Note how the cycle-time then determines the agents' response times directly.

```
Agent A:  ...talk talk pTRP   -    -    -    -    -
Agent B:     -    -    -    -   talk talk ta-   -
Agent C:     -    -    -    -    -   talk TALK TALK...
```

This is achieved in the simulation by making the internal state of 'nobody is talking' to be updated with the latest (previous cycle's) input only *after* the decision to talk in any cycle. So the state reflects the reality not of the last cycle, but the previous one, prior to the last. And that is precisely how such starting agents will always be at least one cycle late than those waiting for the speaker to stop (actually, waiting for any silence), and would only speak at least two cycles after the speaker stops. This will be clearer in the first detailed procedures (§4.4.2).

In contrast, agents that decided to speak but are waiting for the speaker to stop will not start if somebody else takes the turn first. They are waiting for any silence in order to start, so they monitor the latest input before acting. Supposedly, they have a concept of what to say already, and are just looking to start 'in the clear'.

4.1.5   TRP pauses

When finishing an utterance, the speaker decides whether it wants to continue talking if the *sort* of utterance is the general 'pTRP'. If it was 'More', of course one *is* continuing; and if it is SELECT(X) or 'Any?', one is not. If continuing, the agent starts another utterance after a variable short pause, provided nobody else starts first.

This is a *transition-relevant pause*, or simply a pause. It is assumed that the mean length of such pauses between utterances of the same speaker is a characteristic of each one: so some conversants may pause longer than others, in general. Each agent has an attribute (§4.2.2) indicating this mean length.

Procedurally, agents make a pause after each utterance irrespective of whether they decide to continue or not. As they are one-cycle reactive automata, if they did not pause when *not* continuing to talk they might decide to speak again dumbly in the *next* cycle just after having relinquished speakership. This would defeat the purpose of deciding whether to continue talking at the end of each utterance. The obligatory pause when not continuing is thus a sort of 'time off' interval to avoid immediate restarting.

Having decided to continue and while nobody has started to speak in the mean time, the agent starts another utterance after the pause, hence extending its turn at talk. This represents the situation of one finishing an utterance and deciding one has more things to say, which the just-finished utterance was just a part. The pause is then to catch one's breath and monitor uptake.
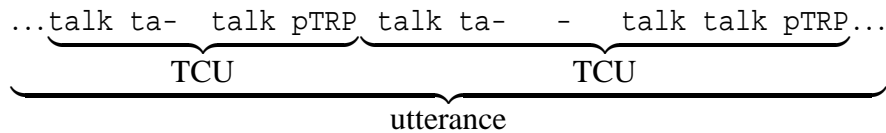
A different situation altogether (in the simulation and in reality) is when the speaker does *not* intend to continue, then stops, but nobody speaks thereafter. After the 'time off' pause, if still nobody is speaking, that agent (actually everybody) can start to talk. If the same speaker restarts, it is be because nobody spoke, which sometimes signify, like negative feedback, that the utterance was not well understood or that listeners are waiting for more to clarify. Whether or not this happens, the resulting silence is then what Sacks et al. (1974) called a *lapse*, not a pause or a silent gap.

4.1.6 Definition of utterance

Now is a good time to summarise what I am calling *utterance*. It is the *unit of talk* of the speakers in this simulation, and of the model's *turn-making* component. Various utterance definitions were discussed by Traum and Heeman (1996). Mine was also proposed or used by Nakajima and Allen (1993), Takagi and Itahashi (1996), Ferrer et al. (2002).

A speaker agent starts setting itself to produce one utterance before pausing and deciding whether to continue. Midway through it, hesitations (silent or voiced) and intermediary points of possible completion (`pTRP`) can occur. The utterance ends with a completion point signalled by one of the pre-TRPs and an optional short 'tail' of clearly finishing talk (`tal_`). It is then followed by silent behaviours representing a TRP pause. If the agent has decided to continue and nobody spoke theretofore, it starts another utterance after the pause.
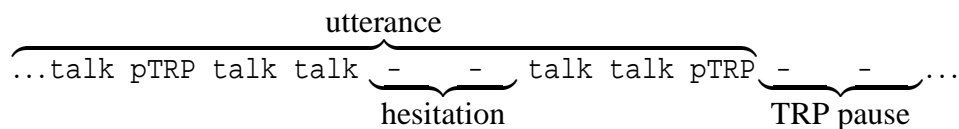
So, an utterance is a burst of uninterrupted talk, or if we consider hesitations, uninterrupted *engagement in talking*. It represents anything from a mumble to one or more words, phrases, clauses and even sentences without pause inbetween. As it may have intermediary points of possible completion, one utterance can actually contain several TRPs, and thus be formed up of several TCUs, turn-construction units (Sacks et al. 1974). In summary:

$$\underbrace{\underbrace{\text{...talk ta- \quad talk pTRP}}_{\text{TCU}} \underbrace{\text{talk ta- \quad - \quad talk talk pTRP...}}_{\text{TCU}}}_{\text{utterance}}$$

And as a speaker may continue producing utterances while nobody else speaks, a *turn* can thus be formed up of several utterances, or possibly even an incomplete one if the speaker is interrupted and cut-off midway through. An example of the first:

$$\underbrace{\underbrace{\text{...talk talk talk pTRP tal}}_{\text{utterance}} \underbrace{\text{- \quad - \quad -}}_{\text{TRP pause}} \underbrace{\text{talk talk talk ...}}_{\text{utterance}}}_{\text{turn}}$$
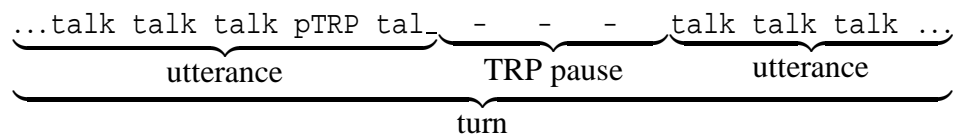
In a sense, the whole thing only gets complicated when silent hesitations are included. As they 'break up' talk, one could argue that, actually, each sequence of talk bounded by silence (even though incomplete) would be an 'utterance' if one would only distinguish talk from silence. So the point here is that I take utterance to incorporate the concept of a *completed message* followed by pause, which *can* be interspersed with silence that is, then, hesitation. And none of these silences 'in the middle' of talk—and other hesitation as well (`ta- `)—is a normal turn-taking locus, a TRP. Although they *can* lead to speaker transition, that is a different turn-taking decision, with a different probability, to which I am referring to as (a type of) *interruption*.

$$\text{...talk pTRP talk talk} \overbrace{\underbrace{\text{- \quad - \quad}}_{\text{hesitation}} \text{talk talk pTRP}}^{\text{utterance}} \underbrace{\text{- \quad -}}_{\text{TRP pause}} \text{...}$$

4.1.7 Pre-TRPs

The four pre-TRP behaviours signal (project) the possible completion of four sorts of utterance *with regards to turn-taking*. That is, in consequence of the contents of the utterance so far, four constraints emerge to subsequent turn-taking—to the overall interaction actually—as shown in table 4.1. How the agents make the different decisions to talk in these cases is described later (§4.4.5). Here they are only outlined. In reality, a wider and finer range of obligations from the infinitely variable contents and contexts of talk may probably exist, but hopefully these four types are representative generalizations.

Table 4.1: The four pre-TRPs: constraints to turn-taking.

`pTRP`: allow turn-taking	`SELECT(X)`: disallow turn-taking[5]
`More`: discourage turn-taking	`Any?`: encourage turn-taking

The first type (`pTRP`) is the normal, free turn-taking. Most of the intermediary points of possible completion in the utterances are of this type too (with exception for what is described in the next subsection). It projects TRPs without constraint as to who is allowed to speak. The only constraint then is who actually *wants* to speak and whether it is appropriate or polite to do so at that point. That some possible completions would actually be stronger points of turn-taking than others—because of subtle differences in semantics, intonation and paralanguistic features of the talk—cannot represented with the simple behaviours here.

The second type (`SELECT(X)`) represents the completion of utterances constructed with a 'select-next' technique as described in chapter 2. These can involve for example a question whose addressee is indicated by gaze, or explicit addressing such as in Sacks et al. (1974, slightly modified from p.717):

> Sharon: Oscar did you work for somebody before?
>
> Oscar: (0.5) Yeh, many many. (3.0) Canned Heat for a year.

Here the selected party responds after a brief pause, and pauses even more before a second utterance. The turn-taking would be like this in the simulation (without matching lengths precisely):

```
                    turn              hesitation
Agent A:   ...  talk talk AgtC  -       -    -
Agent C:         -    -     -      -     -   talk talk ...
                                               turn
```

Although a party is selected and so in theory that party is 'powerfully constrained' to speak, it actually may not end up speaking at all after being selected. If it hesitates for too long, others can start to speak and take the turn instead. But I am assuming here that *at the first moment* after the completion of the utterance, the floor and 'obligation' to speak are with that selected party. Only if he or she soon does not start talking

---

[5]Except for the party thus selected.

coherently (without hesitation) is that others may barge in—although they may do it anyway if for some reason they want to speak before the other does.

This might seem the same kind of interruption as when the speaker is hesitating midway through an utterance—after all, in both cases the speaker is hesitating—, but in fact it is not. One is not in the 'middle of talk' here, in the middle of an incomplete utterance. Instead, a party was selected but has not actually started speaking: it is hesitating. So the selecting-of-next is quite alright a TRP whose *immediate* turn-taking is not free. As it turns out that the selected party hesitates, it then *becomes like* a TRP in which the speaker is pausing between utterances, as he or she has not yet started.

The decision to speak in such a case is then the same as in pauses between utterances, *once it is recognized* the selected party is hesitating. This is covered in the procedures for hesitation and interruption (§4.4.6). Of course, non-selected participants in reality may not have anything to say, as it might be something that only the selected party would know or would be able to answer adequately anyway, but that is not something the present simple simulation without contents of talk can comprise.

The third type (Any?) is similar to selection-of-next, except that no specific party is targeted, like in questions-to-all of the type "Has anybody seen that movie?". This type of utterance can only occur in multi-party talk (with more than two participants) wherein a distinction is possible between *one* other and 'the others'. It can elicit a flood of simultaneous talk, or what Edelsky (1981) described as an apparently collaborative *free-for-all* type of floor, such as (p.386):

> Rafe: OK, let's talk about Tuesday.
> Len: ⌈ Well-
> Carole: ⌊ OK, Tuesday-
> Sally: As long as we're out by four...

In the simulation, it would be something like this:

```
Agent A: ...talk talk Any?  -     -     -
Agent B:     -     -    -    talk talk pTRP
Agent C:     -     -    -    talk talk TALK...
```

As with selection-of-next, it does not necessarily entail that everyone will always start to speak forthwith; in the simulation, it is more *likely* that someone (anyone) will. So

the decision to talk has to be different than the normal turn-taking (pTRP). But as decisions are based on probabilities, it may still happen that nobody speaks immediately though, resulting in a *lapse* of silence in the discussion.

Note finally that TRPs of the sort of 'Any?' and SELECT(X) cannot be followed in normal circumstances by positive feedback. As the speaker is explicitly *relinquishing* the floor, continuer signals would not make sense then (though as a show of understanding or agreement they eventually might).

4.1.8 Not stopping when yielding the floor

The two turn-yielding pre-TRPs (SELECT(X) and 'Any?') entail that the speaker is going to stop speaking at some point. But it may not perforce be *now*, at that pre-TRP. Rather, the speaker may continue to talk for a while without pause, usually to justify or clarify the request (§2.2.2). Examples, the first from Power and Martello (1986, p.34):

> Have you seen Mike, Jane? I want him to help me move the piano.
> Anybody seen my hat? I couldn't find it anywhere.

The second sentences in each of these utterances if spoken out of their contexts would not probably be turn-yielding. But they come after a turn-yielding possible completion of the utterance[6] in which the speaker did not stop and allow the selected party, or anybody, to talk. It is as if the turn-yielding TRP is herewith *postponed* onwards, until the speaker finally stops to allow the turn transition to take place.

But this only occurs if the subsequent turn-unit or units (TCUs) in the utterance are 'neutral' in that regard, so that they get overriden by the postponed turn-yielding. Because, although unusually, they could further change the utterance's intention and turn-taking constraints completely, as in these (admitedly contrived) examples:

> Have you seen Mike, Jane? Ignore him because blah blah blah . . .
> Anybody seen my hat? I don't want it anymore, and blah blah . . .

In any case of course, the affected agents can still decide to speak right at that TRP without waiting to see if the speaker will stop. What happens thenceforth is simultaneous talk to be resolved as usual. But if they decide to wait instead, they could face a different prospect later on (if the turn-yielding mode is reversed).

---

[6]Note there are no pauses between the sentences, so they form *one* utterance.

Long utterances such as those above may actually be constructed in multiple units, each one formulated on its own. But when and while the speaker has more things to say from the 'initial concept' of the utterance, he or she may go over to the next unit immediately without pausing to discourage any potential overtaking.

In this model, the turn-yielding sort is determined 'for the utterance' at its beginning, and output at its end. Intermediary pre-TRPs springing up in the middle, if any, are meant to represent unintentional possible completion points that listeners may recognize, and are normally '`pTRP`'. At any of these points, however, the agent may decide to *reveal* the turn-yielding sort in advance, to emulate the examples above. This would not preclude it from being output at the end too as normal, to represent the neutral last completion point that is then overriden, to represent the first example pair in this subsection. It would be like this:

$$\overbrace{\text{mid-point turn-yielding utterance}}$$

Agent A: `... talk talk talk AgtB talk talk talk talk AgtB`

And whenever the turn-yielding pre-TRP is output in advance, it can be changed (in a fixed low probability) to reflect the kinds of 'changes of heart' of the second example pair above. In the simulation:

$$\overbrace{\text{mid-point turn-yielding unit only}}$$

Agent A: `... talk talk talk AgtB talk talk talk talk pTRP`

These cases in this simple model, as with the TRPs themselves, are signalled unambiguously by the speaker and as such are not subject to listener (variable) interpretation.

4.1.9 More-to-come

The last type of pre-TRP (`More`) represents a variety of cases of which those shown here are but a part. It signals TRPs where turn-taking is discouraged because the utterance so far indicates that *more* is to come. So the speaker is continuing to talk to convey it. One possibility involves grammatically incomplete utterances, lacking for example the direct object or the main part of a subordinated clause. Selting (2000) gives examples of *if-then*, *when-then* and other constructions such as this (p.506):

> Ida: ...either (0.7) live at home and work for a car
> Nat: ⌈ mhm.
> Ida: ⌊ or live here and work for a flat.

Another subordinated construction broken up in two utterances could be:

> A: Although I agree that the process should continue
>
> B: yeah.
>
> A: (0.5) I think that ...

Which in the simulation would be represented something like this:

```
Agent A: ...talk talk More  -    -   talk talk...
Agent B:    -    -    -    uhuh  -    -     -
```

Again, it does not follow necessarily that no one would ever start to speak and interrupt more talk to come by the speaker. Merely that such utterances implying more-to-come would *discourage* turn-taking, as it is evident the speaker is continuing to talk. So it is a matter of politeness versus one's eagerness to talk. The decision whether to speak is then one of *interruption* in the same way as in hesitation; in both cases it is clear the speaker has the floor and is *engaged* in talking. Therefore, the two are determined the same way in the simulation.

Another possibility is when the speaker is explicitly projecting or proposing some form of extended talk in a multi-unit turn by devices such as 'first of all,' (Schegloff 1982) or other list-initiating methods:

> A: There are three things you have to know. (0.5)
>
> B: mm.
>
> A: One is ...

Or story prefaces (Sacks 1974) in which the whole utterance is devoted on doing the projecting of more-to-come, of multiple utterances in an extended turn:

> A: You won't believe what happened to me yesterday!
>
> B: What.
>
> A: I was walking in the park as always. ⌈ Then ...
>
> B: ⌊ mm hm.

This model aims to represent only the local immediate management of turn-taking: only what happens *next* at each possible completion point. The latter examples, however, evidence the existence of higher levels of turn-allocation (actually, discourse) obligations characterized by specific types of dialogue exchanges and conventionalized

sequences. These involve not only the *next* turn, but further turns as being projected or 'appointed' afterwards as well, in a higher-level structuring of the discourse.

Cases like story prefaces or preliminaries such as 'Can I ask you a question' (Schegloff 1980), instead of projecting a single (possibly multi-utterance) turn, entail a multi-*turn* sequence in which a listener or listeners are allowed turns to accept or reject the 'proposal' with the understanding that the floor will come back to the requester for continuation or other replying. Coming from other systems outside the turn-taking system (i.e. discourse and dialogue moves or exchanges), they expand the local turn-order bias of A-B-A unto A-[any or all of the others taking turns to reply]-A in groups. Though it affects speaking order, this is out of the scope of this model to systematize.

4.1.10 Backchannel feedback

A listening agent only decides to give backchannel feedback when *not* wanting to talk at a pre-TRP, or when deciding to wait for the speaker to finish. And then only if it is appropriate: positive feedback after turn-yielding pre-TRPs (`SELECT(X)` or `Any?`), as argued earlier, is not. Giving feedback is thus a strong sign that one does not want to talk, since the best opportunity for it is being skipped. Though one could still start at the next cycle after the feedback, if there is silence, one risks losing the opportunity to whoever may decide to speak immediately at the TRP.

If giving feedback, the agent makes two further decisions: whether to *delay* its vocalization, and whether it will be positive or negative. The former is decided on the inverse likelihood of giving feedback in the first place, *only* if the agent does not want to talk. The assumption is that the more likely someone is of giving feedback, the more likely it will be prompt and not delayed. And if the agent wants to talk and is waiting for the speaker to stop, a possible feedback there must be prompt. The latter decision is (randomly) biased by a fixed probability: for example, negative feedback on 10% of the time (the default). This is taken from a global parameter in the simulation, as it is not something relevant and characteristic enough to be a feature, an *attribute*, of each participant (see next section).

'Delaying' feedback means actually starting to output it around an assumed range of 0.2s to 0.8s later (maybe too narrow a range). This simulates what, it is also assumed, happens in reality: that feedback does not always come right at the point of uptake

that triggers it, but lags depending on the listener's response-time of attention, interest, understanding of the utterance, and so forth—otherwise, all feedback would be unissonous. So, without 'buffering' input to carry the TRP state over to subsequent agent activations (cycles) in order to represent the participants' differing response-times, a variable delay is simulated instead. For example:

```
Agent A: talk talk pTRP talk talk talk talk pTRP  -   ...
Agent B:  -    -    -    -   uhuh uhuh  -    -     -
Agent C:  -    -    -   uhuh  -    -    -    -     -
```

Moreover, the backchannel is assumed to vary in length between 0.2s and 0.6s, so it is one or more of 'uhuh' or 'huh?' in sequence that forms the whole vocalization.

Regarding their *results*, positive feedback is for now generated only for the sake of simulating it, because the agents do not 'interpret' them in any way. Negative feedback, on the other hand, has to be responded to by the appropriate speaker. This feedback is much like a short select-next utterance that happens to select the very speaker to whom it is obviously a reaction, as Sacks et al. (1974) pointed out (§2.1.2).

But therein lies a difficulty for a representation of talk that does not involve meaning and visual clues (such as gaze) to identify *to whom* a "huh?" is 'obviously' intended, aggravated by the possibility that it may be delayed. More than one agent may be speaking at the same time: there is no way then to identify to whom it is intended. The speaker may or may not continue to talk after that pre-TRP. Others may have started to speak then. And finally it is only at the *end* of the vocalization, some cycles later, that the speaker of that earlier pre-TRP has to react to it.

Most of these problems are dealt with in the model by everybody registering who is the speaker at every pre-TRP. This is later remembered when the backchannel is completed (§4.4.3). Of course, a problem still remains that multiple speakers will mark themselves as speaking and would later all react to the negative feedback. But this cannot be solved without contents of talk and agents capable of inferring intention, to whom it was intended. Lastly, in case the identified agent is already speaking (i.e. continuing), it hesitates briefly before starting a new utterance in response to the feedback (so we can recognise it as well):

```
Agent A: ...talk pTRP talk talk talk talk  -   talk talk...
Agent B:    -    -     -   huh? huh?  -    -    -
```

Example from Sacks et al. (1974, p.708):

> Roger: Are you just agreeing because you feel ⎡ you wanna uh-
>
> Jim: ⎣ Hm?
>
> Roger: You just agreeing?

4.2 Attributes

Decisions in conversation such as whether to talk at a certain point, continue talking or to give feedback are external to the system of turn-taking. They depend on the contents of the conversation and personal reasons of the participants at the various points in the interaction. As such, they are associated with the higher-level 'cognitive systems' such as the language and dialogue processing, planning of goals and agendas, emotional states, and so forth. These systems make use of the low-level turn-taking systematics as a means to coordinate the interaction, the exchange of talk.

For example, agreeing or disagreeing with what was said, or having more to say about the topic, are clear reasons for deciding to talk, or to continue talking at a certain point. Psychologically, decisions also involve other variables related *to whom* one is talking: acquaintanceship, liking and empathy, status, gender, and age of the interactants and so forth (Dabbs and Ruback 1987). Although it is arguable that these do have an effect on turn-taking (e.g. on the EAGERNESS parameter)—and I do not deny them, indeed I am acknowledging their interdependency—, they nonetheless are way outside of the scope of the turn-taking system.

As such, they fall out of the scope of this model as well. It was intended to generalize the turn-taking systematics and issues closely related to it so as not to depend on specific contents of talk and real participants. The previous section has precisely described the behaviours that represent this interactional (turn-taking) level.

In the same vein, conversants in the form of agents of the simulation are modelled in a simpler interactional profile. The agents are defined by a set of *attributes* that give parameters individual to each agent. These parameters are probabilities of their making the various interactional (turn-taking related) decisions in the simulation. Decisions are thus random, but biased by the probability contained in the appropriate attribute of each agent, which can then be different from the others.

This concept has two advantages in a strictly analytical model. First, it simulates

interactants' decisions in a variable way. Even if agents have the same probability in an attribute, they are not going to make the same decisions all the time. They will have a *tendency* for the same decision on average in the long run, but it would still be reasonably variable. Second, the attributes provide a way of outlining the *interactive profile* of participants and groups in discussion, which might be useful for investigators of group process and their dynamics.

For example, one attribute, TALKATIVENESS, gives the probability of an agent wanting to talk at each TRP. With it one can model a participant that is highly talkative (a 'leader') or less talkative, perhaps less involved or interested in the discussion. The whole group can be made more or less participative, with individuals standing out as either. Individual and group profiles can be simulated in this way, and probabilities could be taken from the statistics of a real discussion to reproduce its dynamics.

Probabilities are real numbers in the range [0,1]: for example, a value of 0.3 for an agent's TALKATIVENESS means that it would want to talk 30% of the time at each TRP. The greater the value the more likely that the agent will make the decision each time and the more frequently that it will be made overall. Only one attribute does not hold probabilities, but a time length in seconds: TRPAUSING, the mean length of TRP pauses (between utterances).

Some attributes relate to listener decisions, or decisions associated with *turn-taking*. Others relate to speaking characteristics, therefore to the *turn-making* simulation. Table 4.2 summarizes the attributes of the model.

4.2.1 Turn-taking attributes

Attributes associated with listening decisions that shape the turn-taking process:

- TALKATIVENESS: likelihood of wanting to talk at a free TRP (`pTRP`). In the case of 'Any?', 'More' or at hesitations, this is modified by EAGERNESS.

- EAGERNESS: a general measure of *how much* one wants to talk at any given moment (when decided to talk). It is the likelihood of starting to speak forthwith after the pre-TRP, instead of 'politely' waiting for the speaker to stop. It is also part of the decision to talk when interrupting at hesitations or after 'More', and used as an *incentive* to talk after 'Any?'.

Table 4.2: Turn-taking and turn-making attributes.

TALKATIVENESS	likelihood of wanting to talk
EAGERNESS	likelihood of interrupting and speaking immediately
FEEDBACK	likelihood of giving feedback (positive or negative)
CONFIDENCE	likelihood of persisting to speak in simultaneous talk
VERBOSITY	likelihood of continuing to talk (and having 'More')
TRPAUSING	the mean length of pauses between utterances
INTERACTIVITY	likelihood of making intermediary pre-TRPs
HESITATION	likelihood of hesitating ('ta- ' or ' - ')
SELECTIVITY	likelihood of making select-next utterances

- FEEDBACK: likelihood of giving backchannel feedback, either positive or negative, and the likelihood that, if giving feedback, it will be prompt rather than delayed (so that the more one is inclined in giving feedback the more likely it will be prompt). In previous work (Padilha and Carletta 2002, 2003), the attribute was called 'Transparency'—changed to make its purpose hopefully more, err, transparent. The likelihood of everyone deciding to give negative instead of positive feedback comes from a global parameter (NOTUNDERSTAND) as mentioned previously: it is not something in general intrinsically 'characteristic' of each participant but contingent on a discussion, I suppose.

- CONFIDENCE: likelihood of persisting to speak when simultaneously with others, directed to restore one-at-a-time talk by the individual agent decisions of stopping or continuing to speak individually. It represents both a measure of confidence in one's own talk and (conversely) attention to the others while speaking, whether recognising or ignoring that others are speaking too. Also, it is both a turn-taking decision (*who* ends up with the turn) and a turn-making one (how the utterance goes thereupon). Lastly, the various procedures described in §4.4.4 make use of this attribute in related decisions: whether to 'step-up' talk (TALK) or to hesitate (ta-) in simultaneous talk.

Let us take a brief look at how these are used. At normal, free TRPs—i.e. after 'pTRP' and silent intervals that follow it: pauses, not hesitations—, listeners decide whether they want to talk. Each agent's TALKATIVENESS gives the likelihood of doing so: if greater than a generated random number between [0,1], the decision is 'yes'.[7]

If wanting to talk, the agent then decides whether it wants to talk so badly that it will not wait for the speaker to stop: the likelihood of doing so comes from EAGERNESS. If deciding not to speak immediately, the still intending-to-speak agent waits for an upcoming silence in case the pre-TRP is really the last one of the utterance—possibly after a 'tail' of finishing talk as previously described (§4.1.4).

However, those waiting for the speaker to stop may face the possibility that that pre-TRP was not the end of the talk after all, but an intermediary TRP. If such is the case, the agent has then missed its opportunity to speak because it will 'forget' the previous intention at the next pre-TRP, deciding anew whether it wants to talk. Decisions to maintain an intention-to-speak across pre-TRPs—the turn-construction units (TCUs) that constitute an utterance—is not taken here. It would require balancing the importance to the listener of each of these TCUs—its contents—to the changing focus of attention during listening (keep intending to point that thing out, or move on?).

After 'More' or when the speaker hesitates continuously for some time, the decision to talk (to interrupt) is given by the *product* of the agent's TALKATIVENESS and EAGERNESS. This is a way to represent decision to talk at 'less ratified' points of possible turn-taking: as probabilities are numbers smaller than one, the result is then a fraction of the agent's normal TALKATIVENESS according to its EAGERNESS. If deciding to talk at an hesitation, an agent will start immediately of course. In the case of 'More', which only occurs at the end of utterances, the agent decides as normal whether to speak immediately or wait for the upcoming silence.

If the pre-TRP is 'Any?', agents should be more willing to talk than normal, so the likelihood of the decision is the *sum* of the agent's TALKATIVENESS and EAGERNESS. This is a generic representation of how utterances of the type 'Anybody...?' would increase everyone's likelihood of talking. It may seem too generic to reflect any fundamental underlying relationship, but the point of these different *sorts* of pre-TRPs (which is the model's only incursion into the contents of talk) is to recognize the *fact*

---

[7]Decisions involving likelihoods, which are most of the decisions in this simulation, are made in this manner. I will refer to it as 'testing' an attribute (or a combination thereof), particularly in the procedures: e.g. 'test TALKATIVENESS to decide to talk'.

that there are different sorts of turn-taking constraints and that they affect turn-taking differently. Having decided to speak, the agent also decides on EAGERNESS whether to start immediately or wait for the speaker to stop, as 'Any?' can come in the middle of the utterance as well, if 'anticipated' (§4.1.8).

Finally, if one agent is selected (`SELECT(X)`), initially only that agent will 'want' to talk. No TALKATIVENESS decision is made. The agent just decides on its EAGERNESS whether to start immediately or wait for the selecting-speaker to stop—again remembering that `SELECT(X)` can come in the middle of the utterance if 'anticipated'. But when starting to speak, the selected agent further decides whether to pause or hesitate instead: since it was granted the floor, it does not have to guarantee by way of immediate talk. If it then hesitate too much without continuous 'talk', others may then interrupt by starting to speak.

So, speaking goes in a two-step process here: first it is decided whether an agent *wants* to talk, then whether it *starts* immediately or waits for silence. If its EAGERNESS is 1, it is as if the decision was the simple one-step test of TALKATIVENESS because the agent will always start promptly. With EAGERNESS as 0, the agent would always wait for the speaker to stop; it would then take less turns at talk than intended because it is always in disadvantage in the 'competition to talk' against more eager others.

If, however, the agent does *not* want to talk at a pre-TRP, it decides whether to give backchannel feedback through its FEEDBACK attribute. If the pre-TRP is 'Any?' or `SELECT(X)`, only negative feedback is appropriate.

4.2.2 Turn-making attributes

Attributes associated with speaking decisions that affect the making of the turns:

- VERBOSITY: likelihood of wanting to continue to talk at the end of utterances of the 'pTRP' sort. When the speaker decides to continue, it will then pause, (supposedly) to take breath and monitor reaction before starting another utterance. If deciding to continue and nobody has started to speak during the intervening pause, the agent begins a new utterance. As the floor is already established there, it may also hesitate before starting: exactly the same situation (and consequences) as when one is selected to speak. This attribute also determines whether an utterance will be of the 'More' sort when starting one; if so, the agent *will* be

wanting to continue at its end.

- TRPAUSING: not a likelihood, but the mean length of pauses between utterances (TRP pauses), allowing 'characteristic' pausing by different speakers. Even when the speaker has decided *not* to continue talking it pauses as sort of a 'time-off from talking' to prevent the agent deciding to speak just after having decided not to, when out of the speakership state (§4.1.5). Actual pauses should vary in length around that mean, from at least one cycle (of cycle-time length) up to possibly the double of it.

- INTERACTIVITY: likelihood of having intermediary pre-TRPs roughly at each two seconds of talk (an assumed interval). Also, the likelihood of having a 'tail' of finishing talk at the ends of the utterances. So this is a measure of the 'opportunities for interaction' given off while one is speaking. The more interactive the agent, the more likely that at roughly each 2 seconds of talk (varying in each utterance) intermediary pre-TRPs will occur. This means that utterances are more (or less) frequently formed up of recognisable turn-construction units (TCUs) without pause inbetween, rather than longer units.

- HESITATION: likelihood of hesitating in the middle of the utterance and at TRPs when having the floor already: either when selected to speak or after a TRP pause when at the point of beginning another utterance. The decision of whether to hesitate with ' - ' (silence) instead of 'ta- ' (voiced hesitation or self-interruption) is made with another 'test' on the same attribute: the assumption here is that the more likely one is of hesitating, the more likely it will be silently (and longer). Note that, in a sense, hesitation is additional 'interactivity', so this is related to the previous attribute.

- SELECTIVITY: likelihood of making a select-next utterance, which means deciding a SELECT(X) pre-TRP; the agent to be selected is chosen at random. This is a less relevant parameter for simulating group discussion, not least because it can change the *genre* of the verbal interaction being created if the likelihoods or one or more agents are high. If so, the resulting interaction would resemble less free group discussion (which usually has little or no selecting-next utterances) and more like a meeting with participants delegating turns to others.

When starting to speak, the agent first determines if the sort of the utterance will be SELECT(X), based on its SELECTIVITY attribute. If not, then it decides for 'Any?'—

only if the group has more than two participants—based on the likelihood given by a global parameter (named ASKANYBODY). As with choosing negative over positive feedback, the frequency of utterances of type 'Anybody. . . ?' seems hardly something intrinsic to each participant, and rather contingent on the discussion. This is what justifies it being a global parameter instead of an attribute of each agent.

When not choosing 'Any?', then the agent decides for 'More' based on half the likelihood of VERBOSITY. Otherwise, the sort of the utterance will be 'pTRP'. The assumption made here is that the more verbose one is, deciding more often to continue to speak, the more likely one's utterances will be of the 'More' sort, but in half the extent of other neutral utterances, which are certainly more common than it. So, for example, if an agent's VERBOSITY is 0.6, its 'More' utterances will occur 30% of time and 'pTRP' 70% of time (excluding the portion of turn-yielding utterances). If its SELECTIVITY is 0.1 and the global likelihood of 'Any?' also 0.1, then 'More' utterances will occur 24% of the whole time, and 'pTRP' 56% of time (also see figure 6.30 in the evaluation chapter, §6.4).

When starting to speak, the *length* of the utterance is also determined. Two possibilities were considered: a normal distribution with a mean length for each agent, or simple random lengths up to a certain maximum. The first method requires each agent to have another attribute with the mean length of the utterances it produces (say, UTTERANCES[8]). But it has the advantage of modelling their utterance lengths in a statistically meaningful way. One can then take the average utterance lengths of participants in a real discussion as input parameters. And a normal distribution with a high standard deviation could ensure reasonably 'realistic' variation.

If this is not required, however, the second method of just setting lengths randomly up to a maximum may be equally valid. It is simpler in that it would require only a global parameter (say, MAXUTTERANCES) with the maximum length of everybody's utterances as a characteristic *of the discussion*. What justifies it is that utterance lengths in a group discussion really cannot seem to be characterizable by a normal distribution, even with high standard deviation. They vary so much depending on the contents of talk and the various moments and topics of the conversation, independently of participants, that it does not seem meaningful to model them on the basis of these participants alone, without considering what they are talking about and why: the context, their knowledge and choice of what to say, syntax used, and so on.

---

[8]In previous work (Padilha and Carletta 2002, 2003), this was the INTERACTIVITY attribute.

Table 4.3: Global parameters of the simulation.

CYCLETIME	the simulated time of each cycle, in seconds
NOTUNDERSTAND	likelihood of choosing negative feedback
ASKANYBODY	likelihood of making 'Any?' utteraces
MAXUTTERANCES	maximum possible length of all utterances (if generating random utterance lenghts)

Moreover, does utterance length matter as far as a turn-taking system is concerned? The answer seems to be only insofar as longer utterances increase the possibility of more 'points of interaction': more intermediary pre-TRPs and hesitation. Speaking of which, note that the utterance's length set at the beginning is only the 'intended' talk, as the actual utterance may well be extended or cut short by other factors. Hesitations and, possibly, episodes of simultaneous talk can extend it much beyond that initial length; and the speaker can be cut-off short of completing the utterance by others starting to speak at intermediary TRPs.

4.2.3   Global parameters

For the sake of completeness, the global parameters of the model that have been mentioned in passing throughout the last subsections are listed here in table 4.3. They are parameters of the *simulation* and of the discussion, as opposed to the attributes which model individual agents.

A few other mean lengths and probabilities which are *assumed* by this model are incorporated directly into the procedures (§4.4.2 onwards): eg. the mean interval of possible intermediary pre-TRPs, the mean lengths of finishing talk, backchannel feedback and its delays, the mean length of hesitation when having the floor and starting to speak, and how long to wait before interrupt someone hesitating, and others . These are supposed to reflect general speaker characteristics of (arguably) one culture, not individual participants or discussions.

Table 4.4: Agent variables and interactional states.

speaker	the agent is speaking; listening to no one; listening to X
feedback	it is the middle of delaying or giving feedback; or not
huhfor	the speaker at each pre-TRP: to whom 'huh?' is intended?
wannaTalk	(when listening) waiting for silence to start to speak
	(when speaking: pausing) decided to continue to talk
hesitate	(when listening) count continuous hesitation to interrupt
	(when starting) count hesitation when having the floor
sort	(when listening and speaking) the *sort* of pre-TRP
length	starting to speak; middle of utterance; finishing; pausing
midTRPs	interval of possible intermediary pre-TRPs in this utterace
tail	length of the 'tail' of finishing talk (between 0.1–0.6s)
pause	length of the pause after an utterance (from TRPAUSING)

4.3 Interactional states

There are a number of *interactional states* that participants in any talk-in-interaction go through. These are associated with the various contexts and roles acted out during the interaction: whether one is talking, listening, giving feedback, pausing, hesitating, etc. Procedurally, agents in the simulation record these states in their own internal *variables*, which are retained (and 'recovered') across activations at every cycle. These variables, and the interactional states they record, are summarized in table 4.4.

Notwithstanding their strong procedural bias (prefacing the procedures of next section), the relevance of (most) the interactional states should not be underrated. No matter how a system is implemented they will be present in one way or another—or at least those related to listening and turn-taking; those associated with speaking (the lower half of the table) relate to specific assumptions of this model: its structure of the utterance separated by pauses, the 'tail' of finishing talk, etc. In more complex conversational systems that would actually have speech processing and understanding, these states would be, not explicitly, but implicitly underlying the high-level control or

decision modules of the system. The present model only bares them out here.

4.3.1 Listening states

The primary state (or role) in any verbal interaction is of course speakership. A participant at any moment is either a *speaker* or a *listener*. The latter has also been variously called 'hearer' or 'auditor' (Duncan 1974). I use 'listener' because it more clearly entails that one is involved in the equal-status discussion, and not just auditing what is being said or overhearing it while doing something else. Even when nobody speaks, participants are still actively 'listening' (monitoring) for signs of the possible next speaker or to be able to speak without clashing with another starter. At least in the restricted scope of this simulation (§3.4), there are no overhearers, so agents are all equal participants, either a speaker or listener at any moment.

A listener agent further registers internally to *whom*, if anyone, it is paying attention. It is plausible that a participant has this internalized state and gives more focus to that one speaker than others, as opposed to monitoring all input equally at every moment to determine anew to whom to pay attention. Listeners do this anyway, but it is likely the case that attention already directed to one speaker put other sensory input in second place. In this model, input *is* scanned every cycle, but to check if anybody talks *besides* the currently focused speaker (or itself if speaking). The *speaker* variable just registers this speaker, or a null value (0, as it is an integer) indicating no one. If it is set to the agent itself, then that agent is speaking.

This concept of directed attention leads to the concept of *having the floor*. Collectively, a speaker has the floor when having the attention of the others: the more listeners, the more clearly and surely that one has the floor. Speakership is an individual state, but the floor is a collective, emergent state. When more than one agent starts speaking, it is possible that some listeners pay attention to one and some to another. In this case, there is not one clearly defined floor anymore until a sole speaker emerges again and listeners put their attention on that one, which will then have the floor. Note that the state of 'no one speaking' allows to distinguish it from a momentary silence (a pause or hesitation) in one's talk, who may still have the floor, still considered as 'the speaker'.

When listening, an agent may also decide to give backchannel feedback. As it can be longer than one cycle, a variable is required here to count this length, and thereby

record whether the agent is giving feedback or not. The *feedback* variable is set to the length of the feedback and decremented every cycle; so when it is 0, the listener is *not* giving feedback. Delayed feedback (§4.1.10) is realized the same way: first delay, then when it reaches 0, start the backchannel proper, thus simulating the parties' varying response-times. Another variable, *huhfor*, registers who is the speaker at every pre-TRP so everyone identifies to whom a possible negative feedback is intended when it finishes several cycles later (and it can indeed be *much* later, since feedback may be delayed).

A listener's boolean variable *wannaTalk* records if it has decided to talk; if not immediately, then the agent is waiting for the speaker to stop, for any silence. When speaking, *wannaTalk* marks whether the agent has decided to continue talking in another utterance after the pause. Notice then that the decision whether to continue is taken *before* the speaker pauses, at the last pre-TRP. Why not after the pause?

The TRP pauses are meant to be pauses between utterances of the same speaker, and very short in general (less than a second), just the time to take breath and monitor reaction. If one does not have something planned to say when finishing an utterance, he or she will take longer than that to speak again. Moreover, speakers intending to continue do employ techniques to retain the floor (as in §2.4 and §2.6) from the finishing talk onwards to discourage potential turn-claimers. So it is reasonable to think the decision to continue in such cases is taken *before* pausing, when finishing an utterance.

Listener agents use another variable, *hesitate*, to count the amount of continuous hesitation so they know when to decide whether to interrupt it: on a certain threshold, assumed to be 0.8s. The same counter is used when a starting speaker already having the floor—when it was selected to speak or is pausing between utterances—decides to hesitate before talking. The agent then counts a short (random) amount of hesitation, silent or voiced, before actually doing 'talk'.

With silent hesitation in the picture, listeners must be able to contextually distinguish it from TRP pauses. Their *sort* variables record the sort of pre-TRP the speaker just output, but not only for guiding the subsequent turn-taking decisions. Until someone starts to 'talk' again (not 'tal␣'), listeners must know they are in a TRP with the specific constrains of the indicated sort. For example, if someone was selected to speak, silence thereafter should be taken as part of that participant's turn, not that 'no

one is speaking' so anyone could start. If it is 'More', then listeners are discouraged to speak until the current speaker resumes talk. Hence, decisions for all the following cycles after the pre-TRP are restrained by *sort*. And when someone finally goes on with 'talk', *sort* must be cleared so that any subsequent silence after 'talk' is taken as hesitation, not a TRP pause.

4.3.2   Speaking states

A speaker in this simulation is in one of the following states at any time: starting an utterance, middle of the utterance, final pre-TRP, tail of finishing talk, or pausing in a TRP. The *length* variable indicates this; it is set to the utterance length when it is started and decremented at every 'talk' and intermediary pre-TRP. When it reaches 0 the speaker outputs the last pre-TRP—also defined at the beginning and kept in the *sort* variable. Then *length* keeps indicating that the (bulk of the) utterance has ended while *tail* and *pause* are respectively set to the (possible) tail of finishing talk and the subsequent pause. Thence until the decreasing *tail* reaches 0 the speaker is finishing talk; then until *pause* reaches 0 the speaker is pausing.

When starting an utterance the agent decides its length, *sort*, and, if already having the floor, whether to hesitate. It also defines in another variable (*midTRPs*) the interval with which potential intermediary TRPs will appear in the utterance: a random value around two seconds. It is at these intervals throughout the utterance that the agent will 'test' its INTERACTIVITY attribute in order to have intermediary pre-TRPs.

In the next cycle just after starting, the agent will first check if it is speaking simultaneously with others. A phase of 'starting talk' could be distinguished from 'middle of talk' if starting speakers are to have less confidence in continuing to talk simultaneously than those who are already in the middle of talk (such as with a reduction in their CONFIDENCE), since they would be then second starters or interrupters. Or if a speaker is continuing with another utterance after a pause, and finds itself simultaneously with others, it should likely be less 'confident' to persist talking simultaneously, and instead yield the turn to the new speaker(s). These states all have to be indicated by other variables described in §4.4.4.

Figure 4.1: Basic agent procedures.

If I am speaking (internal state):
   Talk for the length of an utterance with possible intermediary TRPs,
   but if someone talks too, decide whether to stop: if so, listen to that one.
   At the end of the utterance, signal a TRP and decide whether to continue,
   then 'tail' any finishing talk, then pause for a certain time:
   if anyone starts to speak then, switch into listening to that one.
   When finished pausing, if I decided to continue: start a new utterance;
   else, set no one as speaking and switch into listening.
Else, I am listening:
   If anyone is talking or starts to talk, listen to that one;
   else if I decided to talk earlier: start a new utterance;
   else, set no one as speaking then decide whether to talk or listen.
   When the speaker reaches a TRP (indicated by a pre-TRP behaviour),
       decide whether to talk: if so, decide to start immediately or wait.

Starting an utterance: determine the utterance length
              and the interval of possible intermediary TRPs,
              then set myself as the speaker, and talk.

4.4 Procedures

All the agents follow the same procedures at each cycle. It is their attributes and state variables (detailed in the previous two sections) which gives them different interactional profiles and transient states. Figure 4.1 presents a general description of the basic procedures. A more detailed pseudo-code specifying variables and attributes appears in §4.4.2, followed by extensions to backchannel feedback in §4.4.3, simultaneous talk resolution in §4.4.4, sorts of TRPs in §4.4.5, and hesitations in §4.4.6.

The basic procedures of figure 4.1 describe what is to be accomplished through several activations in subsequent cycles, not just in one go: read the 'then' therein thus. The process is serialized across the cycles via the intermediary states held in the variables. These are consulted at each activation so that agents make contextually appropriate decisions (with their attributes). The procedures are then very much like a decision tree where the variables, the input from the previous cycle and the attributes drive the agent to ever more specific branches (contexts) down to the final 'leaf' wherefrom the behaviour for the cycle is determined. There is no iteration: one path only on the tree is traversed every time.

The procedures are naturally divided in two parts: when the agent is speaking and when it is not. The speaking part (*turn-making*) goes according to the speaking unit being modelled: an utterance formed of 'talk' with occasional intermediary pre-TRPs ending in a final one, followed by a possible 'tail' of finishing talk and a pause; in the latter, input is monitored and the speaking state is exited in the case of any talk. The listening part (*turn-taking*) involves 'listening' to a speaker, if any, changing attention to another one, deciding whether to talk at TRPs or when nobody is talking, and setting the utterance up when starting one.

In that part, when an agent decides to talk when nobody is talking, the decision is taken in one cycle but the agent will actually *start* in the next one—again, only if nobody then starts.[9] This means that agents who have already decided to talk previously—for example, at a TRP but are waiting for the speaker to stop—have an advantage over those who decide to speak because nobody is. The latter would take a further cycle of monitoring before actually starting to speak. It simulates a participant who decides to talk in a moment, but has to forgo it in the next because someone else has then started (which is sometimes visible in a mouth opening and audible inhalation).

It should be noted that the procedures above can be simplified at the end of the speaking part: the last three lines of pausing-monitoring replicate, and thus can be merged with, the three beginning lines of the listening procedures. It is possible this is more a descriptive redundancy than a reflection of any real identity of cognitive states between pausing-monitoring and just listening, but here it goes.

If anyone starts talking when the speaker is pausing, the speaker can immediately 'fall through' to the listening part and follow the procedures there, which are the same. Likewise, when the speaker is not continuing to talk it can just fall through to the listening part because, again, the procedure is the same. And we can equate the decision to continue to talk at the end of an utterance with, when listening, the decision to talk but first wait for the speaker to stop, both held and triggered by the *wannaTalk* variable (§4.3.1). It only remains that the speaker should continue to pause so long as no one talks. Otherwise, or when the pause ends, execution just goes through to the listening procedures, which eliminates the last three lines of the speaking part. This simplification is incorporated in the detailed procedures that follow, in §4.4.2, after a description of the input processing, next.

---

[9] The 'decide whether to talk or listen' in the listening part of figure 4.1 does not really imply *start* to talk, which only happens in 'start a new utterance'.

4.4.1   Input processing: a 'cognitive filter'

The first thing agents need to do at the beginning of each cycle is to identify if anyone is talking in the array of input from the previous cycle. This is assumed to be an 'automatic' process that takes place whether the agent is speaking or listening. It represents our monitoring 'sort-out' mechanism (a 'cognitive filter') that occurs in parallel and irrespective of the interactional states participants are in, supposedly.

In regard to it, this model does not simulate any variation of attention or in the cognitive capacity of agents. In reality, it may be that attention focused elsewhere or a cognitive 'load' (when initially formulating an utterance, for instance) would prevent or hinder this sort-out of sensory input, affecting one's awareness of others talking, for example.

At the beginning of each cycle the agent scans the array of inputs from the *other* participants (excluding itself) for non-feedback talk behaviour: i.e. 'talk', 'ta- ', 'TALK', 'tal␣' and pre-TRPs. If more than one party is found to be talking, it picks one randomly that is not the current *speaker*, the one being paid attention to. This assumes that agents always recognise when *others* start to talk while listening to a speaker— which may not be the case all the time in reality, as attention fluctuates. The end result is a single internal variable (*talked* in the subsequent procedures) indicating a now-speaking participant that may be the same current speaker, if nobody else besides it is talking, or another participant beyond it (including when the agent *itself* is the current speaker).

A couple of considerations about this. First, backchannel feedback is *automatically* sorted out from talk behaviour by virtue of being represented differently. This is of course a simplification that comes along with the discretization of behaviours in this model, since in reality any vocalization is a continuous acoustic signal interpreted at various linguistic levels until it is finally recognised (possibly only when it ends) as one thing or another. But this mechanism of cognitive recognition is not aimed in the model here, which justifies this 'automatic' distinction as an abstraction of that process. Recognition and reaction to negative feedback (positive feedback is only generated) is included at this stage of input processing (see §4.4.3).

Second, the scanning for possible talkers excludes the agent itself and whoever the agent is paying attention to. If the agent is speaking, then anyone else can be identified as talkers too, so that the agent can decide whether to stop because of the simultaneity. If, however, the agent is listening to a then-current speaker, then anyone *else* who talks

must be singled out so the agent can decide whether to switch attention to that one. As seen in the following procedures, attention is always switched to *any new speaker* without other consideration—which is a simplification of reality, as already mentioned, since listeners may be more or less 'immersed' in the current talk to do it or not.

4.4.2 Detailed basic procedures

Pseudo-code detailing the basic procedures is in figure 4.2. They show how the input, state variables (in *italics*) and attributes (in CAPITALS) determine the resulting behaviours at each cycle. These procedures are the result of extensive testing with the implemented model (appendix 1) that evolved from previous work (Padilha and Carletta 2002). This is the source and justification for all the detailed decisions taken in this and most of the next subsections, that have been described right from the beginning of this chapter. In the whole, the try to represent the patterns and possibilities alluded in all the reviewed work on turn-taking (chapter 2), a synthesis of most of it.

Read all "say xxxx" as 'return behaviour xxxx and exit this activation', and all "test ATTRIBUTE" as a test of whether a randomly-generated number between 0 and 1 is smaller than the attribute's likelihood (is random-number < ATTRIBUTE?). If it is, the test is successful; else it fails. Finally, "around Xs" means generating a value within a random normal distribution around the mean X in seconds; this is of course translated to a number of cycles according to the CYCLETIME parameter.

As just described, the array of possible input is scanned at the beginning of every activation, resulting in a single 'summary' pointing out whether anyone *talked* in the previous cycle, and who. This variable in turn will affect the *speaker*, unless the current speaker chooses to ignore simultaneous talk or a listener decides to talk (becoming the speaker itself).

When speaking, anyone else having *talked* triggers the agent into deciding whether to persist speaking simultaneously or not. This excludes 'finishing talk' (tal_) in order to prevent that an otherwise valid overlapping turn-taking is stopped by the finishing speaker being recognised as 'talking', as if it were in the middle of the utterance (§4.1.4). In reality, it is cues like pitch and loudness of the talk plus its contents (and visual behaviours in face-to-face) that distinguish utterance finishing from its middle; here, simply abstracted in different behaviours.

Figure 4.2: Basic procedures detailing state variables and attributes.

Search input for anyone except me who had any 'talk' or a pre-TRP in the last cycle:
  set *talked* to that agent; if there is more than one, choose not the *speaker*.

If I am *speaker*:
  while (decreasing) utterance *length* > 0:
      if anyone *talked* (except for 'tal‿'), test CONFIDENCE,
        then stop if failing it: set *speaker* to *talked* and say ' - ' (cut-off)
      at *midTRPs* intervals, test INTERACTIVITY to say 'pTRP'
      otherwise, say 'talk'
  when *length* $= 0$: (only once)
      set *wannaTalk* to a test of VERBOSITY in deciding whether to continue
      set *tail* (finishing talk) to around 0.4s, or 0 if failed INTERACTIVITY
      set *pause* (length) to around TRPAUSING
      say 'pTRP'
  while (decreasing) *tail* > 0, say 'tal‿'
  while (decreasing) *pause* > 0 and nobody *talked*, say ' - ' (pause)

(else, I am listening:)
  if nobody *talked* and I *wannaTalk*: StartTalk
  if nobody is *speaker*, test TALKATIVENESS to StartTalk
  set *speaker* to *talked*
  if *speaker* (if any) reached a TRP (signalled for now only by 'pTRP'):
      set *wannaTalk* to a test of TALKATIVENESS
      if I *wannaTalk*, test EAGERNESS to start now: if so, StartTalk
  otherwise, say ' - ' (silence).

StartTalk: set *length* to a random number of cycles up to a maximum or a mean
        set *midTRPs* (interval of middle pre-TRPs) around 2 seconds
        set *speaker* to myself, then say 'talk'.

While speaking, *length*, *midTRPs*, *tail* and *pause* regulate the ongoing utterance. The first two are defined at the start of it; the last two at its end. The aim in grouping these definitions together, with lengths set in advance, was implicity. In reality, actual utterance length comes about from within the process of talk itself, not in a pre-planned way, with hesitations possibly extending it further. Another simplification is that the length of the upcoming pause is set together with the 'finishing' tail of talk, if any.

Setting the utterance length right at its outset is a concession to the clarity of the model. If we were to generate actual language as in reality, the agent would instead have at the outset just a mental proposition of what it wants to say, or a topic or concept to be elaborated that is coherent with what has just been said. This proposition or topic or concept would then be developed to more specific components and translated into syntactic structures and then words in the course of the next cycles, like in a series of parallel cascading processes (Levelt 1989). Failure to carry on this process smoothly in the given slices of time would yield the various hesitations, filled pauses, self-repairs, or otherwise fragmentation that is common in talk.

As mentioned elsewhere (§4.1.5), the speaker 'pauses' even when *not* continuing to talk. That is a 'time off' from talking, or else the agent could decide to speak again immediately afterwards, which would defeat the purpose of 'not continuing'—an operational way to emulate the short-term fact of having talked just recently. Because of the condition in the last line of the speaking part in figure 4.2—that the speaker continues pausing only if nobody *talked*—, execution should 'fall through' directly to the listening part if otherwise (so these two parts are not mutually exclusive, it should be clear). There, the agent would set *speaker* to the 'other at talk' to effectively disengage from speakership.

The boolean state *wannaTalk* is set to true when the speaker decides to continue to talk after a pause, or when a listener decides to talk either when no one is (so it starts at the next cycle if the situation persists) or at a TRP but it is waiting for the next silent opportunity. In any of these cases, the agent will only start if nobody is speaking. A pausing 'speaker' who has decided to continue will go through the listening part when the pause ends in much the same way as if somebody else started to speak. On the other hand, a listener who wants to talk at the next silent opportunity may never get to it if the current speaker does not finish talking but reaches another TRP instead. In this case, listeners just make *new* decisions and 'forget' the previous ones—they are assumed to have short 'memories'.

Note that deciding to talk when no one is speaking already is based on *speaker*, not *talked*; and *speaker* is updated only after this decision. This means that agents may begin to speak even if someone has already started in the last cycle. The decision is based on a state of 'nobody is talking' that existed in the cycle *prior to the last one*. Therefore, second-starts by one cycle are possible.[10] This gives agents who already decided to talk at a TRP an advantage over those who decide to talk in the following silence: the first are quicker to speak.

Finally, it is useful to concentrate the procedures for starting talk in one sub-routine, called StartTalk. Not only because it is needed in three situations (more than three, later), but because it is where speaking practices such as hesitation and pre-starter other beginnings like "well," (Sacks et al. 1974) could be added. The utterance length is then determined by a random normal distribution from mean lengths held in agent attributes, or—if one considers that utterance length is to all purposes arbitrary, as I implemented in appendix 1 and evaluated in chapter 6—random lengths up to a certain maximum taken from a global parameter.

4.4.3 Feedback

Procedures for positive and negative feedback are in **bold** face in figure 4.3 to distinguish them from the previous ones. Besides the more relevant *feedback* and *huhfor* variables described in §4.3.1, one other state (of *occurringHuh*) is needed to mark whether negative feedback is occurring, so that the appropriate agent (*huhfor*) reacts to it when it ends. This is activated in the first bold line of the procedures, dealing with feedback recognition.

When the negative feedback ends, the agent who identifies itself as its target will respond to it by starting to talk. If it is already (or still) talking, it first makes a silent break of one cycle before beginning a new utterance, which is the self-interruption that characterises a response to the feedback (as described in §4.1.10).

Next, everybody's *huhfor* is updated to the current *speaker* when it outputs a pre-TRP: thenceforth the supposed target of any negative feedback. When such a backchannel occurs and some cycles later ends, the agent who identify itself in *huhfor* will respond to it. It is possible as a result that more than one, or the 'wrong' one, would do as

---

[10]Which again connects the length of each cycle to the agents' response-time that is being simulated here, a dependency that can make their reaction to be 'super fast' if that length is too short.

Figure 4.3: Procedures with backchannel feedback.

Search input for anyone (except me) who had any 'talk' or a pre-TRP in the last cycle:
  set *talked* to that agent; if there is more than one, choose not the *speaker*.
Mark if anyone except me gave negative feedback (`huh?`).
When the marked feedback is over (i.e. no 'huh?' anymore):
  if *huhfor* = me (last one to 'emit' a pre-TRP), the feedback is for me:
      StartTalk (interrupt talk with ' – ' if I am talking now).
If *speaker* had a pre-TRP: set *huhfor* with *speaker*

If I am *speaker*:
  while (decreasing) utterance *length* > 0:
      ⋮
  when *length* = 0: (only once)
      ⋮
  while (decreasing) *tail* > 0, say '`tal_`'
  while (decreasing) *pause* > 0 and nobody *talked*, say ' – ' (pause)

While (decreasing) *feedback* > 0:
  continue the backchannel or its delay: repeat the last behaviour I made
When *feedback* = 0 and I was delaying it (with silence): GiveFeedback.

(else, I am listening:)
  if nobody *talked* and I *wannaTalk*: StartTalk
  if nobody is *speaker*, test TALKATIVENESS to StartTalk
  set *speaker* to *talked*
  if *speaker* (if any) reached a TRP (here only signalled by '`pTRP`'):
      set *wannaTalk* to a test of TALKATIVENESS
      if I *wannaTalk*, test EAGERNESS to start now: if so, StartTalk
      test FEEDBACK to give feedback:
      if so, test again (or if *wannaTalk*) to start now: GiveFeedback,
        or else delay it for around 0.5s (just set in *feedback*)
  otherwise, say ' – ' (silence).

StartTalk: set *length* to a random number of cycles up to a maximum or a mean
          set *midTRPs* (interval of middle pre-TRPs) around 2 seconds
          set *speaker* to myself, then say '`talk`'.
GiveFeedback: set *feedback* around 0.4s, then say '`uhuh`' or '`huh?`'
                          (according to NOTUNDERSTAND).

such. As backchannels can last for some cycles and on top of that can be delayed, other agents might speak in the mean time, and output other pre-TRPs. So this simple mechanism is not entirely failsafe—but none will be unless agents are able to infer intention or recognize visual clues as who someone is addressing.

Agents always try to answer negative feedback in this model, so problems of hearing or understanding are given a higher priority over other talk that carries on the discussion (Sacks et al. 1974, p.720), at least from the party they are (supposedly) directed. Whether the involved speaker continues afterwards in the face of any simultaneous talk is left to their normal decisions of CONFIDENCE. It is possible therefore that the speaker responding to a negative feedback give up talking soon in the face of simultaneous talk, or continue and cut-off the others.[11]

From the listeners' side, deciding to give feedback (positive or negative) only occurs if the agent does *not* start at a TRP, whether wanting to talk or not. The actual backchannel can be started immediately at this TRP in a second test of FEEDBACK, *or* it can be delayed by around half a second (from 0.2s to 0.8s)—only if the agent does not want to talk. The assumption here is that backchannels are not always prompt, but bear on the varying delays of understanding and attention of the listeners. And that the more likely someone is of giving feedback the more likely it will be prompt.

The actual realization of feedback goes in the middle of figure 4.3 between speaking and listening. The same *feedback* counts down the states of 'backchannelling' or 'delaying' while the behaviour of the previous cycle is copied over: feedback ('uhuh', 'huh?') or silence (' - '). The 'copying' of the behaviour is just an operational simplification to do away with another internal variable: for the states of 'delaying' or 'vocalizing' each type of backchannel.

When delaying, *feedback* is set to the length of the delay and the agent returns silence as normal: it is then copied throughout the delay, and the feedback proper starts after it. To realize it, the counter is set to a length (around 0.4 seconds, or varying between 0.2 and 0.6 seconds) and the type is decided: positive, or negative according to the parameter NOTUNDERSTAND. The behaviour is output and again copied throughout the count of *feedback*.

---

[11]A marker identifying the recent (possibly second-) starter as a respondent of negative feedback could be added to the balance of simultaneous talk resolution procedures (next subsection) to weigh one down prioritarily, since these procedures normally favour first-starters.

Figure 4.4: Two-step simultaneous talk resolution.

⋮

If I am *speaker*:
  while (decreasing) utterance *length* > 0:
        if anyone *talked* (except for 'tal⎵'), test CONFIDENCE,
          if failed but I last 'talk'ed: just say 'ta- '
          if failed but I last hesitated (ta-), stop talking:
            set *speaker* to *talked* and say ' - '
        at *midTRPs* intervals, test INTERACTIVITY to say 'pTRP'
        otherwise, say 'talk'

⋮

4.4.4 Simultaneous talk

Resolution of simultaneous talk in the previous procedures is very simple: just decide
whether to stop or continue every time someone else talks. This results much too
frequently and much too soon in both speakers quitting together and no one continuing
afterwards (see also the evaluation results in §6.3).

But speakers in simultaneous talk in fact do more than just ignore the other(s) for
a while and quit. Various 'resources' are employed, both consciously and uncon-
sciously, to deal with non-backchannel simultaneous talk (Schegloff 2000). Although
seemingly arbitrary behaviour like both speakers quitting at once cannot be wholly
eliminated in this simulation with random probabilistic decisions, the incorporation of
such resources shall furnish agents with bolder and more 'realistic' behaviour.

In particular, agents could take longer when deciding whether to stop in order to figure
out if the other talk is not just a short aside or complement. One frequent consequence
or 'resource' to deal with a simultaneity are 'hitches' in the talk, as reviewed in §2.5:
e.g. hesitation, self-interruption, and repeated stop-restart-stop. Another is to step-up
to 'competitive' talk to mark a stronger stance of wanting to keep the turn: speaking
louder, changing tempo, speaking syllables more clearly, etc.

Initially, simultaneous talk could be resolved in a two-step process as shown in figure
4.4: first hesitate if failing a test of CONFIDENCE, then if failing again and the agent is
already hesitating, quit talking. If succeeding any test, the speaker continues as normal.
Agents would thereby respond to simultaneous talk mostly with a few hitches—though
not repeatedly. Yet this still yields poor results (in testing the implemented version) as

more often than not both agents decide on the same thing: to hitch then quit, or the same cycle of hitch-restart-hitch-restart and so on.

So here are seven procedures that can be incorporated (independently or together) in the resolution of simultaneous talk, as outlined in figure 4.5 (in **bold**):

1. It is reasonable to assume that *second-starters*—i.e. those who start to speak slightly after someone already has—, and also those who are already the speaker continuing after a pause, would have less 'confidence' in persisting talking than the others (§2.1.4), who started first or are trying to take a turn from the current speaker. Or at least, they would have a 'penalty' to their normal CONFIDENCE. The first two and the last two bold lines in figure 4.5 encode this with the states *firstStarter* and *continuing*. At StartTalk, a second-starter would have *talked*, when different than *speaker*, indicating that someone has already started (the *speaker* would be finishing talk). If continuing, the agent would *be* the speaker already. But second-starters ought not to be penalized, or rather made more confident instead, if it is a response to negative feedback (which would require different StartTalk 'modes'): this is one situation cited by Sacks et al. (1974) where second-starters can get the turn.[12]

2. Only worry about the simultaneous talk if what the speaker has (remaining) to say is relatively 'long': e.g. longer than 1 second. If *length* is equal or less than this, the agent needs not bother about being simultaneous; what it is saying is perhaps an aside, a complement or accompaniment to the other talk, intended or unconcerned with the simultaneity, *or* it was a longer utterance that is now nearing completion whereat the agent will just not quit now.

   Admittedly, the justification for this is mine own, from observation in a real group discussion (though a 'laboratory' one, not casual) that short utterances or when at the end of one, are simply not left abandoned because of simultaneous talk: it seems like in a 'final stage' of utterance production, when it is fully formulated, speakers would just *finish* it. This then partly determines the evaluation in 6.3, such as the reduction in the number of incomplete utterances and false-starts (figure 6.29). It is also in line with criticisms to the supposed one-at-a-time

---

[12]Alternatively, first- or non-continuing starters could have their confidences boosted instead; or starting speakers could have varying 'confidences' individually for each attempt to talk, according to the circumstances: interest, now or never, etc. In the end I did not implement this procedure in the programmed version (appendix 1, model 2) because it is .

$\vdots$

Figure 4.5: More procedures for simultaneous talk resolution.

$\vdots$

If I am *speaker*:

   while (decreasing) utterance *length* > 0:

        if anyone *talked* (except for 'tal_'):

           if there was a *firstStarter*, decrease CONFIDENCE **(e.g. by 20%)**

             or if I am *continuing* after a pause, decrease it too (by 20%)

           count the extent of simultaneous talk so far (*simultalk*)

           if *length* is 'long' (i.e. what is left to say >1s):

             test CONFIDENCE **to continue: if fail, say 'ta- ' (a hitch)**

             else, test again to continue: if fail, say 'TALK' (step-up talk)

             (or else continue as normal down below with 'talk' or a pre-TRP)

           BUT if I already hesitated (ta-), test CONFIDENCE **to stop,**

             or stop immediately if the other said 'TALK':

             set *wannaTalk* to a test of TALKATIVENESS **(retake at the next opp.)**

             set *speaker* to *talked* and say ' - '

           OR if I said 'TALK' already: continue with 'TALK', but

             if the other said 'TALK' too, test CONFIDENCE **to say 'ta- '**

        when no one *talked* (=0) and *simultalk* was 'long' (>1s):

           repeat overlapped talk: increase *length* and say 'ta- '

        when no one *talked* (=0):

           clear both *simultalk* and *firstStarter* (set to 0)

        at *midTRPs* intervals, test INTERACTIVITY to say 'pTRP'

        otherwise, say 'talk'

   $\vdots$

$\vdots$

StartTalk: **save *talked* in *firstStarter* if $\neq$ *speaker* (not with negative feedback)**

          mark *continuing* if I am *speaker* already

          set *length* to a random number of cycles up to a maximum or a mean

          set *midTRPs* (interval of middle pre-TRPs) around 2 seconds

          set *speaker* to myself, then say 'talk'.

'perscription' of talk, as discussed in §2.2 and §2.2.1 (Edelsky 1981, O'Connell et al. 1990).

3. If failing a first CONFIDENCE test, the agent hesitates (`ta-`)—and does not count the hitch as part of the utterance: i.e. do not decrement *length*, or actually re-increment it (not shown in figure 4.5). If succeeding, test CONFIDENCE again to ignore the simultaneity and continue talking as normal (`talk`), or, if failing, to 'step-up' talk (`TALK`). Ignoring the simultaneity is the strongest stance (Schegloff 2000), after succeeding two tests (see §2.5.4). Stepping-up talk, by succeeding first and failing second, means that one is confident enough not to be affected by the other talk (no 'hitch'), but not just to ignore it. Anyhow, the aim is to get three outcomes from a decision test that gives only two.

4. BUT (and this is a capital BUT because it should actually precede the previous procedure in real code) if already hesitating, the agent should quit talking if failing a first test—the previous two-step process incorporated here—, *or* quit immediately if the other said '`TALK`' (has already stepped-up talk). If not failing this test, the agent follows the other procedures as normal: thus it can hesitate repeatedly as in the actual patterns of hitch-restart-hitches, by failing in one cycle then succeeding in the next (or next few ones). Or the agent might hesitate, then step-up to competitive talk or resume the normal talk (see §2.5.2).

5. If quitting, test TALKATIVENESS to decide whether to 'retake' the turn at the next opportunity (next TRP), by setting *wannaTalk* accordingly. This is a common strategy (Oreström 1983, Schegloff 2000); see §2.5.5.

6. But after stepping-up to '`TALK`', the agent continues until the other speaker(s) quit or step-up talk too: in the latter case, they hesitate if failing a further test of CONFIDENCE. Once others quit talking and leave alone just one speaker who has stepped-up talk, this one will naturally readjust itself back to normal talk (by the normal non-simultaneous procedure); see §2.5.6.

7. Finally, when getting out of the simultaneity this way, still retaining speakership, if the whole episode was relatively 'long' (say, using the same threshold as in the second procedure above: longer than 1 second), the agent should try to 'recycle' part of the talk obscured by the simultaneity, as seen in §2.5.6. In the simulation, the agent will then just increase *length* by the length of the simultaneity and

interrupt itself with 'ta- ' to indicate this.[13] The extent of the simultaneity needs to be counted in order to determine whether it was a 'long' one, which is done with the variable *simultalk*.

Several other practices, 'resources' or courses of action in and after simultaneous talk were described by Schegloff (2000) and Oreström (1983) that cannot be represented or do not make sense in this simple simulation without contents of talk. Just to cite two: addressing the other overlapped talk just afterwards, try to 'shut someone down' or clip redundancy.

4.4.5 Sorts of utterance

Procedures for the specific *sorts* of utterance are shown in figure 4.6. The extensions are again in **bold** over the basic code of figure 4.2. The previous extensions of feedback and simultaneous talk resolution could all be added cumulatively with these—and in the implementation of appendix 1, model 3, they are—, but were not included here for simplicity.

At StartTalk, the beginning speaker determines the *sort* of its utterance. It chooses SELECT(X) based on its SELECTIVITY likelihood with a target agent chosen at random. Else, it chooses 'Any?' based on a global-parameter likelihood (ASKANYBODY), or then 'More' based on half its VERBOSITY likelihood. Otherwise, the variable *sort* is set to the general 'pTRP'. This is in the last four bold lines of figure 4.6.

How much participants select others to talk next in a discussion is assumed to be characterizable of each conversant and relevant enough of each interaction to be made into individual agent attributes. It enables a simulation of parties with more or less frequent (apparent) roles of delegating talk to others, as if 'chairing' the discussion. The frequency of 'Any?' utterances, on the other hand, does not seem to be characterizable or even useful in this general way. Hence a global parameter making it a characteristic of the discussion, not of individuals.

Choosing 'More' through the VERBOSITY attribute comes from it being a decision to *continue* talking—and in this case to indicate it somehow. But it is a decision made *in*

---

[13]In the implemented program (appendix 1, model 2), the behaviour returned is 'ta/ ' just for our visual recognition of this. Likewise, quitting talk because of simultaneity was changed to ' .. ', just a visually different silence—although used by the evaluation program too (appendix 2), to facilitate some contextual identifications.

Figure 4.6: Procedures for the various sorts of TRPs.

⋮

If I am *speaker*:

    while (decreasing) utterance *length* > 0:

        if anyone *talked* (except for 'tal_'), test CONFIDENCE,

          then stop if failing it: set *speaker* to *talked* and say ' - '

        at *midTRPs* intervals, test INTERACTIVITY to say 'pTRP'

        (in e.g. 20% of cases if *sort* is SELECT(X) or 'Any?', say it;

        in e.g. 20% of doing this, 'change heart': set *sort* to 'pTRP')

        otherwise, say 'talk'

    when *length* = 0: (only once)

        set *wannaTalk* to a test of VERBOSITY **if *sort* = 'pTRP',**

        because if *sort* = 'More' set *wannaTalk* true,

        else, if *sort* = SELECT(X) or 'Any?', set it false

        set *tail* (finishing talk) to around 0.4s, or 0 if failed INTERACTIVITY

        set *pause* (length) to around TRPAUSING

        say (chosen pre-TRP from) *sort*

    while (decreasing) *tail* > 0, say 'tal_'

    while (decreasing) *pause* > 0 and nobody *talked*, say ' - ' (pause)

(else, I am listening:)

    if nobody *talked* and I *wannaTalk*: StartTalk

    if nobody is *speaker*, test TALKATIVENESS to StartTalk

    set *speaker* to *talked* **(if 'no one', set only if *sort* = 'pTRP' or 'Any?')**

    if *speaker* (if any) reached a TRP **(set *sort* to the pre-TRP):**

        if *sort* = 'pTRP', test TALKATIVENESS

        if *sort* = 'More', test TALKATIVENESS*EAGERNESS

        if *sort* = 'Any?', test TALKATIVENESS+EAGERNESS

        if *sort* = SELECT(X), I *wannaTalk* only if X = me

        set *wannaTalk* to the result of any of the tests above

        if I *wannaTalk*, test EAGERNESS to start now: if so, StartTalk

    otherwise, say ' - ' (silence).

StartTalk: **test** SELECTIVITY **to SELECT(X): set *sort* (random X),**

        else, decide *sort*='Any?' by testing ASKANYBODY**,**

        else, decide *sort*='More' by testing VERBOSITY/2

        or else, decide for *sort*='pTRP'

        set *length* to a random number of cycles up to a maximum or a mean

        set *midTRPs* (interval of middle pre-TRPs) around 2 seconds

        set *speaker* to myself, then say 'talk'.

advance of realizing the utterance, in its conceptualization as part of a multi-utterance message or discourse. The overt result is of course that it is hedged against turn-taking in one way or another (synctatically, semantically, pragmatically, etc). But the covert consequence is that no decision to continue needs to be made at the end, since the speaker *is* continuing to talk.[14]

So a portion of the agent's VERBOSITY likelihood chooses 'More' under the assumption that the more one is verbose, wants to keep talking, the more likely to indicate it explicitly. How much, it would depend on the rate we can assume speakers make these hedged utterances in discussion, rather than not: I used half the attribute's likelihood. This double use of VERBOSITY is much like the double use of CONFIDENCE, where two tests are made to select between three possibilities; in the case of VERBOSITY: 'More' in advance, or else, when it ends, continue or not a normal 'pTRP' utterance.

So when the utterance ends, the speaker signals the appropriate pre-TRP stored in the *sort* variable. It only decides whether to continue (*wannaTalk*) when the sort is 'pTRP', because it *is* continuing with 'More' and *not* continuing with the turn-yielding types SELECT(X) or 'Any?'.

Before that, at intermediary TRPs, the speaker may decide to 'anticipate' a turn-yielding sort, as described in §4.1.8. It means simply sending out that pre-TRP instead of the usual middle 'pTRP'.[15] Then, in a small likelihood of anticipating, the speaker may 'change heart': that is, reverse the turn-yielding stance being revealed in the middle the utterance to a free one, changing the *sort* to 'pTRP'. This is encoded in the first two bold lines of figure 4.6, but in the end I did not include this in the implemented version (appendix 1, model 3).

When listening agents 'hear' a pre-TRP, it is first saved in *sort*. If it is SELECT(X), only the agent identified in X will be wanting to talk. Else, the decision to talk uses appropriate likelihoods: the agent's normal TALKATIVENESS for 'pTRP', or reduced by EAGERNESS in the case of 'More' (a possible interruption), or amplified by it in the case of 'Any?' (an encouragement to talk).

---

[14]Indeed one might argue that in some real talk, even when no indication of continuity is made explicitly in an utterance (no 'More'), the speaker may have already decided to continue talking before its end, depending on his or her planned discourse or intentions.

[15]This description may appear misleading with relation to what it represents in reality: it is not that a real speaker chooses to 'anticipate' a turn-yielding message, but simply formulates one and follows it closely, without pause, with explanatory or justifying talk in a single utterance with different TCU 'sorts'.

The decision whether one wants to talk in reaction to that TRP is then stored in *wannaTalk*. Not only for the immediately next decision to talk, but for the next moments until someone takes the turn. If the sort is SELECT(X) or 'More', then *the floor is not free* in the subsequent TRP, while the selected or the current speaker does not (re)start to talk. As a consequence, the silence afterwards cannot be taken by the other listeners as 'no one is talking so I can start'; they could do it only if deciding to interrupt. This is ensured by preventing the *speaker* state to be set back to 'no one' in the silent interval unless *sort* is of the 'free floor' type: 'pTRP' or 'Any?'. Hence the *speaker* variable keeps indicating the agent who has the floor even if it is not speaking (yet).

Lastly, one point regarding feedback: positive feedback should be constrained in turn-yielding TRPs (SELECT(X) and 'Any?'). The speaker is explicitly yielding the floor, so continuers like "hm-hmm" do not make sense—although other signs of approval or endorsement for the utterance might be possible, like: "John, what are the items today?", "Yeah, let's see them!".

4.4.6   Hesitation and interruption

Procedures for hesitation and its interruption are, again, in **bold** in figure 4.7. This time they extend those of the previous subsection because they depend on *sort* for recognizing when a silence is hesitation (mid-utterance) or a TRP pause. One more variable is required here (*hesitate*) for counting the length of hesitation at TRPs and for listeners to decide when they can interrupt it.

The assumptions embodied here are reasonable ones, I hope—like the observation that hesitations may be interrupted after some minimal time. Others are mentioned in the literature, such as hesitating when being selected to speak (pausing). The operational details, like the states and constraints distinguishing hesitation from pausing, surfaced as necessary during the testing of the simulation (appendix 1).

So, a speaker may hesitate in the middle of an utterance (while talking) or at a TRP when starting to speak having the floor already: i.e. having been selected to speak or after a pause before beginning another utterance. The decision to hesitate comes from the likelihood in HESITATION.

At StartTalk, the agent checks if it has the floor: if *sort* is SELECT(me), or the agent is the *speaker* already. This is done before choosing the sort of the new utterance,

Figure 4.7: Hesitating in mid-utterance, or at TRPs when having the floor.

⋮

If I am *speaker*:
   while (decreasing) utterance *length* > 0:
      if anyone *talked* (except for 'tal‿'), test CONFIDENCE,
        then stop if failing it: set *speaker* to *talked* and say 'ta- '
     at *midTRPs* intervals, test INTERACTIVITY to say 'pTRP'
     (in e.g. 20% of cases if *sort* is SELECT(X) or 'Any?', anticipate;
      in e.g. 20% of anticipating, 'change heart': set *sort* to 'pTRP')
     If (decreasing) *hesitate* > 0 OR test HESITATION succeeds:
      increase *length* (hesitation not part of the utterance)
      say either 'ta- ' or ' - ' (test HESITATION to choose)
     otherwise, say 'talk'
   when *length* = 0: (only once)
   ⋮

(else, I am listening:)
   if nobody *talked*, and I *wannaTalk*, **and *sort* ≠ 'talk'**: StartTalk
   if nobody is *speaker*, test TALKATIVENESS to StartTalk
   set *speaker* to *talked* (if 'no one', only if *sort* = 'pTRP' or 'Any?')
   if *sort* = SELECT(X) set *speaker* to X
   if *speaker* last 'talk'ed, set *sort* to 'talk'
   if *speaker* hesitates for at least e.g. 0.8s (count in *hesitate*):
      test TALKATIVENESS*EAGERNESS to StartTalk
   if *speaker* (if any) reached a TRP (set *sort* to the pre-TRP):
   ⋮

StartTalk: **I have‿the‿floor if *sort* = SELECT(me)**
        OR when I am *speaker* already (continuing),
      test SELECTIVITY to SELECT(X): set *sort* (random X),
      ⋮
      set *speaker* to myself,
      if I have‿the‿floor, test HESITATION:
        if success, set *hesitate* around 0.5s and say 'ta- ' ("well…")
      else, say 'talk'.

because *sort* will be changed then; actually, the important distinction here is that there *are* two (or more) *modes* of 'starting to speak': one that affords hesitation and other that does not (in normal turn-taking). If at the end of StartTalk the agent decides to hesitate, *hesitate* is set to a length around 0.5 of a second: roughly between 0.2s and 0.8s (again, maybe too short a variation). In the subsequent cycles the hesitation is realized while this is counted down.

In the middle of the utterance, the speaker hesitates by the same procedure, but its HESITATION is tested at each cycle of talk now. It is then generated as a punctual, one-cycle event in the middle of talk, representing disfluencies that mostly reflect difficulties in formulation. Whatever way, *length* must always be incremented so the hesitation is not counted as part of the utterance length, as part of talk. And the choice between voiced or silent hesitation is made by a second test of the same attribute, the assumption being that the more hesitant someone is, the more the hesitation will be voiced: filler pauses ("erm", "umm", "you know"), self-interruptions, etc.[16]

While listening, it is assumed that the agents would only decide whether to interrupt after a certain extent of continuous hesitation: a threshold for interrupting, here taken to be 0.8s. They count this amount of continuous hesitation in the same *hesitate* variable.

But there is a problem there. How do listeners distinguish a silence (`-`) that is hesitation from a TRP pause? The former comes after '`talk`' and the latter after a pre-TRP, being recorded in *sort* by the procedures of the last subsection (intended as part of figure 4.7). Just as the floor is not free with certain TRP sorts (`SELECT(X)` and '`More`'), the it is not as well when the speaker begins an utterance but gets silent before completing it. The distinction is made by registering in the same *sort* that a speaker has started talking (`talk`), thus ending the TRP and marking that any subsequent silence must be hesitation.[17]

Thus, the occurrences of silence when *sort* = '`talk`' (and of '`ta-`' too) is cumulated in *hesitate* and cleared with any talk. If it reaches 0.8s or more of continuous hesitation, the listener tests whether to interrupt using the same condition of the other previous

---

[16]One might think the opposite could as well be true too. The (admittedly intuitive) reasoning for assuming the reverse here is that silent hesitations seem to derive more from difficulties or temporary 'blocks' in formulation (searching for a word), independently of the talker being 'characteristically' hesitant or not, which is sort of what HESITATION represents. So the more one would be hesitant (here), the more it would show through voiced hesitation.

[17]Of course, real listeners make this crucial distinction anyway when (in a state of) 'understanding talk'. It is just that the simulation having no contents of talk so far did not need it, as talk was just checked, automaton-like, for its presence or not.

type of interruption: the product of TALKATIVENESS and EAGERNESS. In reality, there may be several other 'appropriateness' gradations for interrupting, based on the contents of talk, the listener's interest in it, and the degree with which the utterance has been developed: e.g. it is probably 'easier' to interrupt when the speaker has not said much of the planned utterance so far ("I think that it's, you know, I mean, err") than when one has already realized a substantial part of it.

The indication of 'middle of talk' in *sort* then prevents listeners from thinking that the silent hesitation is the next TRP (when they have decided to talk and are waiting for any silence to start) *or* that it means 'no one is talking', so anyone could start. The extra condition of *sort* ≠ 'talk' when *wannaTalk* is true guarantees the first case. The second is ensured by keeping *speaker* unchanged while in hesitation, when *sort* = 'talk', just as when it is 'More'.

And when someone is selected to speak as indicated by *sort* = SELECT(X), *speaker* is changed to that X immediately, as if it is already speaking, to prevent others from starting because 'no one is talking' (like 'More'). Hesitation at a select-next TRP, therefore, is considered as from that selected 'speaker', and after the same threshold the agent would decide whether to interrupt it in the same way as in mid-utterance hesitations. *However*, the speaker hesitating after a usual 'pTRP' pause and before beginning another utterance continues to be a free-floor TRP like 'no one is talking', so listeners will test their normal TALKATIVENESS then, even after voiced hesitations, until any 'talk'.

One last note: in real discussions listeners sometimes give positive feedback ("hm-hum") when the speaker hesitates—and of course they do in *elicits* like "you know" (§2.3), represented here by 'ta- ' as well. This seems either an automatic response to the interruption of the talk or an encouragement for the speaker to go on (a continuer). It may even be more common than thought at first, but was not included in the implementation here, for simplicity.

4.5   Summary of assumptions

1. Talk goes in *utterances* separated by (silent) *pauses* of varying lengths. Each utterance is a complete unit of talk grammatically and prosodically; that is, it ends in a TRP (transition-relevance place) and may have other *intermediary* TRPs in

the middle that correspond to what listeners may recognize as possible comple-
tion places of the utterance. This means that utterances can be formed of one or
more of the turn-construction units (TCUs) of Sacks et al. (1974). And turns can
contain one or more utterances.

2. Utterances may end in a 'tail' of finishing talk of varying length (around 0.4 of a
second), that goes after the last TRP. This is the talk that is overlapped in speaker
transitions when a turn taker decides to speak immediately after the TRP without
waiting for the speaker to stop. In general, it corresponds to finishing talk spoken
in a lower pitch and loudness.

3. Possible utterance completions, or TRPs, are indicated explicitly by pre-TRP
behaviours, rather than being a listener interpretation of the cues in speech such
as syntax and intonation, and nonverbal behaviours like gaze.

4. The speaker determines whether it wants to continue talking in a new utterance at
the end of each one. Then it makes a pause whose average length is characteristic
of each conversant. Then, if no one has started to speak in the mean time, it
would start another utterance if it has decided so, or exit the speakership state.
Hence, agents stay in the 'pause' even when deciding not to continue talking,
as a time-off interval to prevent them from starting to speak immediately after
having decided *not* to continue talking.

5. *Who* is talking is verified at each cycle by the agents in a sort of 'cognitive
filter'. The cycle-time (how long each cycle represents) is therefore more than
the minimal unit of time: it represents the *response-time* of conversants, as there
is no 'buffering' of input to emulate delays in recognition (varying attentions).
If it is more than one, only one is considered for any reaction by each agent
(excluding itself, of course); preferably not the current *speaker*, so that the agent
can change focus to a new one if it wants. In the procedures of the model, agents
always promptly change focus to new speakers. If there is more than one talking
apart from the current focused speaker, one is chosen at random.

6. Feedback can be delayed depending on the inverse of the likelihood to give feed-
back itself. That is, the more one is wont to give feedback, the more it will be
prompt, immediately after the pre-TRP; else, it is variably delayed around half
second of time. Feedback itself has a length varying around 0.4 of a second.

7. Negative feedback causes the agent that was the last to emit a pre-TRP to respond

to it and start to speak, whether someone else is already speaking or not. If the agent itself is speaking (continuing to), it first makes a short silent pause before starting a new utterance, marking the interruption of the prior talk.

8. Starting to speak occurs in a two-step process: first the listener agent decides whether it wants to talk, then whether it should start immediately after the pre-TRP (at the possible utterance completion) or wait for the speaker to stop. If speaking immediately, it may overlap finishing talk or talk in the middle of the utterance (if it is not its end after all): if the current speaker gives up and stops speaking, the new taker has interrupted it (him or her); but on the contrary, if the new speaker gives up, it was just a false-start.

9. Deciding to talk because no one is talking already only occurs after a moment of silence when nobody speaks. Such starters are assumed to be not as quick as those who have decided to speak earlier and are just waiting for the current speaker to stop: these will start immediately at the first sign of silence, whereas the others shall wait a little longer to see whether 'no one is talking'.

10. The various assumptions of simultaneous talk are in §4.4.4.

11. Utterances may change the nature of subsequent TRPs with regards to turn-taking liberties. There are various sorts: a particular party is selected to talk next, or anyone is encouraged to talk, or the utterance indicates that there is 'more to come' and so turn-taking is discouraged, or the TRP is an ordinary one. Starting to speak after a more-to-come utterance or when someone else is selected to speak next, for example, should be taken as a different decision than that of free TRPs.

12. Hesitation can occur in the middle of talk: both silent or voice hesitation (discontinuities, repetition, fillers: "erm"). Silent hesitation is distinguished from pauses by the context: whether the immediately prior talk ended in a TRP or not. Listeners may decide to interrupt a hesitation (when it is past a certain minimum), in a similar 'kind' of decision as after a more-to-come utterance.

13. Speakers may also 'hesitate' (or take longer to start) when they already have or are granted the floor: when they are selected to speak next or after pausing before starting a new utterance. Listeners would interrupt there too: hence, when an agent is selected to speak next and does not start immediately, it may be interrupted.

4.6 Conclusion

This chapter presented a model of turn-taking intended as a symbolic, simulational representation of the process and its associated issues. It is made of four components: behaviours, agent attributes, interactive states and procedures. The 'behaviours' only represent whether interactants are talking or not, and what sorts of contraints to turn-taking are involved in the TRPs occurring in the talk. They are broken up in moment-sized units as the smallest unit of simulation and *response time* of the participants, forming up the TCUs, utterances, hesitations, pauses and, ultimately, turns at talk.

Conversants are modelled by agents with a set of attributes containing the probabilistic parameters of their turn-taking behaviour (broadly speaking) in the discussion: how much they want to talk, to give feedback, and so forth. They also have a set of 'internal' variables recording their various interactive states throughout the discussion, which are relevant (some of them, at least) for any system that needs to make the distinctions made here: speakership, pauses and hesitations, normal turn-taking or interruption, etc. Procedures that determine the agents' behaviour (in the broad sense) include: giving and recognizing backchannel feedback, practices to deal with simultaneous talk, the constraints posed by various sorts of TRPs, hesitation and types of 'interruption'.

In this model, TRPs are signalled explicitly. A possible future work would be to extend it with representations of the various actual *cues* of possible utterance completion according to the listeners' individual, possibly varying, interpretation (misspotted TRPs, etc), and not as explicit signals emitted by the speaker. Actual cues (§2.6) would be syntactic completion, intonation, prosody (rhythm and tempo), visual behaviours, loudness, and others.

Another possible extension, or use of this framework of conversant multi-agents, is by adding contents of talk and the reasons for interaction as in dialogue systems, following theories such as in §2.7. This would mean systematically replacing the various probabilistic attributes by more principled reasons for why agents decide to talk at each moment, such as obligations from the *moves* (in a dialogue and discourse sense) contained in the utterances.

Chapter 5

Examples

The simulation was implemented in Java: its code is listed in appendix 1. The basic model, `model.java`, includes a simple front-end that shows the results cycle-by-cycle vertically: one cycle (or 'round' of simulation) per line and various columns for the agents in the group. This form of output is useful for going one cycle at a time to understand what is happening in the simulation, but it produces very long printouts.

Therefore, the examples shown here were created with the front-end of `hmodel.java`, listed at the end of appendix 1. It runs the same simulation (in whatever extension it is compiled) showing the results *horizontally*, in 80-character-width 'blocks', with one line per agent of the group. Thus, the behaviours on each block continue unbroken on the next block coming down below: the last behaviours in one (last column) are followed by the first behaviours (second column) in the next.

The first column in all blocks lists the agent names—only important when a name is used for selecting the next speaker at the end of an utterance (in §5.3). The first line in each block identifies the cycle separation with marks at every second. There can be more or less cycles per second according to a global parameter that determines the resolution of the simulation, what time each cycle represents. By default, it is 0.2s, which gives 5 cycles per second.

For the benefit of those who may have come here directly without reading the previous chapters, a quick description of the behaviours is useful. Intuitively, 'talk' means that an agent is talking at a given moment (cycle), talk that is not complete yet; 'tal_' is the extended or redundant finishing talk that can be overlapped by a new speaker without much import or even any notice of it; and 'ta- ' is a disfluency in the talk, like a

voiced interruption ("uh"), the beginning of a self-repair, or a 'filler' hesitation such as "erm" or "you know". This 'ta- ' will only appear in the examples in §5.2 and §5.4.

Empty slots of ' - ' or ' . ' and also ' .. ' are all silence. The first is the general listening silence; the second is the *pausing* silence after an utterance, while the agent still has not disengaged from speakership, maybe even wanting to resume talk; and the third is the 'cut-off', abandoning-talk silence, when quitting from a started utterance due to simultaneous talk. They are distinguished just to facilitate our identification of different states and state changes in the agents, but are all recognised as (and can be made) the same. The cut-off silence is particularly helpful to distinguish from hesitations, when silence after talk does not necessarily mean the speaker is quitting.

Talk occurs in *utterances*, variable-length sequences of 'talk' and 'ta- ' and silences that represent hesitations in their middle (as in §5.4). Silence also occurs in between utterances, when the agent is pausing (as speaker) or is listening. In the first model, utterances always end in a 'pTRP' behaviour, which is talk indicating the projection of their completion (a transition-relevance place, TRP). This can be followed by a short but variable-length *tail* of additional or redundant finishing talk (tal_), which goes out typically in a lower(ing) pitch and/or loudness.

Backchannels occur in sequences of 'uhuh' (positive feedback such as "uh huh" and the like) and 'huh?' (negative feedback such as "huh?" and the like). It is not that *each* of these behaviours represents one separate utterance, but it is the whole sequence of the same behaviour that forms one "uh huh" or "huh?" vocalization, just as with 'talk'.

5.1 Basic simulation with backchannel feedback

The basic simulation in model.java corresponds to the procedures of figure 4.3 explained in sections 4.4.2 and 4.4.3: feedback generation with a very simple resolution of simultaneous talk and no other sorts of TRPs or hesitations.

5.1.1 A 3-party example

The first example is a 3-party discussion with agents having TALKATIVENESS of 0.1, 0.2 and 0.4 respectively, FEEDBACK of 0.2, 0.3 and 0.1, and everyone with EAGER-

NESS of 0.4, CONFIDENCE of 0.6 and VERBOSITY or 0.3. These are the meanings of the parameter digits in the java command-line below.

```
$ java hmodel 3 t124 f231 e4 c6 v3
Turn-taking model in a simulation of small group discussion v.22/07/04

____|0s__|____|____|____|____|1s__|____|____|____|____|2s__|____|____|____|____|
AgtA| -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtB| -   talk talk talk talk talk talk talk talk talk talk talk pTRP tal_  .
AgtC| -   talk  ..   -    -    -    -    -    -    -    -    -    -    -    -

____|3s__|____|____|____|____|4s__|____|____|____|____|5s__|____|____|____|____|
AgtA|uhuh uhuh uhuh  -    -    -    -    -    -    -    -    -    -    -    -
AgtB| .    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtC| -    -   talk talk talk talk talk talk talk talk pTRP tal_ tal_ tal_  .

____|6s__|____|____|____|____|7s__|____|____|____|____|8s__|____|____|____|____|
AgtA| -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtB|talk talk talk talk talk talk pTRP talk talk talk talk talk talk talk talk
AgtC| .    -    -    -    -    -    -    -   uhuh uhuh  -    -    -    -    -

____|9s__|____|____|____|____|10s_|____|____|____|____|11s_|____|____|____|____|
AgtA| -   talk talk talk talk pTRP  .    .    -    -    -    -    -    -   uhuh
AgtB|pTRP  .    -    -    -    -    -   uhuh uhuh  -    -    -   huh? huh?  -
AgtC| -    -    -    -    -    -    -   talk talk talk talk pTRP talk talk talk

____|12s_|____|____|____|____|13s_|____|____|____|____|14s_|____|____|____|____|
AgtA|uhuh  -    -    -    -    -    -    -    -    -    -    -    -    -   talk
AgtB| -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtC| -   talk talk talk talk talk pTRP tal_ tal_  .    .    -    -    -    -

____|15s_|____|____|____|____|16s_|____|____|____|____|17s_|____|____|____|____|
AgtA|talk talk talk talk talk talk talk talk talk talk pTRP  .   talk  ..   -
AgtB| -    -    -    -    -    -    -    -    -    -    -    -   talk talk talk
AgtC| -    -    -    -    -    -    -    -    -    -    -    -    -   talk  ..

____|18s_|____|____|____|____|19s_|____|____|____|____|20s_|____|____|____|____|
AgtA| -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtB|pTRP tal_  .    .    .   talk talk talk talk talk pTRP  .    -    -    -
AgtC| -    -    -    -    -    -    -    -    -    -    -   talk talk talk talk

____|21s_|____|____|____|____|22s_|____|____|____|____|23s_|____|____|____|____|
AgtA| -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtB| -    -    -    -    -    -   talk talk talk talk talk pTRP  .    .    .
AgtC|talk talk talk talk talk pTRP tal_ tal_ tal_  -    -    -    -    -   uhuh
```

The attributes in CAPITALS are the probabilistic parameters that model each agent (table 4.2): TALKATIVENESS is how often an agent wants to talk at the appropriate turn-taking points (at silences and at TRPs: that is, after 'pTRP'); FEEDBACK is how much one gives backchannel feedback when not talking, either 'uhuh' or 'huh?' (chosen in a fixed proportion); EAGERNESS is how eager one wants to speak when deciding

so, either right away in a TRP, or (politely) after waiting the speaker to stop; CONFI-
DENCE is the likelihood of persisting to talk simultaneously with others, which may
determine who 'wins' the simultaneity; and VERBOSITY is how much one wants to
continue making new utterances while as speaker.

In the example, right at the second cycle (at 0.2s) two agents decide to start simulta-
neously; one, `AgtC`, gives up immediately at the next cycle (marked by ' .. ') while
the other, `AgtB`, continues to produce a full utterance. `AgtA` gives positive feedback
at 3s ('`uhuh`' for 0.6s), just a little after the end of that utterance, while `AgtC` starts to
speak at 3.2s. After it finishes, at 6s, `AgtB` starts another, longer utterance containing
an intermediary TRP signalled by '`pTRP`' at 7.2s.

This indicates a possible completion of the utterance—listeners cannot know precisely
yet, at that moment—, or a place of some uptake (the recognition of an idea or message,
end of a clause), so they can decide either to speak or to respond to it in some other way.
Only `AgtC` responds with a short positive feedback of 0.4s at 7.6s. If an agent decided
to speak there, its EAGERNESS would give the likelihood of starting immediately or
waiting to see if the speaker was really finishing.

Two turns later (after two speaker changes), an intermediary TRP is responded with
two sorts of feedback: negative feedback at 11.4s (`huh?`) then positive feedback a little
later at 11.8s (that is, delayed by 0.4s from the '`pTRP`'). Negative feedback requires the
speaker to respond, so `AgtC` stops its utterance midway (at 12s) and restarts—the short
break then indicates the start of a new utterance addressing the problem of hearing or
understanding that was raised.

After this utterance ends, nobody talks for one second (a short lapse actually), until
`AgtA` begins at 14.8s. It finishes its utterance then makes a *very* short pause of 0.2s—
like, maybe, a quick inhalation—before resuming at 17.4s. The problem is that the
other two agents also take cue of the TRP signalled at 17s to speak as well, almost
simultaneously. In this case, `AgtB` is the only one who continues whilst the others
immediately withdraw. We can see that talkers have very sharp reactions here; it is
possible that in real life people would sometimes speak a little more (perhaps 0.6s or
more) until realizing the simultaneity and reacting to it.

`AgtB` talks in two short utterances separated by a medium pause of 0.6s (noted by
' . ' silences). Then `AgtC` *latches on* perfectly to the second one at 20.2s by starting
a new utterance right when the prior one finishes, at the indication of its (possible)

completion which did not have any extra finishing talk (another latch occurred earlier at 9.2s). The new utterance is later overlapped when `AgtB` also starts forthwith at the utterance's first possible completion (22.2s). That was not quite its completion yet: there was still 0.6s of finishing talk which is then overlapped by the new start.

5.1.2 A 4-party example

With more participants, the chances that someone will want to talk at each TRP or silence increases, even with low TALKATIVENESS. In this example, agents have reasonably low TALKATIVENESS and FEEDBACK, but low CONFIDENCE as well (0.4 for all), which makes them more likely to stop if talking simultaneously with others.

```
$ java hmodel 4 t1234 f2121 e4623 c4 v5
Turn-taking model in a simulation of small group discussion v.22/07/04

____|0s__|____|____|____|____|1s__|____|____|____|____|2s__|____|____|____|____|
AgtA| -    talk  ..   -    -    -    -    -    -    -    -    -    -    -    -
AgtB| -    -    talk talk talk talk talk talk talk talk pTRP tal_ tal_ tal_  .
AgtC| -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtD| -    talk  ..   -    -    -    -    -    -    -    -    -    uhuh  -    -

____|3s__|____|____|____|____|4s__|____|____|____|____|5s__|____|____|____|____|
AgtA| -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtB| .    .    -    -    -    talk talk  ..   -    -    -    -    -    -    -
AgtC| -    talk  ..   -    -    -    -    -    -    -    talk talk talk talk
AgtD| -    talk  ..   -    talk talk talk  ..   -    -    -    -    -    -    -

____|6s__|____|____|____|____|7s__|____|____|____|____|8s__|____|____|____|____|
AgtA| -    -    -    -    -    -    -    talk  ..   -    -    -    -    -    -
AgtB| -    -    -    -    -    -    -    talk talk talk pTRP tal_ tal_  -    -
AgtC|talk talk talk pTRP tal_ tal_  .    .    -    -    -    talk talk talk talk
AgtD| -    -    -    -    -    -    -    -    -    -    -    -    -    -    -

____|9s__|____|____|____|____|10s_|____|____|____|____|11s_|____|____|____|____|
AgtA| -    -    -    -    -    -    -    -    -    -    -    -    -    -    uhuh
AgtB| -    -    -    -    -    talk  ..   -    -    -    -    -    talk  ..   -
AgtC|talk talk pTRP  .    .    talk talk talk talk talk talk pTRP talk  ..   -
AgtD| -    -    -    -    -    -    -    -    -    -    -    -    -    -    -

____|12s_|____|____|____|____|13s_|____|____|____|____|14s_|____|____|____|____|
AgtA|uhuh  -    -    -    -    -    -    talk talk talk talk talk pTRP talk talk
AgtB| -    -    -    -    -    -    -    -    talk  ..   -    -    -    -    -
AgtC| -    -    -    -    -    -    -    -    -    -    -    -    -    talk  ..
AgtD| -    talk talk talk talk pTRP  .    .    -    -    -    -    -    -    -
```

In the beginning three agents start to speak more or less simultaneously. The model allows just one cycle before agents starting at silences *recognise* there are others speak-

ing; so starts of talk can be one cycle apart at most. It is simply that they check the previous cycle, not the current, when deciding to speak in the next one—which sort of simulates they are busy thinking up what to say then. However, once they start they recognise immediately if there are others speaking.

In this initial case, only the late starter (`AgtB`) continues; the others stop at the first hearing of the simultaneity, because of each other—what in the next chapter I will call a 'collective' false-start. The simple resolution of multiple talk here does not provide for any preference of the type 'first to start takes the turn'; it is the agents' CONFIDENCE that determine whether they insist to talk. As both first starters stopped right away, the second one (`AgtB`) could continue without actually being in any simultaneity.

Then at 2.4s there is a very short backchannel feedback (like, say, "m": a closed-mouth vocalization) at the utterance's finishing talk, followed by two groups of collective false-starts at 3.2–3.4s and 4–4.4s. In the second instance, the agents persist for one or two cycles before stopping together: if only one had stopped, the other would have continued speaking to a full utterance.

Further on, two agents start at 7.4s, just 0.2s after `AgtC` stopped speaking. This can only mean they have decided to talk at the TRP of the previous utterance (one cycle after, at 6.8s) and were waiting for it to finish: because an agent starting from silence (when nobody is talking) would wait one cycle more. So, listeners already intending-to-speak are quicker in starting after someone stops than others who will only decide to speak when there is silence. At 7.6s, one of the two starters gives up immediately recognizing the simultaneity, while the other continues producing a short utterance that is overlapped by `AgtC` at 8.2s.

At 10s another agent tries to speak together with the current speaker (`AgtC`) restarting after a pause, but gives up. The simple resolution of multiple talk here also does not provide for any preference of the type 'current speaker continues *unless* another self-selects' from rule 3 of the turn-taking systematics (§2.1.4), if it would mean *even when* current and another self-selecting speaker start more-or-less simultaneously. Here, it is much like any other simultaneous start: anyone might stop or continue speaking depending on their CONFIDENCE.

Finally, at 11.4s, another agent starts to speak forthwith at a TRP, causing both speakers to stop just afterwards. In this case the speaker was cut off, interrupted by a start of talk in the middle of its utterance. Of course, it is a random decision based on the

agents' CONFIDENCE attributes, but it illustrates the bluntness that a simple resolution of simultaneous talk generates. This is much improved in the model that is extended with a more elaborate resolution of simultaneous talk, to be seen next.

5.2 Elaborated simultaneous talk

The model with more simultaneous talk procedures corresponds to those of figure 4.5 described in section 4.4.4, and implemented in `model2.java` of appendix 1.

```
$ javac model2.java
$ java hmodel 4 t2345 f2341 c3475 e3642 v6
Turn-taking model in a simulation of small group discussion v.22/07/04

____|0s__|____|____|____|____|1s__|____|____|____|____|2s__|____|____|____|____|
AgtA|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtB|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtC|  -   talk talk talk talk pTRP tal_ tal_  -    -    -    -    -    -    -
AgtD|  -   talk talk ta-  TALK TALK TALK talk talk talk talk talk talk pTRP   .

____|3s__|____|____|____|____|4s__|____|____|____|____|5s__|____|____|____|____|
AgtA|  -    -    -    -    -    -    -    -    -   uhuh uhuh  -    -    -    -
AgtB|talk talk pTRP   .    -    -    -    -    -    -    -    -    -    -    -
AgtC|  -    -    -    -    -   talk talk pTRP TALK TALK TALK talk talk talk talk
AgtD|  -    -    -    -    -    -   talk talk talk pTRP tal_ tal_  -    -    -

____|6s__|____|____|____|____|7s__|____|____|____|____|8s__|____|____|____|____|
AgtA|  -    -    -    -    -    -    -    -    -    -   talk TALK TALK ta-   ..
AgtB|  -    -    -    -    -   uhuh uhuh uhuh  -   talk talk talk talk talk pTRP
AgtC|talk talk talk pTRP tal_ tal_ tal_  .    .   talk ta-  ta-  TALK TALK TALK
AgtD|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -

____|9s__|____|____|____|____|10s_|____|____|____|____|11s_|____|____|____|____|
AgtA|  -    -   uhuh uhuh  -    -    -    -    -    -    -    -    -    -    -
AgtB|tal_ tal_  -    -    -    -    -    -    -    -    -    -    -    -    -
AgtC|TALK talk talk ta/  talk talk talk talk talk talk talk talk talk talk talk
AgtD|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -

____|12s_|____|____|____|____|13s_|____|____|____|____|14s_|____|____|____|____|
AgtA|  -    -    -    -    -    -    -    -   talk talk pTRP tal_  -    -    -
AgtB|  -   talk talk talk talk talk talk pTRP talk talk TALK TALK talk talk talk
AgtC|pTRP   .    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtD|  -    -   uhuh uhuh uhuh  -    -    -   talk ta-  ta-   ..   -    -    -
```

In this example, two agents start to speak at 0.2s and both continue simultaneously, first ignoring each other then one of them (AgtD) hesitating briefly ('ta- ') and upgrading or 'stepping-up' talk to a more competitive stance (such as louder), which is

represented by 'TALK'. But since the other speaker's intended utterance is short (a little over 1s when completed), its producer is not going to stop.

That is one of the assumptions embedded in these practices of simultaneous talk: if the speaker is at 1 second or less of finishing its utterance—not counting an additional tail of 'tal_'—, it could just as well complete it. Short utterances usually go simultaneously with other talk without disturbing it too much; generally it seems their producers just do not mind or ever consider stopping in the middle because of its simultaneity. Similarly, when a longer utterance is in its final stages (around 1s or less of completion), and there is other talk for whatever reason, it would not make sense to stop then, after most of the utterance has been realized already.

In the present case, both speakers continue and when AgtD finally emerges from the simultaneity alone, it readjusts its talk back to normal. This happens at 1.4s yet still when the other was finishing talk, as the agents do not take 'tal_' as conflictive (so that they do not stop when overlapping a finishing speaker).

Another similar simultaneous episode with competitive talk begins at 4.2s, but the one more worth looking begins at 7.8–8s: three agents competing for the floor. Two of them step-up their talk and the other (AgtB) goes on to ignore the first two; its utterance is also within the threshold of a short utterance so it is not going to stop anymore. Soon, one of the other two (AgtA) gives up at 8.8s as it recognized that another speaker, AgtC, also stepped-up its talk.

This one, the remaining speaker after the simultaneous episode, decides to repeat part (or all) of its talk that was obscured by the simultaneity. That is another assumption included in the procedures here. It happens in the simulation when there is a 'long' simultaneity of more than 1s (the same threshold again). The agent then breaks its talk with 'ta/ ' at 9.6s and restarts talking (supposedly to repeat it). This behaviour is intended to be just a 'ta- ', a self-interruption, made different in the implemented program so that we know it corresponds to a decision to recycle (part of) the talk obscured by the simultaneity: a distinction transparent to the agents just like with ' .. '.

There is another simultaneity later at 13.6s, this time in the middle of an utterance. Two listeners decide to speak at a TRP in the talk (perhaps thinking the speaker was finishing): one makes a short utterance while the other hesitates before quitting, leaving a false-start behind.

5.2.1   A 5-party example

In the following 5-party simulation, agents are given medium-to-high parameters of
TALKATIVENESS that could create a busy discussion (0.2 to 0.6) and some higher
CONFIDENCE values for the more talkative agents this time (respectively 0.2, 0.4, 0.7,
0.5, 0.8) for more simultaneous talk conflicts. Other attributes not included in the
command-line are left at default likelihoods of 0.5.

```
$ java hmodel 5 t23456 e46273 c24758
Turn-taking model in a simulation of small group discussion v.22/07/04

____|0s__|____|____|____|____|1s__|____|____|____|____|2s__|____|____|____|____|
AgtA|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtB|  -   talk ta-   ..   -    -    -    -    -    -    -    -    -    -    -
AgtC|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtD|  -   talk talk ta-   ..   -    -    -    -    -    -    -    -   talk talk
AgtE|talk talk talk ta-  ta-  talk talk talk talk talk talk talk pTRP  .    -

____|3s__|____|____|____|____|4s__|____|____|____|____|5s__|____|____|____|____|
AgtA|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtB|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtC|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtD|talk talk pTRP talk talk talk talk pTRP  -    -    -    -    -    -    -
AgtE|  -    -    -   talk talk ta-  talk ta-  talk talk talk talk talk talk talk

____|6s__|____|____|____|____|7s__|____|____|____|____|8s__|____|____|____|____|
AgtA|  -    -   uhuh uhuh uhuh  -    -    -    -    -    -    -    -    -    -
AgtB|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtC|  -   huh? huh?  -    -    -    -    -    -    -    -    -    -    -   uhuh
AgtD|  -   talk talk talk talk talk talk ta-   ..   -    -    -    -    -    -
AgtE|pTRP tal_ tal_  -   talk talk TALK TALK TALK talk talk talk talk pTRP  .

____|9s__|____|____|____|____|10s_|____|____|____|____|11s_|____|____|____|____|
AgtA|  -    -    -    -    -    -    -    -    -   talk talk talk talk talk pTRP
AgtB|talk talk ta-   ..   -    -    -    -    -    -    -    -    -    -    -
AgtC|uhuh talk talk pTRP tal_ tal_  -    -    -   huh? huh?  -    -    -    -
AgtD|  -   talk ta-   ..   -    -    -    -    -    -    -    -    -    -    -
AgtE|talk talk talk talk ta-  talk talk talk pTRP  .    -    -   talk talk talk

____|12s_|____|____|____|____|13s_|____|____|____|____|14s_|____|____|____|____|
AgtA|tal_ tal_  .    .    -    -    -    -    -    -    -    -    -    -   uhuh
AgtB|  -    -    -   talk ta-   ..   -    -    -    -    -    -    -    -    -
AgtC|  -    -    -   talk talk pTRP talk TALK ta-   ..   -    -    -    -    -
AgtD|  -    -    -    -   talk TALK TALK TALK TALK talk talk pTRP tal_ tal_  .
AgtE|pTRP  -    -    -    -    -    -    -    -    -    -    -   uhuh uhuh uhuh
```

Right at the beginning in this example, three agents start to speak more or less si-
multaneously, with only one continuing (`AgtE`); the other two make false-starts. The
persisting agent just hesitates briefly at 0.6–0.8s because of the other talk. We can take

repeated 'ta- ' as a *continuous* hesitation just like with 'talk' or 'uhuh' (say, a long "ummm"), not necessarily multiple discrete breaks.

A decision to stop is now taken in two steps: first the agent is likely to hesitate with low confidence, then stop afterwards in a second adverse decision. But it is also possible that an agent hesitate then talk then hesitate then talk again and so on, simulating what is sometimes observed in real simultaneities: a cycle of stop-restart-stop-repeat caused by other talk, while one supposedly attempts to speak 'in the clear'. That is observed in the next simultaneous talk episode, at 4s and 4.4s: AgtE hesitates twice before the other speaker completes its utterance. We know that the AgtD's 'pTRP' at 4.4s is the end of its utterance and not simply an intermediary TRP, since there is no ' .. ' or 'tal_' thereafter and the agent did not hesitate first.

The next conflictive simultaneity (from 6.8s onwards) is interesting because it is the only case here that an agent may start to speak not in a TRP, but because of negative feedback. Here it happens that another agent started at the possible completion of the recent utterance, when a little later its producer also speaks responding to the 'huh?' vocalization (when it ended). The same thing also happens at 11.4s.

Finally, the last interesting simultaneity begins at 12.6s with another multiple start. This time both agents eventually step-up their talk to a competitive mode, but one of them immediately hesitates (AgtC at 13.6s). This means it will stop next, because when one simultaneous speaker upgrades talk, the other(s) either step-up too or end up quitting—unless they ignore the simultaneity as when finishing or just producing a short utterance, 1s or less to end. Thus when an agent upgrades its talk it has a slight edge over the others in that it will not quit as easily then, unless the other(s) ignore or upgrade their talk too.

Note how, in relation to the previous model, speakers do not quit talking as easily as before now. Simultaneous or partly simultaneous utterances are common in many real discussions: people do not give up as promptly as in fractions of a second. In many circumstances there are practices and hitches they use to insist and get their message across anyhow.

5.3 Sorts of TRPs

The third model or extension to the basic one (`model3.java` in appendix 1), corresponds to the procedures of figure 4.6 described in section 4.4.5. It incorporates the above elaborated simultaneous talk and adds other 'sorts' of pre-TRP behaviours that signal TRPs with different turn-taking restrictions.

Hitherto, TRPs were indicated only by '`pTRP`'; now there is also '`More`', which is when the utterance entails there is more the speaker is going to say; '`Any?`', when the speaker encourages anyone to speak such as in questions like "Has anybody…"; and, when an agent's name ends the utterance, it indicates that that agent was 'selected' to talk next, as when one asks a question addressed to someone specifically.

```
$ javac model3.java
$ java hmodel 5 t12345 c2 f5 e5 v75864 s64753
____|0s__|____|____|____|____|1s__|____|____|____|____|2s__|____|____|____|____|
AgtA| -   talk ta-   ..   -    -    -    -    -    -   talk talk talk AgtB tal_
AgtB|talk talk pTRP tal_  -    -   talk ta-   ..   -    -    -    -    -    -
AgtC| -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtD| -    -    -    -    -    -   talk ta-   ..  talk talk ta-   ..   -    -
AgtE|talk talk talk talk talk Any?  .    -    -    -   talk ta-  ta-   ..   -

____|3s__|____|____|____|____|4s__|____|____|____|____|5s__|____|____|____|____|
AgtA|tal_  .    .    -    -    -    -    -    -    -    -    -    -    -    -
AgtB| -    -   talk talk talk talk talk AgtC  .    .    -    -    -    -    -
AgtC| -    -    -    -    -    -    -    -    -   talk talk talk talk talk talk
AgtD| -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtE| -    -    -    -    -    -    -    -    -    -    -    -    -    -    -

____|6s__|____|____|____|____|7s__|____|____|____|____|8s__|____|____|____|____|
AgtA| -    -   talk talk talk talk talk More  .    .   talk talk talk  -   talk
AgtB| -    -    -    -    -    -    -    -    -   uhuh uhuh uhuh  -    -    -
AgtC|talk AgtA tal_  .    -    -    -    -    -    -    -    -    -    -    -
AgtD| -    -    -    -    -    -    -    -    -    -   huh? huh?  -    -    -
AgtE| -    -    -    -    -    -    -    -    -   uhuh uhuh  -    -    -    -

____|9s__|____|____|____|____|10s_|____|____|____|____|11s_|____|____|____|____|
AgtA|talk talk talk talk talk Any? tal_ tal_  -    -    -    -    -    -    -
AgtB| -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtC| -    -    -    -    -    -   talk talk ta-  ta-   ..   -    -    -    -
AgtD| -    -    -    -    -    -    -    -    -    -    -   talk talk talk talk
AgtE| -    -    -    -    -    -   talk ta-  talk ta-   ..  talk ta-   ..   -

____|12s_|____|____|____|____|13s_|____|____|____|____|14s_|____|____|____|____|
AgtA| -    -    -    -    -    -    -   uhuh uhuh  -    -    -    -    -    -
AgtB| -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtC| -    -    -    -    -    -    -    -   uhuh uhuh  -    -    -    -    -
AgtD|talk More tal_ tal_ tal_  .    .    -    -    -    -   uhuh uhuh  -    -
AgtE| -    -    -    -    -    -   talk talk talk talk pTRP talk talk talk pTRP
```

The agents in this example are given high VERBOSITY and SELECTIVITY likelihoods, so as to show examples of the three specific turn-taking sorts of utterances. VER-BOSITY is the parameter modelling how much a speaker wants to continue talking in new utterances, and therefore how often it indicates explicitly with 'More': representing for instance an utterance that is a subordinated clause lacking the subordinant, or a beginning like "I want so say some things". SELECTIVITY is another attribute modelling how often agents select others to speak in the discussion. Also, the fixed likelihood with which 'Any?' utterances are chosen was changed here (in the compiled program, for there is no command-line input for it) to 0.4.

Immediately, in the example, we have two simultaneous short utterances, one ending in the normal 'pTRP' and another ending in 'Any?'. This entails a *turn-yielding* TRP: we know its speaker is giving up the floor and, more than that, it is *encouraging* anybody to talk. Two agents then start right away but give up just afterwards. Then another agent speaks, persists through the other false-starts, and makes an utterance that selects AgtB to talk next, with a pre-TRP that is the name of the selected agent.

AgtB then waits the speaker to finish in order to start in the subsequent turn-yielding TRP and selects AgtC with its utterance, who also waits for silence before speaking, in turn selecting AgtA to speak next. This one starts right away without waiting for silence, and makes a short utterance explicitly indicating with 'More' that it is continuing to talk—supposedly to complete the message, idea or syntax announced or opened up in that utterance.

'More' *discourages* anyone to speak in the following restricted TRP, making the other agents to wait and see: to continue listening to what the speaker is going to say. Here they only respond with feedback, including a negative one which makes the next utterance to be interrupted at 8.6s so as to address the problem raised.

The next utterance is again one that encourages anyone to speak. This makes two agents start right away without waiting for the utterance's completion; subsequently, they both stop because of each other. Then another agent (AgtD) makes an utterance indicating explicitly more-to-say, but this time someone (the more talkative agent) decides to interrupt it at 13.2s in a smooth interruption.

So turn-taking is only restricted, not eliminated, at 'More' TRPs. With hesitations, TRPs selecting the next to speak can also be 'interrupted' by some non-selected other if the selected agent does not start to speak immediately, but instead hesitates.

5.4 Hesitations

The fourth model or extension to the basic one (`model4.java` in appendix 1) corresponds to the procedures of figure 4.7 described in section 4.4.6. It includes on top of the previous procedures the generation of hesitations in the middle of utterances and when already having the floor—when pausing before speaking again or when one is selected to speak—, together with the corresponding control by the listeners of when to interrupt or keep listening. Hesitation may be silent, with ' - ' after '`talk`', or with '`ta-` ', which represents voiced disfluencies or hesitation like "erm".

The example below shows various instances of hesitation in the middle of talk. We know that agents do not give up speakership at 0.4s because their behaviours there are ' - ', not ' .. ', which means they are hesitating in mid-utterance. One of them finally gives up a second later at 1.4s while the other (`AgtD`) makes short hesitations intercalated with talk until quitting its speakership at 3.6s. More hesitation is in the long utterance starting at 11.8s and going all the way to 16.8s (which is made much longer because of hesitation).

Listeners *can* interrupt a hesitating speaker, when it hesitates continually for 0.8s or more. Such an extent of hesitation only occurs in this example with `AgtA` at 10s (hesitating from 9.2s), and also with `AgtE` at 7.6s (hesitating from 6.8s because it was selected to speak). But the decision to interrupt has to be much less likely than a normal turn-taking: it is assumed to be just as when interrupting someone after a more-to-say utterance, only if the agent is talkative and/or eager enough.

There are three instances in this example of an agent hesitating *before* starting to speak, when that agent already has the floor. Two instances occur when `AgtE` is selected to speak at 6.8s and 16.8s: in both cases it hesitates before starting to speak (what others have called 'pause'[1]). The agent is granted the floor then, so it does not usually need to hurry up speaking before others do, as is the case in normal turn-taking. The third such instance, finally, is when `AgtE` pauses and hesitates before starting to speak in a new utterance, at 19s. That is a place a little riskier to delay resuming to talk because anyone could start there (less hesitantly) and take the turn instead.

---

[1]A distinction that could be made is whether it is an *intentional* pause or an (apparent) *unintentional* hesitation whilst thinking and formulating what to say.

```
$ javac model4.java
$ java hmodel 5 t24132 h45276 f1 e6 v6 c6 s3
____|0s__|____|____|____|____|1s__|____|____|____|____|2s__|____|____|____|____|
AgtA|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtB|talk ta-   -  talk ta-  ta-  ta-   ..   -    -    -    -    -    -  uhuh
AgtC|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtD|talk ta-   -    -  talk ta- TALK TALK talk   -  talk   -  talk pTRP talk
AgtE|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -


____|3s__|____|____|____|____|4s__|____|____|____|____|5s__|____|____|____|____|
AgtA|  -    -    -    -    -    -    -    -    -    -    -    -    -  talk talk
AgtB|uhuh talk talk talk talk talk More   .    .  talk talk talk pTRP tal_ tal_
AgtC|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtD|  -  talk ta-   ..   -    -    -    -    -    -    -    -    -    -    -
AgtE|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -


____|6s__|____|____|____|____|7s__|____|____|____|____|8s__|____|____|____|____|
AgtA|talk talk talk AgtE   .    .    -    -    -    -    -    -    -    -    -
AgtB|tal_   -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtC|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtD|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtE|uhuh uhuh   -    -    -    -  ta-  ta-  talk talk ta-   -  pTRP tal_   -


____|9s__|____|____|____|____|10s_|____|____|____|____|11s_|____|____|____|____|
AgtA|talk   -    -  ta-  ta-  talk   -  talk talk talk pTRP tal_ tal_   .    .
AgtB|  -    -    -    -    -    -    -    -    -    -    -    -    -    -  talk
AgtC|talk talk talk More tal_ tal_   -    -    -    -    -    -    -    -    -
AgtD|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtE|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -


____|12s_|____|____|____|____|13s_|____|____|____|____|14s_|____|____|____|____|
AgtA|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtB|talk ta-  ta-  talk ta-  talk talk ta-  ta-  ta-  talk   -    -  talk pTRP
AgtC|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtD|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtE|talk ta-   ..   -    -    -    -    -    -    -    -    -    -    -    -


____|15s_|____|____|____|____|16s_|____|____|____|____|17s_|____|____|____|____|
AgtA|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtB|ta-    -  talk talk   -  ta-  talk talk AgtE tal_ tal_   -    -    -    -
AgtC|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtD|  -  uhuh uhuh   -    -    -    -    -    -    -    -    -    -    -    -
AgtE|  -    -    -    -    -    -    -    -    -    -    -    -    -  talk talk


____|18s_|____|____|____|____|19s_|____|____|____|____|19s_|____|____|____|____|
AgtA|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtB|  -    -    -    -    -    -    -  uhuh uhuh uhuh   -    -    -    -    -
AgtC|  -    -    -    -    -    -    -    -    -    -    -    -    -    -    -
AgtD|  -    -    -    -    -    -  uhuh uhuh   -    -    -    -    -    -    -
AgtE|talk   -    -  talk pTRP tal_ tal_   .    .    .  ta-   -  ta-  talk talk
```

5.5   A real example

Several interesting segments of recorded conversations from the ICSI corpus of group meetings (Janin et al. 2003) were selected and examined for a qualitative evaluation that ended up not being included in the thesis. By 'interesting' I mean segments in which there was multiple interaction by several participants in short succession (like the discussions simulated here), instead of the usual one-person-talking-for-a-long-time that was prevalent in that corpus. One segment from the first discussion in the corpus (named 'Bdb001') had its timings measured in detail, and serves as an example of the kind of timings, lengths of utterances and silent gaps that can occur in real small group discussions.

This segment is presented below, with labels indicating the lengths, in seconds, of blocks of silence and talk of three conversants, named A, B, C. Silent intervals are indicated in parentheses, both between talk of the same speaker (pauses or hesitations), and as gaps of speaker transitions. The transcript is broken down in 'blocks' to fit the width of the page. Each block may group together simultaneous or quasi-simultaneous talk from more than one conversant put more closely together. Some indentation is used to show roughly when talk begins and overlaps other talk.

Clearly, lengths of silence in parentheses at the beginning of blocks are relative to the longest talk of the previous one. When it is not obvious to which block they refer or when it is relevant to indicate to which speaker the silence is relative, it is marked with it. For example, (S: XX.YY) means this silence is relative to the previous talk of speaker S. Also, when the interval is *negative* to some previous talk (overlapping talk), but has been wrapped down to the next block because of space, the speaker to whose talk it refers is indicated too.

When talk appears indented on the line without any indication to whom it refers, it is to the simultaneous talk of another speaker in the same block, generally just above. At other places, a length of silence is left at the end of one line indicating an amount of pause or hesitation to the next talk of the *same* speaker on the next line.

Lastly, the transcription of this segment has some (intuitive) conventions: words in capitals are emphasized talk (louder) and the dash (-) is self-interruption. Extended vowels (lengthening) were represented by doubling the letters instead of the usual colons (:). Utterances between '○' had lower intensity and/or pitch. There is one very long utter-

ance that was broken up in different blocks: an equal sign (=) at the end of the first was used to indicate that it continues unbroken on the next one, which contains the length of whole utterance.

C : the transcripts $(0.4s)$ that $(0.13s)$ may or may not change and then the
— 1.0*s* — 0.23*s* — 1.51*s*

C : $(0.7s)$ UTTERANCE
0.53*s*

C : $(0.21s)$ which were these tiny boundaries that may or may not change
2.6*s*

A : $(0.15s)$ oh that's-
0.56*s*

A : $(C : -0.24s)$ that's actually very nicely handled here because you could
2.38*s*

C : $(1.08s)$ ermm
0.29*s*

A : $(0.31s)$ you could- all you'd have to change is the $(0.6s)$ ermm
2.14*s* 0.47*s*

A : $(0.44s)$ TIMESTAMPS in the TIMELINE
1.94*s*

A : $(0.38s)$ without $(0.41s)$ without errr changing the idea
0.4*s* 1.66*s*

B : $(A : -0.31s)$ right
0.24*s*

C : $(0.51s)$ and you'd be able to propagate all of the $(0.59s)$
1.81*s*

C : the information
0.79*s*

B : $(A : 0.28s)$ that's ho- that's WHY you do that extra level of indirection
2.44*s*

1.5*s*

B : (2.42*s*) so that you can just change the timeline

1.59*s*

A : (1.34*s*) 'cept the timeline is gonna be HUGE

2.8*s*

A : (0.26*s*) if you say- suppose you have a phone level alignment (0.15*s*)

C : (0.34*s*)   yeah (*A* : 0.73*s*) yeah specially at the PHONE level (0.4*s*)
0.29*s*   1.66*s*

B : (0.23*s*) yees
0.35*s*

2.0*s*

C : (*A* : −0.24*s*) thewe- we have phone level backtresses (1.22*s*)

A :   you'd have- you'd have-
0.95*s*

B :   (*A* : 0.68*s*) yeah this- I don't think I would do this for phone level
2.29*s*

(continues on the next block)

B : (0.2*s*) I think for phone level you want to use some sort of=

C : amm
0.35*s*

4.52*s*

B : binary representation because it'll be too dense otherwise

C :   (*B* : −0.73*s*) ok, sooow
0.79*s*

2.38*s*

C : (0.66*s*) if you were doing that and you had a sort of companion

0.38*s*   1.55*s*

C : (0.28*s*) aaaah (0.15*s*) thing that gets called up for phone level

$$\overbrace{\text{1.21}s}$$

C : $(0.33s)$ °ah what would that look like° $(0.2s)$

$$\overbrace{\text{0.4}s}$$

C : °how would you-°

B : $(0.34s)$ I WOULD use just an existing
$$\underbrace{\qquad\qquad\qquad}_{1.39s}$$

A : $(0.69s)$ °what-°
$$\underbrace{\qquad}_{0.28s}$$

$$\overbrace{\text{1.12}s}$$

B : $(1.47s)$ °an existing way of doing it°

A : $(1.51s)$ but but why not use it for follow- it's just a matter of
$$\underbrace{\qquad\qquad\qquad\qquad\qquad\qquad}_{2.5s}$$

A : $(0.27s)$ it just a matter of $(0.26s)$ uh being bigger
$$\underbrace{\qquad\qquad}_{0.76s}\qquad\qquad\underbrace{\qquad\qquad}_{0.73s}$$

Chapter 6

Evaluation

In order to evaluate the models quantitatively, a procedure was attached to the agent framework to inspect the behaviours in the blackboard at the end of every cycle. Its function was to count a number of relevant *measures* of the resulting discussions, such as the total time of single talk and simultaneous talk, total number of utterances, and so on. This modified version of the simulation would then run large numbers of group discussions, at various parameters and attribute settings, accumulating the results of each iteration to compute their averages at the end.

A total of 150 discussions of 5 minutes each was chosen for this evaluation. 150 is a sufficiently large number and indeed the results showed only minimal variation—of tenths of seconds or of a few number of occurrences—from one evaluation to another, demonstrating that they are reasonably 'stable' averages. And since agents cannot get *tired* or the simulation change with the time of the discussion, 5 minutes is a sufficient length too: as the results proved stable, increasing the length of discussions would hardly change them significantly.

Because the simulation uses random numbers of a pseudo-random nature (as it is the case in any computer program), the 150 iterations were divided in three groups of 50 discussions intercalated by two prompts for a key press from the person running the program. As the random number generator is seeded with the computer clock-time at each 'run', key presses should infuse a little more outside randomness in the chain of pseudo-random numbers for each group of 50 simulations. The randomness is the (variable) micro-second interval up to the actual key press, taken from the CPU clock each time to seed the random number generator.

This chapter will first describe in the next section the *measures* that were counted. Subsequent sections will discuss the results of the various models in a range of parameters and attribute settings. In addition, appendix 3 lists the full set of charts—containing measures omitted here—in a different arrangement for other comparisons.

6.1 Measures

The measures counted in the evaluation comprise two broad groups: total time and number of occurrences. The first group consists of the following:

- average amount of **single talk** in 5 minutes of discussion (from the 150 runs);

- average amount of **simultaneous talk**, both with and without overlapping transitions (overlaps), plus the **mean** length, the **shortest** and **longest** lengths of non-overlap episodes of simultaneous talk (such as from multiple starts of talk, middle starts and long simultaneous talks);

- average amount of **silences** between non-continued utterances, generally speaker transitions (gaps and lapses), and the **mean** length, the **shortest** and **longest** lengths of these intervals[1];

- average amount of **no talk**, which includes the time of silent gaps above, but also the time of *pauses* between (same-speaker) continuing utterances—excluding, however, silent *hesitations* in the middle of utterances, which are not counted as talk either (though non-silent hesitations like "erm" are);

- average amount of **absolute silence**: that is, the previous total of 'no talk' not having any feedback vocalizations either; and

- average amount of total **feedback** and of just '**backchannels**': i.e. only the feedback that came simultaneous with talk, in the background of it. The difference of these two, thus, gives the amount of feedback occurring in the silences between utterances ('no talk'), which has to be the same difference as between the total time of no talk and absolute silence.

It should be noted that the most interesting measures are arguably the first three: the time of single talk, of simultaneous talk without overlaps, and of silent speaker tran-

---

[1] The *shortest* lengths of these and simultaneous talks always ended up being the minimum resolution of the simulation, its clock-cycle: by default, 0.2s.

sitions. The others, though providing interesting information of the make-up of the discussions, are predictable from the design and structure of the simulation and from the parameters (of the attributes) given to the agents:

- the amount of overlaps in speaker transitions is a linear (though complex) function of the INTERACTIVITY and EAGERNESS attributes of the current and the new starting speakers, respectively, modulated by their overall TALKATIVE-NESS: the higher they are, the more overlaps shall occur (and less silent gaps);

- the amount of pauses between same-speaker utterances is also predictable from the VERBOSITY and TRPAUSING attributes—which determine the probability an agent wants to continue speaking and the mean time of its pauses before starting new utterances—again, modulated (strongly) by TALKATIVENESS: speaking agents can only pause and continue talking as long as the others let it by not deciding to talk;

- and the amount of feedback is a direct function of the FEEDBACK attributes of the agents, modulated by their TALKATIVENESS: the less an agent wants to talk, the more it may give feedback. From the amount of feedback derives the amount of non-backchannel feedback (not overlapped by talk) and, thus, the total of absolute silence in the discussions.

These are all measures not straightforwardly determinable by one single parameter or the structure of the simulation, hence providing relevant information about the resulting discussions. Other possible measures such as, for example, the length of pauses and overlaps or the shortest and longest sequences of 'single talk' are either directly determined by a single parameter or randomly created—like the length of utterances, varying randomly from the minimal possible (two cycles of simulation: a 'talk' and a 'pTRP') to a given maximum which can be changed—and will be, in §6.2.2.

The second group (number of occurrences) comprises the following:

- average number of total **utterances** in the 5-minute discussions (averaged from 150 runs), plus the average distribution of utterances amongst the agents in the group, when different TALKATIVENESS parameters are given to each;

- average number of **continuing utterances** of the same speaker after pausing, plus their distribution amongst the agents in the group (these numbers are part of the previous number of total utterances);

- average distribution of the special 'sorts' of pre-TRPs (`More`, `Any?`, `SELECT(X)`) in the extended model with these (§6.4);

- average number of **silent gaps**, **overlaps** and **latches** in speaker transitions (complementary values for this total);

- average number of **single-speaker starts** of utterances, and of **multiple starts** plus the **mean** number of speakers there[2];

- average number of **middle starts**: when someone started to speak in the middle of another's talk prompted by a middle 'pTRP';

- average number of **simultaneous talks**: which is roughly the sum of multiple- and middle-start episodes, and a few others—though not in a linear way because one simultaneous talk can be composed of various multiple or middle starts in succession, without intervening silences;

- the average number of **long simultaneous talks** (included in the previous total): when the *same* simultaneous speakers persist for a full second or more, plus the **mean** number of speakers there; note that the *longest* found lengths of simultaneous talks (listed in the 'time' measures) may not necessarily be one of these (and often were not), but rather a series of several agents speaking and stopping in succession without intervening silences;

- average number of individual **false-starts**: when agents start to speak but give up shortly afterwards (in less than a full second) because of others talking—again, there may be many more false-starts than the total number of simultaneous talks because more than one false-start (from different agents) may have been (and were) counted per each episode of simultaneity;

- average number of **'collective' false-starts**: when two or more agents start simultaneously (or off by a cycle) but they all stop shortly in less than a second because of each other—with the individual false-starts in such cases counted as just *one* collective false-start, so these two totals are disjoint, complementary;

- and the average number of **incomplete utterances**: when someone was already speaking for one second or more and stopped because of simultaneous talk, usu-

---

[2]Multiple starts were identified as being not only starting on the same simulation-cycle, but also one cycle off: agents who may have started to speak without noticing that others already have, in the immediately previous cycle.

ally from a persisting other initial starter or a middle starter.

So, the threshold of *one second* used to identify the last four measures is then just a simple (though arbitrary) way of distinguishing incomplete utterances from false-starts, and 'long' simultaneous talks from the other (short) simultaneous talks. These measures are further clarified hereinafter with examples.

6.1.1   Output of the results

The evaluation procedure inspected the agent behaviours at the end of every cycle of simulation to identify specific changes such as in the speaker, and events like overlaps, latches or silences, thereby counting totals as the next subsection will show. The procedure maintains along the cycles its own internal states, relative as to who is speaking, who spoke in the last few cycles, and others. This program is listed in appendix B.

There follows below an example of the evaluation results for a group of 5 agents: each with TALKATIVENESS of 0.3, 0.2, 0.4, 0.3 and 0.1 respectively, and EAGERNESS of 0.5 for all. The other parameters are left at defaults: cicle-time of 0.2s (5 cycles per second), maximum utterance length of 4 seconds (varying between 0.4s and 4s), 10% of negative feedback (huh?), and all other attributes at 0.5 for everybody.

```
$ java mtest 5 t32431 e5
Press ENTER, or SPACE+ENTER for nonstop evaluation.
(press ENTER for the randomic seeding)
(press ENTER for the randomic seeding)

Totals: amount of single talk (223.7s), of no-talk/total silence (30.0s/15.8s)
        simult.talk with overlaps (38.9s), feedback/backchannels (71.1s/56.5s)
silent gaps  (80): total (25.5s), mean (0.3s), shortest (0.2s), longest (1.6s)
simult.talks (93): total (30.4s), mean (0.3s), shortest (0.2s), longest (1.8s)
simult.starts(63), mean #speakers(2.4); long sim.talks(4), mean #speakers(2.1)
middle starts(17), false-starts/incompletes(94+24/14), overlaps/latches(24/26)
single starts(21), utterances (107:25+17+30+23+9), continueds (10:2+1+3+2+0).
```

Remember again that these results are *averages* from the generation of 150 simulations of 5 minutes (300 seconds) each, except for the longest and shortest lengths of gaps and simultaneous talks, which are the longest and shortest found in *all* 150 simulations. The mean lengths of gaps and simultaneous talks, mean number of simultaneous starting speakers and speakers in long simultaneous talks are also averaged from all the episodes found in the 150 runs. The two numbers of false-starts are for *indi-*

vidual and *collective* false-starts. And the various numbers summing up to the totals of `utterances` and `continueds` (continuing utterances) are of course those counted for each agent. They are different in this case because agents were given different SMALL CAPS TALKATIVENESS likelihoods.

The `Totals` (times of single talk, silence, etc.) correspond to the averaged amounts of time in the 300-second discussions, and the *numbers* (of silent gaps, simultaneous talks, etc.), to the average number of occurrences in this length of simulation. Times are presented with one fractional digit, but the numbers are truncated to integers for displaying so they lose some decimal precision. That is why some sums do not match up exactly, like the sums of individual agent utterances to the total of utterances, and the total of 'continued' utterances from the individual agents (in the last line of results).

Another run of the evaluation with the same parameters shows how little the results change, which demonstrates they are sufficiently stable averages:

```
$ java mtest 5 t32431 e5
Press ENTER, or SPACE+ENTER for nonstop evaluation.
(press ENTER for the randomic seeding)
(press ENTER for the randomic seeding)

Totals: amount of single talk (223.0s), of no-talk/total silence (30.2s/16.1s)
        simult.talk with overlaps (39.1s), feedback/backchannels (70.8s/56.3s)
silent gaps  (80): total (25.7s), mean (0.3s), shortest (0.2s), longest (1.6s)
simult.talks (92): total (30.5s), mean (0.3s), shortest (0.2s), longest (2.0s)
simult.starts(64), mean #speakers(2.4); long sim.talks(4), mean #speakers(2.1)
middle starts(17), false-starts/incompletes(94+24/14), overlaps/latches(24/25)
single starts(21), utterances (107:24+17+30+24+9), continueds (10:2+1+3+2+0).
```

6.1.2 Counting the measures

Next, let me exemplify how some of the measures are identified, via a 'verbose' mode of the evaluation used originally for debugging. This mode reports when the *number* measures are counted in a cycle-by-cycle basis. The counting of total times is not shown because they were trivial and posed no debugging problems—moreover, listing them every time would clutter the output too much. It was the number measures that required a more thorough check since they involved complex context identification, sometimes along a span of two or more cycles.

Here is a sample of the cycle-by-cycle evaluation indicating the measure count (it is slightly edited to fit in a shorter width):

```
$ java mtest 5 +2 t2
Press ENTER, or SPACE+ENTER for nonstop evaluation.

Time  AgtA|AgtB|AgtC|AgtD|AgtE|
      ------------------------ ____identified measures____
0.0    -    -    -    -    -
0.2    -    -    -    -    -
0.4    -    -    -    -   talk  gap_shortest gap_longest gap_sum+=2 singlestart
0.6    -    -    -    -   talk
```

After an initial silence and when someone starts to speak (which ends that silence), some measures are counted: the shortest and longest silent gaps (so far), the number of silent gaps (implicitly in the gap_sum increasing) and that of single starts of utterance (singlestart). Further onwards:

```
5.2   talk  -    -    -    -
5.4   pTRP  -    -    -    -
5.6   tal_  -   uhuh  -    -    utterance
5.8    -    -   uhuh  -    -
6.0    -   talk  -   talk  -    gap_shortest gap_sum+=1 simstart_many+=2
6.2    -   ..    -   talk talk  false-start
6.4    -    -    -   talk talk  simstart_many+=1
6.6    -    -    -   talk  ..   false-start simtalk_longest simtalk_sum+=3
6.8    -    -    -   talk  -
7.0    -    -    -   pTRP  -
7.2    -   talk  -   tal_  -    utterance overlap
7.4    -   talk  -   tal_  -
7.6    -   talk  -    -    -
7.8  uhuh  pTRP  -    -    -
8.0  uhuh  tal_  -   uhuh  -    utterance
8.2    -   tal_  -   uhuh  -
8.4    -    .    -    -    -
```

Now after an utterance is identified by a 'pTRP' behaviour not followed by 'talk' (which would indicate a *middle* TRP), there is a short gap of 0.2s until someone starts to speak again (never mind the uhuh's, which make it a case of 'no talk' but not absolute silence). The gap is identified as the shortest yet found (gap_shortest) and its number and time counted (gap_sum). The same cycle sees two agents starting: a case of multiple starts (counted and) indicated by the number of multiple starters accumulated for the mean (simstart_many).

Right in the next cycle another agent starts, oblivious of the others. One of the first to start then stops immediately with the simultaneous talk, being recognized as a false-start. Then the third starter who began a little later is next recognized in the additional simstart_many. When the simultaneous talk ends with a second false-start, it is recorded as the longest so far (simtalk_longest) and its length is cumulated in simtalk_sum. Later, an overlap is identified besides a finishing utterance.

```
10.4  -   talk  -    -    -
10.6  -   pTRP  -    -    -
10.8  -   tal_  -    -    uhuh  utterance
11.0  -   tal_  -    -    uhuh
11.2  -   .     -    -    -
11.4  -   .     -    -    -
11.6  -   talk  -    -    -     continued singlestarts
11.8  -   talk  -    -    talk  nosingle simstart_many+=2
12.0  -   talk  -    -    talk
12.2  -   talk  -    -    talk
12.4  -   talk  -    -    talk
12.6  -   talk  -    -    talk
12.8  .   talk  -    -    ..    incomplete longsimtalks_many+=2
```

This time, a continuing utterance (continued) is identified after a pause and taken as a single start (singlestarts). But that is corrected—decremented from the counter—in the next cycle (nosingle) when someone else begins an utterance too, as yet unaware of the other speaker. A case of multiple start is then identified as noted by the number of speakers accumulated for the mean (simstart_many). The pause between utterances, earlier, was not counted as silent gap but as 'no talk'. Further, after some cycles, the second speaker gives up, recognized as an incomplete utterance rather than a false-start because it lasted for a second or more. It is equally recognized as a 'long' simultaneous talk—whose number of speakers is counted (longsimtalks_many)—and also as general simultaneous talk (not indicated but implied).

```
$ java mtest 5 t6 c3 +2
Press ENTER, or SPACE+ENTER for nonstop evaluation.

Time  AgtA|AgtB|AgtC|AgtD|AgtE|
      ------------------------ ____identified measures____
    :
    :
17.4  -   talk  -    -    -
17.6  -   talk  -    -    -
17.8  -   pTRP  -    -    -
18.0  -   .     -    -    -     utterance+=1
18.2  -   .     -    -    -
18.4  -   .     -    talk -     singlestart
18.6  talk  -   talk talk talk  nosingle simstart_many+=4
18.8  talk  -   pTRP ..   ..    false-start false-start
19.0  talk  -   talk talk talk  middlestart middlestart simstart_many+=4
19.2  ..    -   ..   talk talk  false-start false-start
19.4  -   -    -    ..   ..     false-st.. false-st.. botched(-4) simtalk_sum+=5
19.6  -   -    -    -    -
```

In this last, jumbled-up segment, four agents start almost simultaneously after an utterance and a short gap, then two of them stop but restart immediately after an early middle 'pTRP', which identifies the new starts as being middle starts. They all soon

end up stopping because of each other's simultaneity, thus identified as a 'collective' false-start, which appears as '`botched`' in the last line of output (slightly abbreviated to fit on the page). The negative count (-4) is the number of individual false-starts of which the episode comprises, subtracted from that counter. Finally, a total of one second is counted in `simtalk_sum` at the end of the simultaneous talk. It is not identified as a *long* simultaneous talk because it was not one continuous episode, but rather a compound of short ones—though its length might have been recorded as the longest found (not in this case).

This example demonstrates the non-trivial relationship between the number of simultaneous talks in general and other related 'numbers'. On a first thought, the total of simultaneous talks counted by the evaluation would seem to be just the sum of simultaneous starts and middle starts. That is not necessarily the case as this last example shows. There, only *one* episode of simultaneous talk occurs, being formed by *one* simultaneous start and *two* (agents performing) middle starts.

One other case of simultaneous talk that is not a multiple start or a middle start of utterance is when an agent responds to negative feedback ('`huh?`') by (re)starting to speak, when there are other(s) already speaking. This generally causes another occurrence of simultaneous talk.

6.2   The basic model

The basic model (§4.4.2) was evaluated in four small group sizes, from 3 to 6 agents, in a range of parameters for TALKATIVENESS as the primary attribute and in a few settings of EAGERNESS. The effects of two other secondary attributes (CONFIDENCE and INTERACTIVITY) as well as the maximum utterance length were also verified, but focusing only in one or two group sizes (3's and 5's).

In all the charts of this chapter (and those of appendix 3), the horizontal axis indicates the range of TALKATIVENESS parameters given to the agents: the same in all evaluations of this and the next section, but reduced in sections 6.4 and 6.5. The vertical axis, clearly, gives the scale of the measures being displayed: either amounts of time in seconds, or number of occurrences. The variation in a secondary attribute is indicated by a number of shaded bars (instead of just one) for each setting of TALKATIVENESS, as indicated by the legends beside the charts.

The range of TALKATIVENESS parameters consists of two sets. In one, the *same* likelihood was given to all agents in the group, from 0.1 to 0.8, indicated by single '.1' to '.8' labels on the horizontal axis of all the charts. These are intercalated by sets of *different* likelihoods given to each agent in the group, indicated by vertical series of dot-digit labels in the horizontal axis. The lower ones in the charts for 4 or more agents are truncated, so the full sets of different TALKATIVENESS parameters used in the evaluations of this and the next section are listed below—for the sake of clarity only by their decimals (e.g. '112' meaning agents set to 0.1, 0.1 and 0.2):

- for 3-agent groups: 112, 223, 324, 435, 546, 657, 768, 879;

- for 4-agent groups: 1122, 2312, 3243, 4354, 5465, 6576, 7687, 8798;

- for 5-agent groups: 11213, 23123, 32431, 43542, 54653, 65764, 76875, 87989;

- for 6-agent groups: 112132, 223231, 324341, 435421, 546532, 657647, 768759, 879896.

My intention with these sets of different likelihoods was to alternate TALKATIVENESS gradually, not abruptly, like people in real groups would probably measure up to: e.g. from 0.4 to 0.3 to 0.5 to 0.6, instead of 0.4 to 0.2 to 0.8 to 0.1, etc.

It is possible that high TALKATIVENESS likelihoods of 0.7 or up could be deemed unrealistic, unrepresentative of real conversant behaviour. After all, people in small group discussions in general do not want to talk 70% or more of the time at every possible opportunity. Nevertheless, those high values should provide a fuller idea of the model's behaviour in a wide range of probabilities.

6.2.1 Eagerness variation in the small group sizes

The basic model was first evaluated in four small group sizes and in five settings of EAGERNESS for everybody, in the range of TALKATIVENESS parameters described above. The other attributes were given a middle likelihood of 0.5 for all agents, and a maximum utterance length at the default of 4 seconds.

Figure 6.1 presents the averaged amounts of single talk for the 3-, 4-, 5- and 6-agent groups. They peak at about 225s in the 300-second simulations, which corresponds to 75% of the total discussion time that the basic model generated as 'proper' one-at-a-time talk. These totals decrease gradually as TALKATIVENESS increases and at the

Figure 6.1: Total time of single talk in the small group sizes
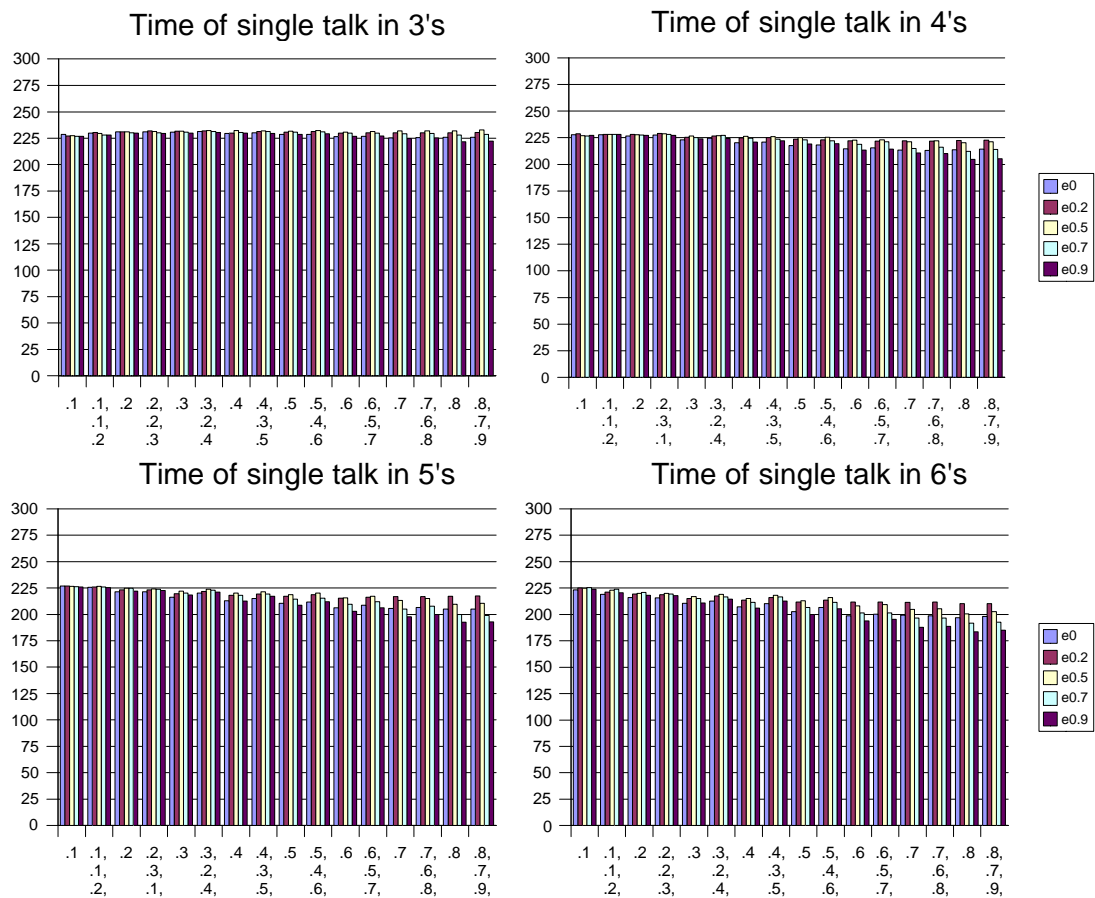
extremities of EAGERNESS, with the trend being more pronounced in bigger groups. They drop most by 42s, down to 183s in the last two TALKATIVENESS settings for 6 agents. The one thing that could change these levels of single talk significantly is the overall maximum length of utterances generated in the discussion, as the next subsection will show (with a change in the maximum utterance length parameter).

The different EAGERNESS values yield an interesting point. This attribute ranging from low to high characterizes agent turn-taking behaviour from *polite* to *eager-to-talk*: i.e. whether someone wanting to talk at each TRP either waits for the current speaker to stop, or starts immediately regardless. The peaks of single talk time at middle values of this attribute and on the higher range of TALKATIVENESS and group size, though small, demonstrate that when turn-taking decisions (*when* to start, in this case) are *different* each time, *the chances of one speaker going first* and taking the next turn at talk increase, consequently keeping the levels of single talk from falling.

Because middle likelihoods for all agents mean their decisions *will vary the most*, so they tend to be different each time. For EAGERNESS, this reduces multiple starting clashes and increases the chances of one agent starting first, and alone, each time.

That is exactly what transpires in the charts of figure 6.2, which present the total time and number of occurrences of simultaneous talk other than overlapping transitions. Time and number are both smaller in the middle values of EAGERNESS, and the trend is more pronounced at higher TALKATIVENESS and group sizes. Although the biggest drop in simultaneous talks (between 0 and 0.2 EAGERNESS) is of only 4s and 11 occurrences with 3 agents, it can be of 14s and 20 occurrences with 6 agents.

Particularly in the larger groups, indeed, simultaneous talk jumps high with zero EA-GERNESS. When all agents always wait for the speaker to finish talking so they can speak, they all start at the same time afterwards, causing much more simultaneous starts (figure 6.7) and, thus, simultaneous talks. EAGERNESS has then an important role in the simulation: to enable different turn-taking timings—though they are just two here—, which significantly reduces simultaneous starts and improves the flow of the interaction. This comes in line with general observations from the empirical literature, such as 'the first to start gets the turn' (Sacks et al. 1974).

Total amounts of simultaneous talk were between 8–35s (2.7–11.7% of the total) with 3 agents, 13–49s (4.3–16.3%) with 4, 18–61s (6–20%) with 5, and 22-70s (7.3–23%) with 6 agents, along the TALKATIVENESS range. Together with number of occurrences they indicate that their mean length remained very short: 0.3s for 3-agent groups, between 0.3s and 0.4s in 4-agent groups, and up to 0.5s in 5's and 6's. It increases in bigger groups because more people wanting to talk at the same time create more simultaneities composed in succession of smaller clashes (as in the end of the previous section), pushing the averages up a little. Actually, the mean length of simultaneous talks is directly affected by the CONFIDENCE of speakers: increase this attribute and the overall length of simultaneous talks should increase too (as in §6.2.3).

The longest episodes of simultaneous talk found in 150 iterations ranged between 1.4–2.2s with 3 agents, 1.4–2.8s in 4's, 1.6–3.6s in 5's, and 2.0–3.8s in 6's, again, along the TALKATIVENESS parameters. These can (and probably were) compounds of shorter simultaneities in succession. The shortest occurrences were always at the minimum possible: 0.2s. Hence, there was a majority of very short simultaneous talks with longer ones in a decreasing proportion (probably exponentially).

Figure 6.2: Time and number of simultaneous talks in the small group sizes
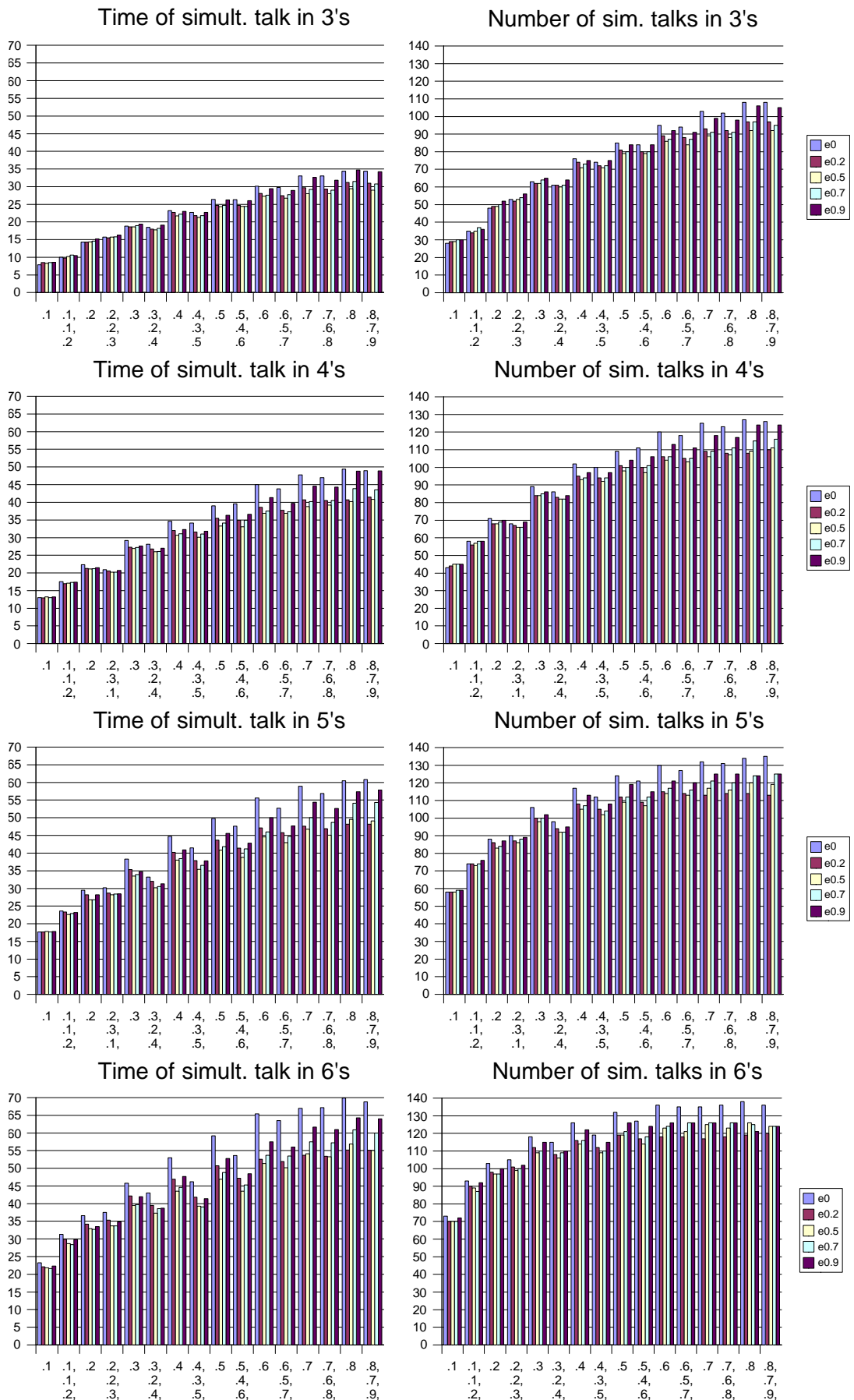
It is also noticeable how groups that were given *different* TALKATIVENESS likelihoods generated smaller totals than those whose agents were all given the nearest same value—one of many such cases to come. The slightly differing TALKATIVENESS likelihoods appear to *drag down* the results of many measures we will see, more visibly in the middle range between 0.3 and 0.6. This is due no doubt to the decreasing bias in the values chosen, yet observing how a group with TALKATIVENESS of (say) 0.4–0.3–0.5–0.4 generated meager totals (of various measures, not only simultaneous talk) than one with all 0.4 likelihoods, the conclusion is that being *less* talkative has a greater (negative) effect in the net results than being equally *more* talkative. Less talkative people increase the possibilities of single starts and single talks, whereas more talkative ones just add to the group of already simultaneous speakers.

The total time and number of silent gaps (also lapses) are presented in figure 6.3. As expected, increasing EAGERNESS progressively reduced the number and therefore the total time of silent gaps, since they were replaced by latched or overlapped transitions (figure 6.4). What is new is that this reduction diminishes significantly to the point of becoming minimal past middle EAGERNESS values, more and more as TALKATIVENESS and group size increase. The reason is that silent gap occurrence is only indirectly affected by EAGERNESS: higher likelihoods may increase the frequency of latches and overlaps (turn-takings leaving no gaps), but this in turn increases simultaneous clashes which will cause false-starts and, before and after them, silent gaps.

Their resulting profile is then very similar in all group sizes, reaching equilibrium limits on both extremities of EAGERNESS as TALKATIVENESS increases. With zero EAGERNESS, the number of silent gaps actually *increase* to a maximum above 120 occurrences: as agents always wait for silence before speaking, the number of (minimal) gaps is then much greater. In other likelihoods, they are replaced by overlaps or latches, decreasing gradually to below 80 (EAGERNESS at 0.2) or 50 occurrences (at 0.5 or more) along the TALKATIVENESS range.

Total time varied from about 35s to below 15s, comprising between 11.5% and 5% of the total time of the discussions, and the number of occurrences varied from 130 to less than 50 at high EAGERNESS for all group sizes. The mean length of silent gaps was thus between 0.4–0.2s; 0.2s only at the highest values of TALKATIVENESS with 4 and 5 agents, and at the middle to high range with 6 agents. The longest occurrences of silent gaps were between 4.2–0.8s along the range of parameters with 3 agents, down to 3.2–0.6s with 4 agents, and 2.6–0.4s with 5 and 6 agents.

Figure 6.3: Time and number of silent gaps in the small group sizes
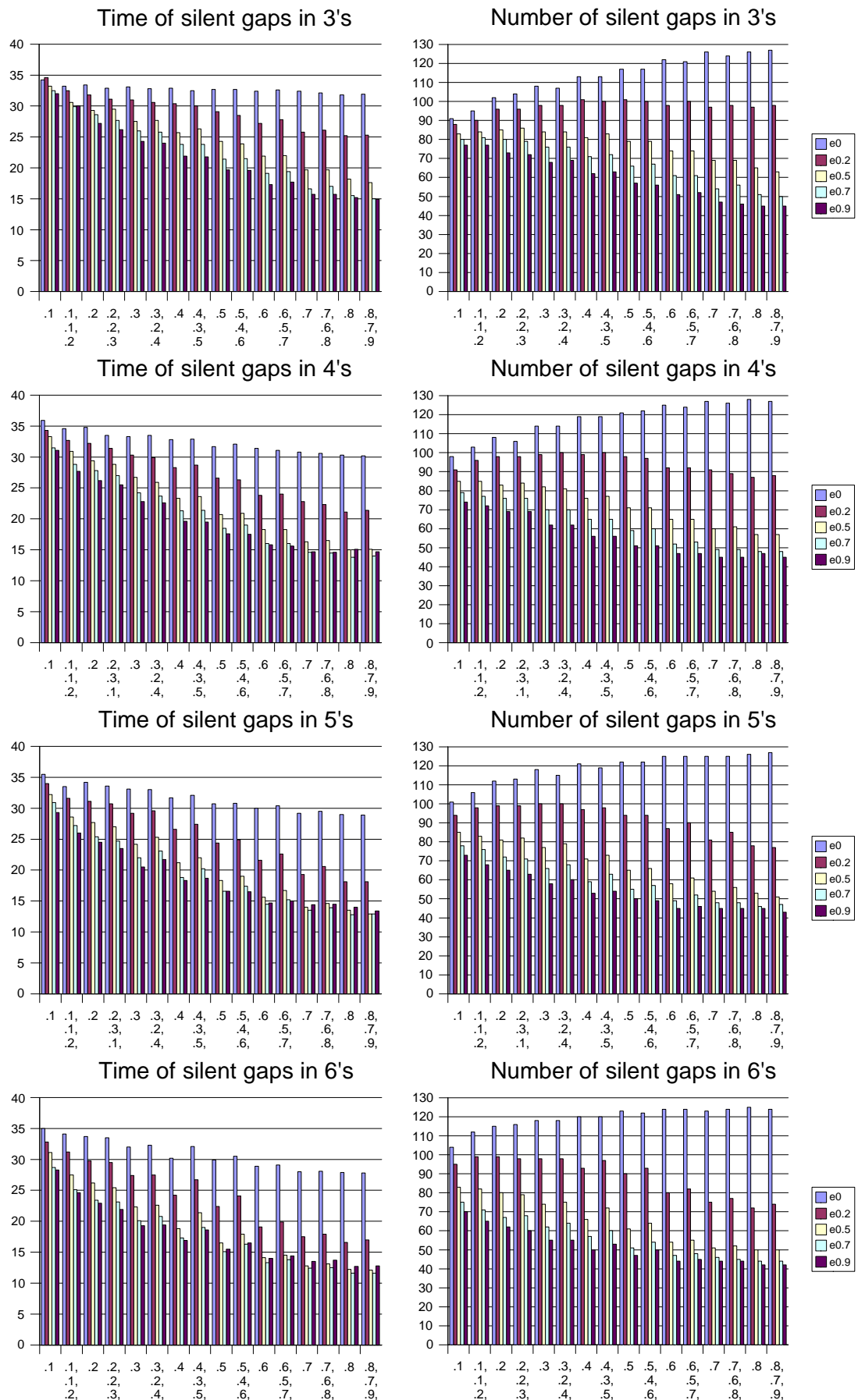
Figure 6.4: Number of overlaps in the small group sizes
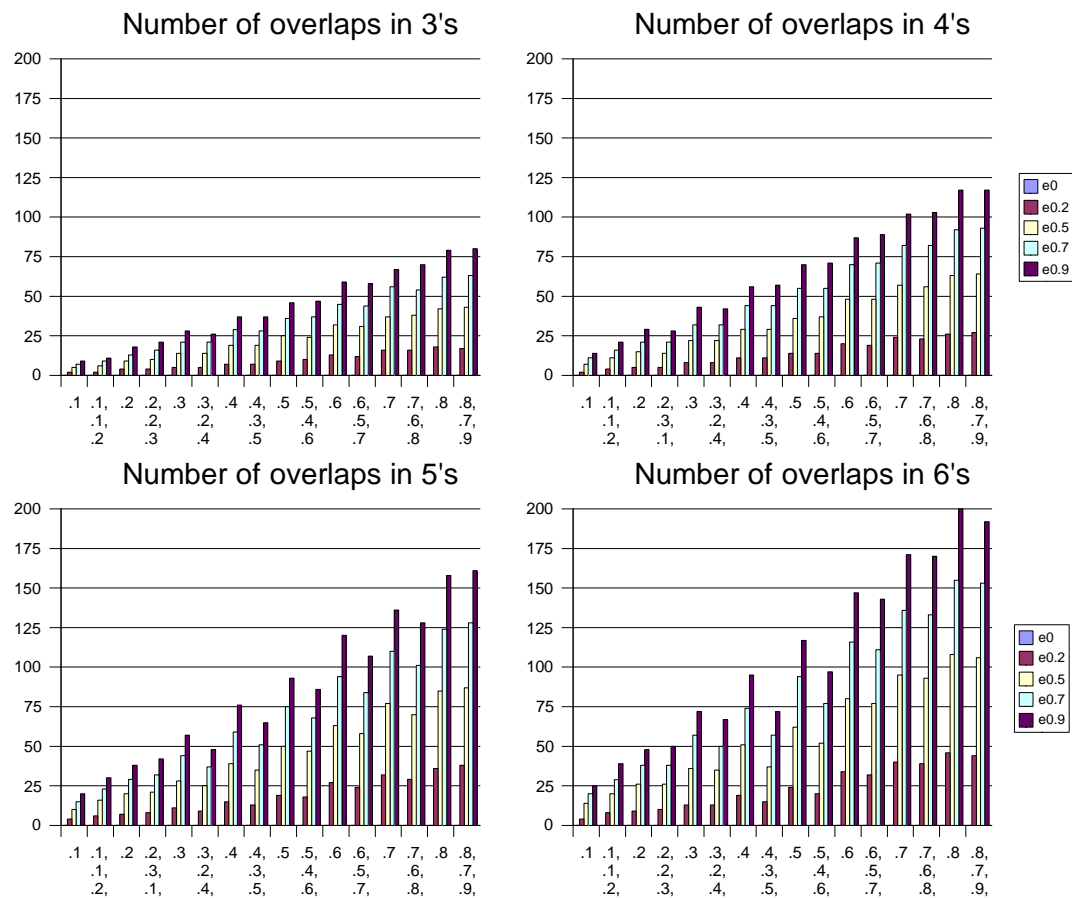
Figure 6.4 shows the increase in the number of overlapped speaker transitions as EA-GERNESS likelihoods of all agents are greater than zero. Besides overlaps, roughly the same number of latches also occur as the attribute increases. There is no need to chart them since they would be pretty much the same, since the INTERACTIVITY attribute of all agents was set at 0.5: meaning that in half the time that an agent produced an utterance, it would leave a trail of finishing talk—which is the simple way that overlaps can happen in this simulation—and in the other half, it did not. Changing this attribute, evenly or unevenly for the agents, would change the proportions of latches and overlaps equivalently when EAGERNESS is greater than zero (as in figure 6.24).

The more overlaps and latches the less silent gaps, but the decrease is not directly proportional as figure 6.3 shows: silent gaps tend to stabilize on a minimum of 45–40 occurrences at the higher end of the parameter range, whereas the number of overlaps (and latches) soar to 100 and more in the larger groups. This is because *more than one*

latch and overlap can be counted in simultaneous talk each time, at the higher values of TALKATIVENESS, when one or more (simultaneous) utterances are overlapped or latched by one or more other starting ones. On the other hand, at any time there can only be *one* silent gap until one or more agents start to speak—no such thing as 'simultaneous gaps'. No matter how high are TALKATIVENESS and group size, there will always be a minimum number of silent gaps at the minimum length possible.

The charts of figure 6.5 present the total numbers of all (complete) utterances, including there simultaneous and continued ones, besides the totals for just the latter in the right column. Number of utterances in 5 minutes of discussion increases from a minimum of 91–97, when all agents have TALKATIVENESS of 0.1, to 125–159 at the high side. These results are for a maximum utterance length of 4s: increase this and number of utterances should drop (figure 6.12 in the next subsection).

Higher EAGERNESS for all agents affects the totals (not of continuing utterances) in a small proportion on the lower side, but gradually more on the higher side of TALKATIVENESS and in larger groups (more about why in figure 6.6). The biggest increases are between 0.5–0.7 EAGERNESS, more visibly in high TALKATIVENESS values, indicating these likelihoods generate slightly more 'productive' discussions—not necessarily always like real ones though.

Continuing utterances may occur when nobody speaks after someone finishes an utterance, so that the same speaker is able to start another if deciding so. With VERBOSITY at 0.5 for all agents (which means that half the time speakers *will* want to continue), the results show an increasingly tinier proportion of continuing utterances in relation to the total, disappearing or becoming minimal in high TALKATIVENESS and group size. Their share of the total is 44% on the lower side of 3-agent groups, falling to 30%, then 20%, 13% and below. In larger groups, the initial proportions (at 0.1 TALKATIVENESS) are 33% in 4's, 25% in 5's and 20% in 6's.

These diminishing shares of continuing utterances from the total are that much strong because (besides the random nature of speaking decisions) in the basic model speakers have no way of securing the floor through what they say in order to continue talking, other than hoping others would not start. This is remedied somewhat in the third model which incorporates 'More' utterances representing those whose contents make it clear the speaker has more to say (§6.4).

Different TALKATIVENESS likelihoods for the agents in the group mean they end up

Figure 6.5: Number of utterances and continuing ones in the small group sizes
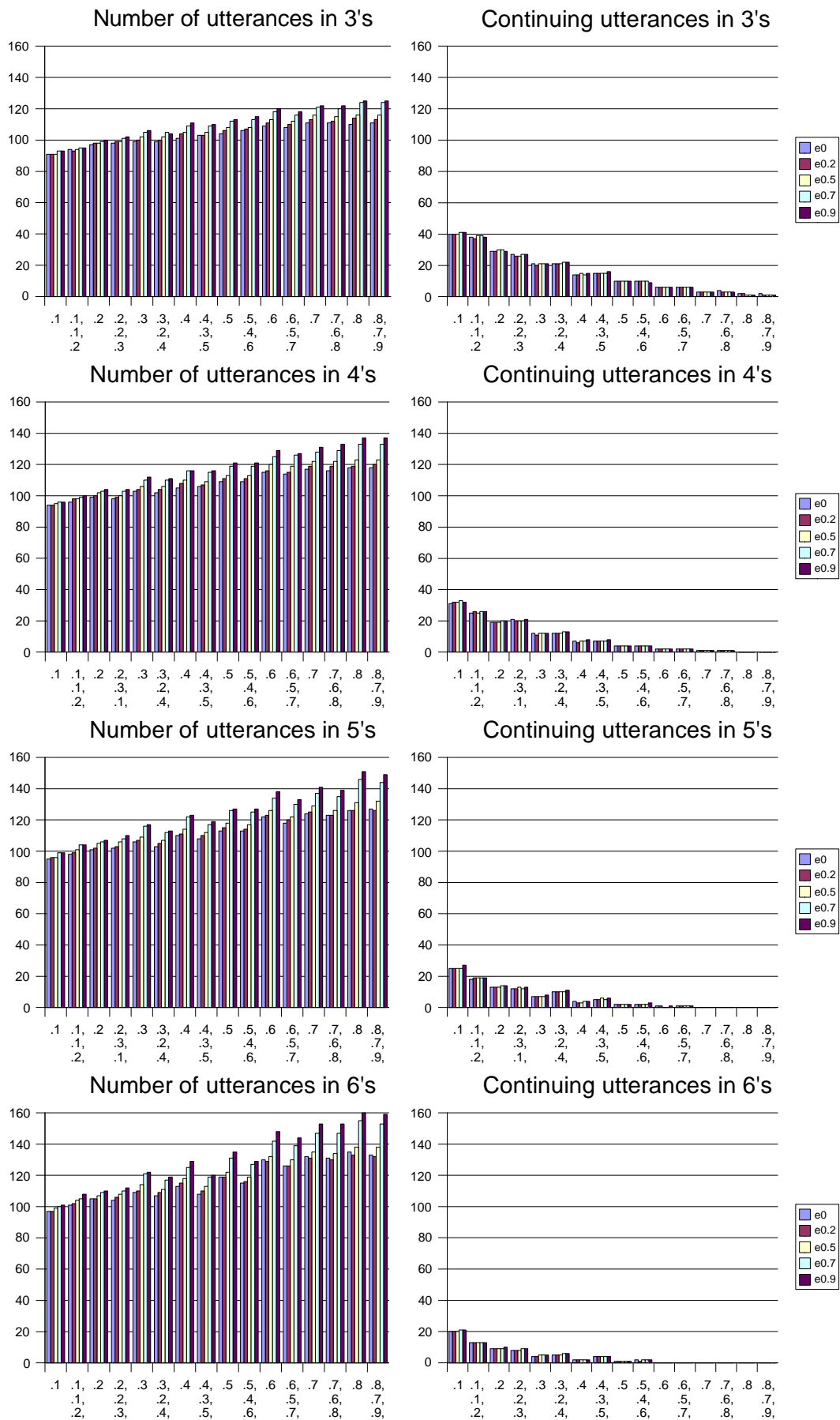
Figure 6.6: Utterance distribution amongst the agents in the small group sizes
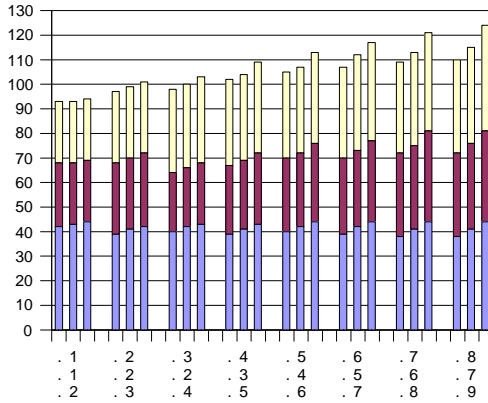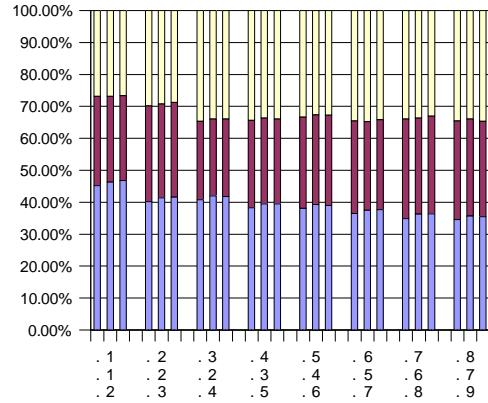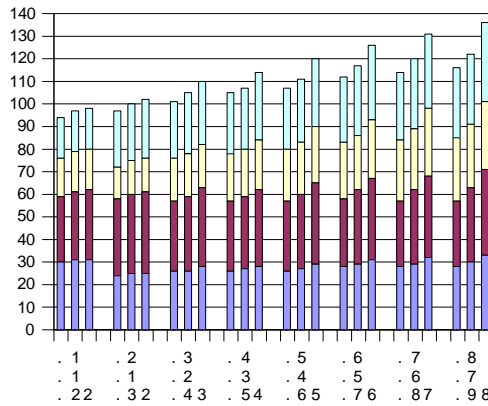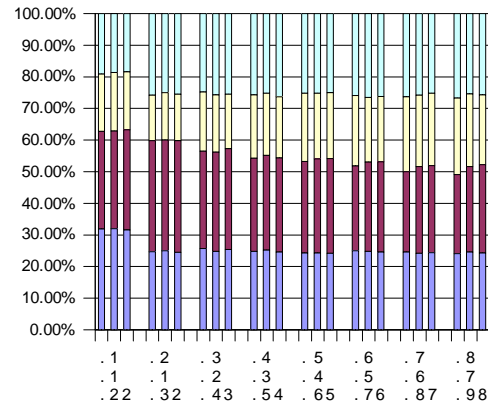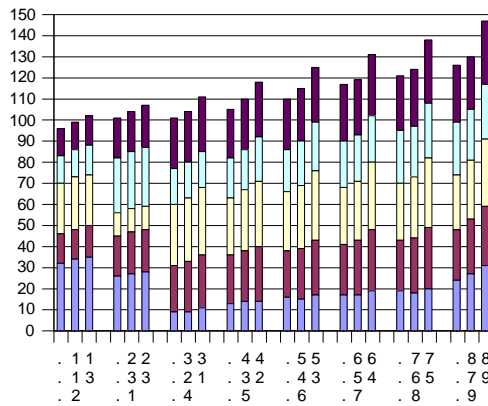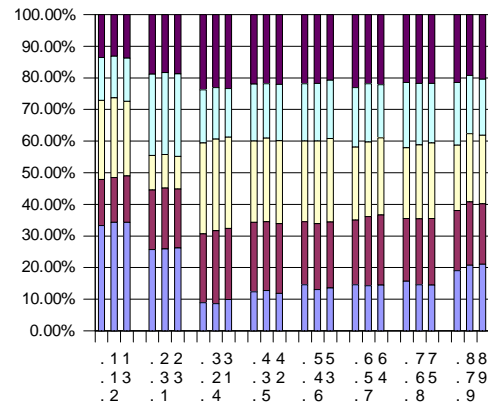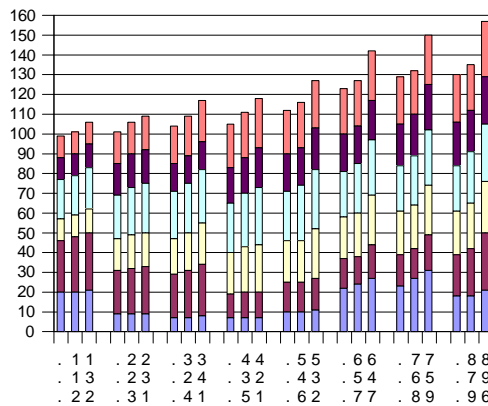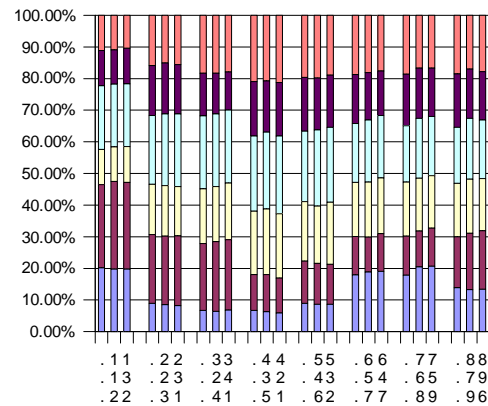
producing different numbers of utterances, as detailed in figure 6.6. The charts on the left display the shares in number of utterances for each agent in the group, and those on the right the same shares in percentages of their totals. They are all aligned vertically with the respective TALKATIVENESS parameters arrayed in the horizontal axis. With 4 or more agents, the order goes down vertically one column of dot-digit values then in a second one: e.g. the first set for 5 agents is ordered 0.1, 0.1, 0.2, 0.1, 0.3, aligned from top to bottom to the chart bars.[3]

The increases in the individual and total number of utterances can be seen not only in the rising TALKATIVENESS but in three EAGERNESS values: the three bars at each setting in the charts of figure 6.6, that correspond to 0, 0.5 and 0.9 EAGERNESS for all agents, respectively. They demonstrate the effectiveness of starting first in the 'competition' for the floor (Sacks et al. 1974): the higher this likelihood the more utterances that agents were able to produce, indicating more successful turn-takings.

Not only in absolute numbers, as can be seen easily in the left charts, but also in the proportions of each agent in the right ones. They increased (though minimally) on those with highest TALKATIVENESS and decreased on those with lowest. So the more talkative (and eager) agents *took more of the share of the less talkative ones*, even those with high EAGERNESS as well. This occurs more with less agents (e.g. with 3 and 4) and the more is the TALKATIVENESS difference between them. For example, the share increases and reductions (as EAGERNESS rises) are more accentuated in the first TALKATIVENESS values for 3 agents (0.1, 0.1, 0.2), though only by a few percent points, because one likelihood is the double of the rest. They are less visible in the others and in larger groups.

Figure 6.7 charts the number of simultaneous starts (including *quasi*-simultaneous ones: off by one cycle) and single starts of utterances. As expected, zero EAGERNESS produces a lot of simultaneity, though more single starts too. However, the latter measure diminishes steadily to become just a fraction of the former (or disappear) as TALKATIVENESS and group size increase. Only with 3 and 4 agents is that the number of single starts is greater than multiple starts at the lower side, but they soon get smaller at middle to high TALKATIVENESS. With 5 agents, the two measures begin level (at the low side) but grow in opposite directions from there.

A shortcoming of the basic model is therefore clear here: superficially, it enables too

---

[3]Unfortunately, the order of different TALKATIVENESS values I have chosen did not result in nicely aligned percentage distributions in the case of 5 and 6 agents as it did for 3's and 4's.

Figure 6.7: Number of multiple and single starts in the small group sizes
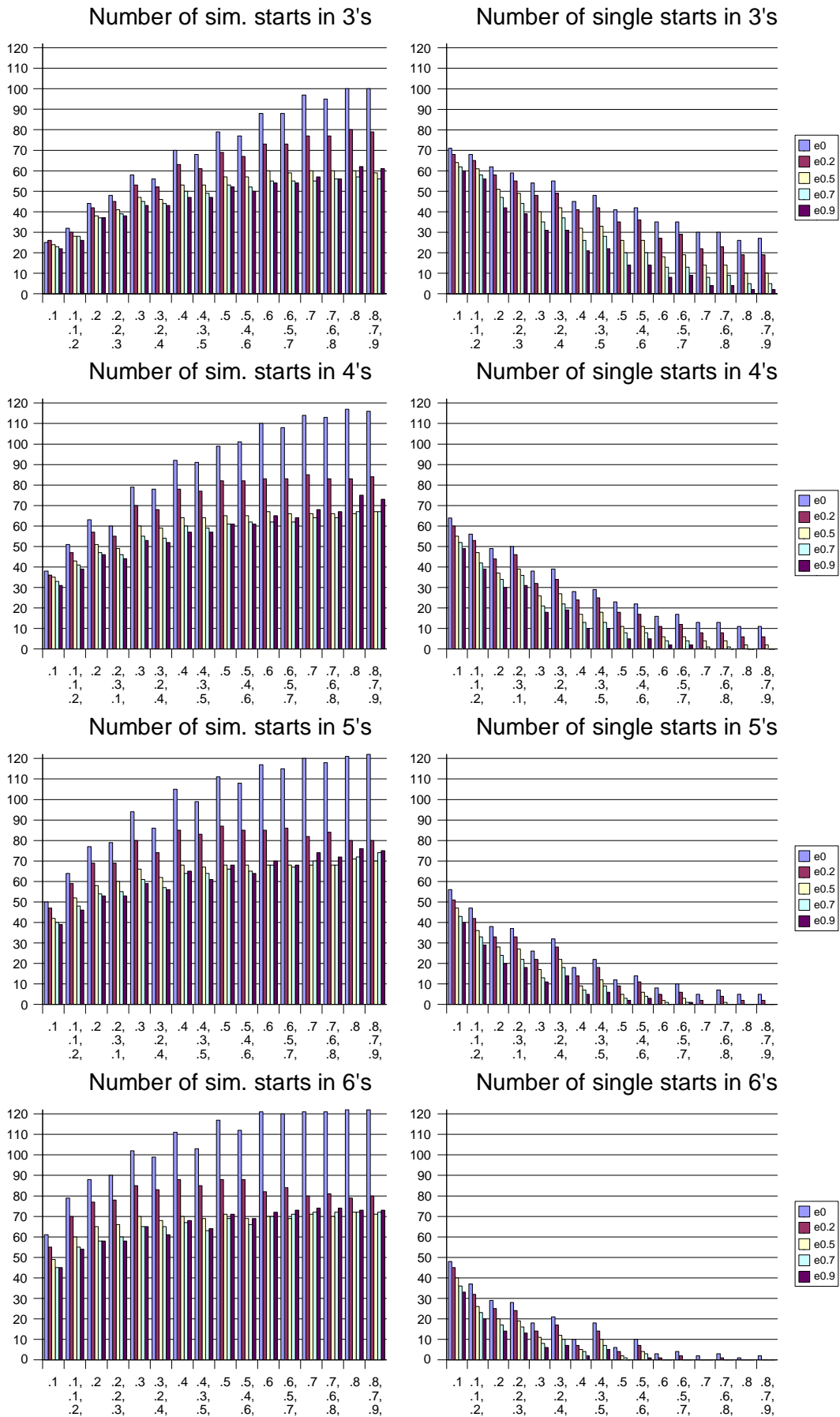
many simultaneous starts and, thus, too few single starts of utterances—even though something like this should be expected in more talkative groups. Notwithstanding the same structural limitations that favour this (no contents of talk, mindless conversants governed by probabilities), the extended models are able to achieve a more balanced distribution (next sections, figures 6.28 and 6.33).

Simultaneous starts fall sharply with any non-zero EAGERNESS specially in higher TALKATIVENESS and group sizes because of the greater competition for the floor. If anything, these falls make it clear again that any variation in turn-taking timings substantially reduces starting clashes. In this case, however, it is only to a certain minimum, in part because there are just *two* timings: speak at a TRP or wait for silence.[4] Hence EAGERNESS likelihoods greater than 0.5 seem to impact no further on the multiple starts; on the contrary, they tend to increase them a little.

This measure indicates the number of *episodes* of multiple starts (such as of single starts), independently of *how many* agents actually started each time. The mean number of speakers in multiple starts goes from 2.0 in 3-agent groups to 2.1 with EAGERNESS at 0.7 and 0.9, then gradually to 2.2, 2.3 and up to 2.7, always at the higher values of this attribute along the TALKATIVENESS range. For the larger groups, the mean number of starters was between 2.1–3.3 along the parameter range with 4 agents, 2.2–4.1 with 5, and 2.2–4.8 with 6: the raises were increasingly greater with more agents. The highest mean of 4.8, for example, corresponds to 4 and even 5 agents starting simultaneously most of the time.

Lastly, two *failure* measures appear in the charts of figure 6.8: the average totals of individual false-starts (short, abandoned starts of talk) and 'collective' false-starts (when the starters all stop because of each other). With group CONFIDENCE at 0.5, the numbers of individual false-starts soar above 100 in 3-agent groups, 200 in 4's, 300 in 5's, and 400 in 6's with high TALKATIVENESS and at the extremeties of EAGERNESS (0, 0.7, 0.9), peaking at 477 with 6 agents. Besides a 'low' CONFIDENCE, the very high totals result from the simple way of resolving simultaneous talk in the basic model, which is substantially improved with more elaborate procedures such as those of the second model (next sections, figures 6.29 and 6.34).

The numbers of collective false-starts converged around 27, with rises as much quicker in bigger groups. They peaked occasionally around 30 at high EAGERNESS—which

---

[4]Yet there is a third one: starts of talk off by one cycle, but which were *assumed* as indistinguishable in determining precedence, thus turn-taking, because of the nearness of just one cycle.

Figure 6.8: Number of individual and of 'collective' false-starts in the small group sizes
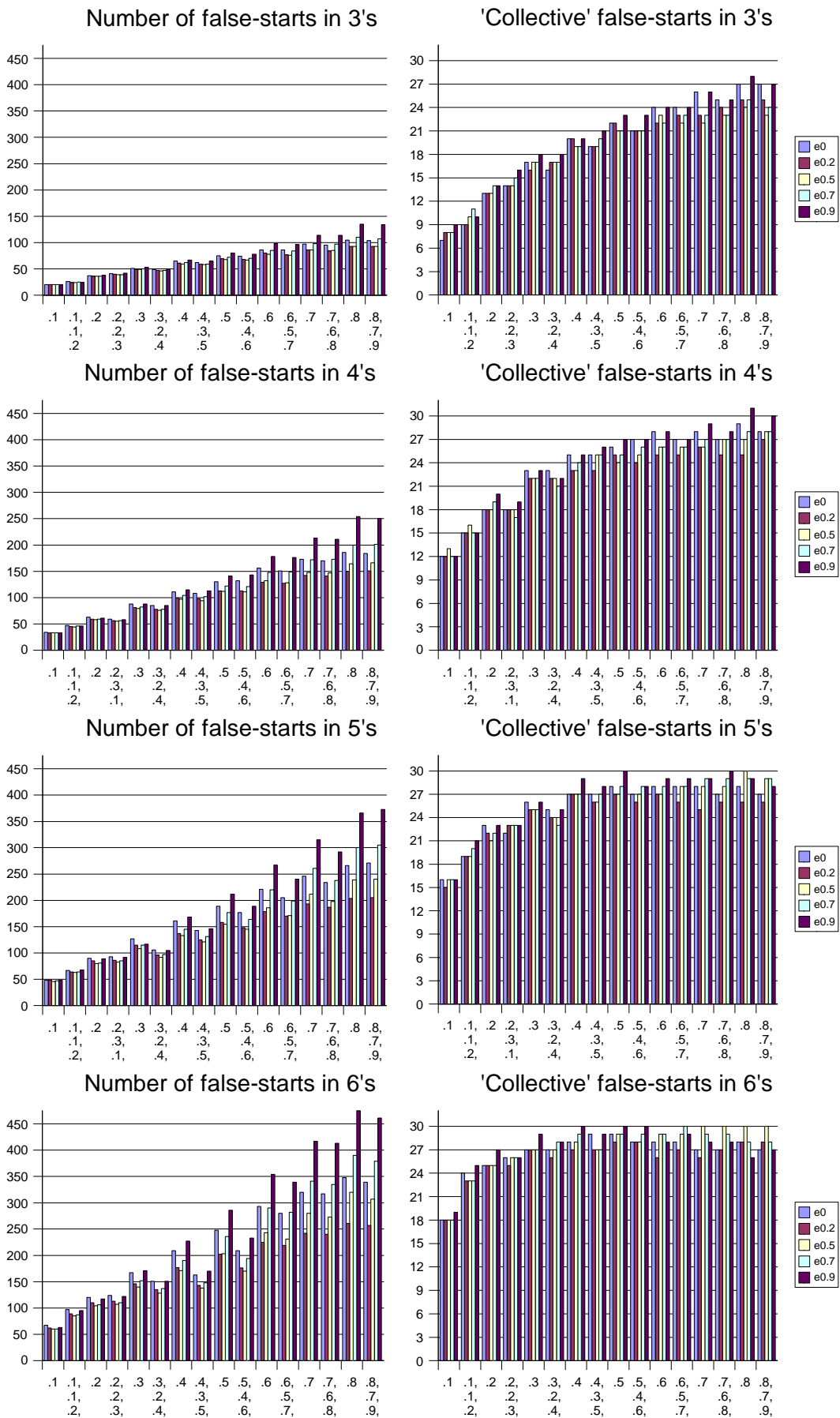
Figure 6.9: Total time of single talk in different maximum utterance lengths
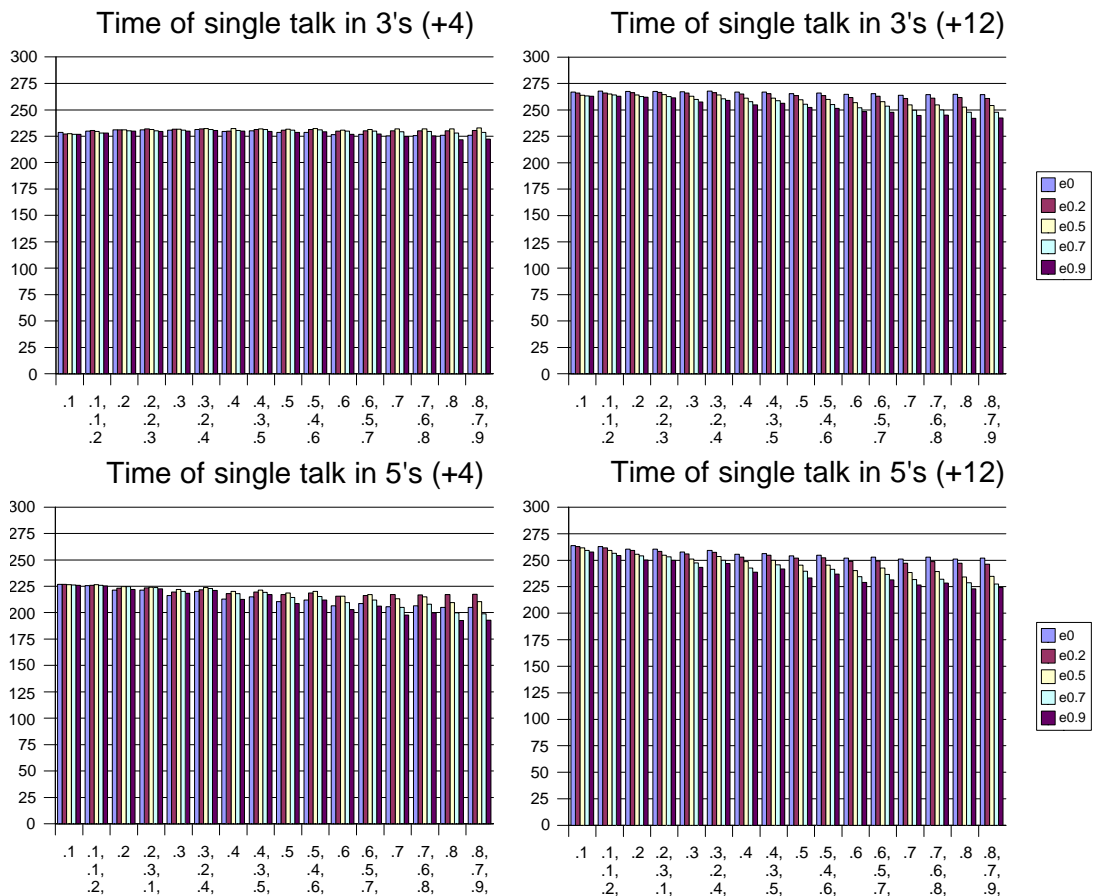
means one each 10–11s in average. Middle EAGERNESS likelihoods also generated fewer totals: both measures then confirming again that when agents vary their turn-taking timings, less starting clashes (and thus, false-starts) would occur.

6.2.2 Longer utterances in 3- and 5-agent groups

Regarding the totals of single talk and a few other measures which 'stabilized' on certain limits, it is useful to see how much an increase in the overall utterance lengths would affect those averages. The charts in this subsection compare the results for 3- and 5-agent groups with the simulation's maximum utterance length parameter increased to 12s from the more 'realistic' limit of 4. 4 seconds is already a long talking span for anyone to sustain in informal situations without any pause for breathing (and without hesitation and fragmentation); 12s then seems awfully long, although it is the

maximum possible: lengths would then vary randomly between 0.4s and 12s.

First, figure 6.9 compares the amounts of single talk for both settings of this parameter, referred to as '+4' and '+12' in the chart titles (their command-line arguments). The charts to the right show that the average amounts of single talk increased some 20–45s, less homogeneously than before, surpassing and around the 250s mark (83% of the total), peaking at 270s at lower EAGERNESS for 3's and lower TALKATIVENESS for 5's. With 3 agents, totals remained between 270–250s, dropping below 250s in a few high values of EAGERNESS and TALKATIVENESS. With 5 agents, they only remained between 260–250s with lower parameters, but soon falling between 250–225s.

The effect of increasing EAGERNESS with longer utterances is now different: single talk time just falls, to a low of 225s in 5-agent groups. With utterances potentially three times as longer, having thus more intermediary TRPs within them, agents more eager-to-talk at more TRPs would cause a progressive increase in the number of *middle* starts of talk (and more incomplete utterances, see figure 6.13). This in turn generates relatively more simultaneous talk, as figure 6.10 shows—more than the reduction in the opportunities *between* utterances would warrant (the effect of the shorter setting).

The charts on the right of this figure (6.10) give the total time and number of simultaneous talks (overlaps excluded) in the longer maximum length parameter. Charts on the left show the previous totals for comparison. Although simultaneous talks are always less with longer utterances, they grow very close to the same totals (particularly in numbers) as EAGERNESS rises. They now just increase with this attribute because of the increase in middle starts coupled with the sheer reduction in the number of utterances (figure 6.12), and consequently, on the opportunities for turn-taking which is the main locus of the simultaneous talks.

The mean lengths if simultaneous talks continued to be short: between 0.3–0.5s. They were only 0.3s in 3-agent groups, increasing to 0.4s midway through 5-agent groups, and reaching 0.5s at a few higher values of TALKATIVENESS and EAGERNESS. This increase is actually visible in the charts on the higher side of TALKATIVENESS for 5 agents: though the *number* of simultaneous talks seems to reach an upper limit, their total *times* keep increasing along the same range, which means the mean lengths rise a notch there, too. Lastly, the longest simultaneous talks that were found now ranged between 1.2–2.1s for 3 agents and 1.4–2.6s for 5: only the latter is smaller than the previous results (it was 1.6–3.6s for 5 agents in the maximum length of 4s).

Figure 6.10: Time and number of multiple talks in different maximum ut. lengths
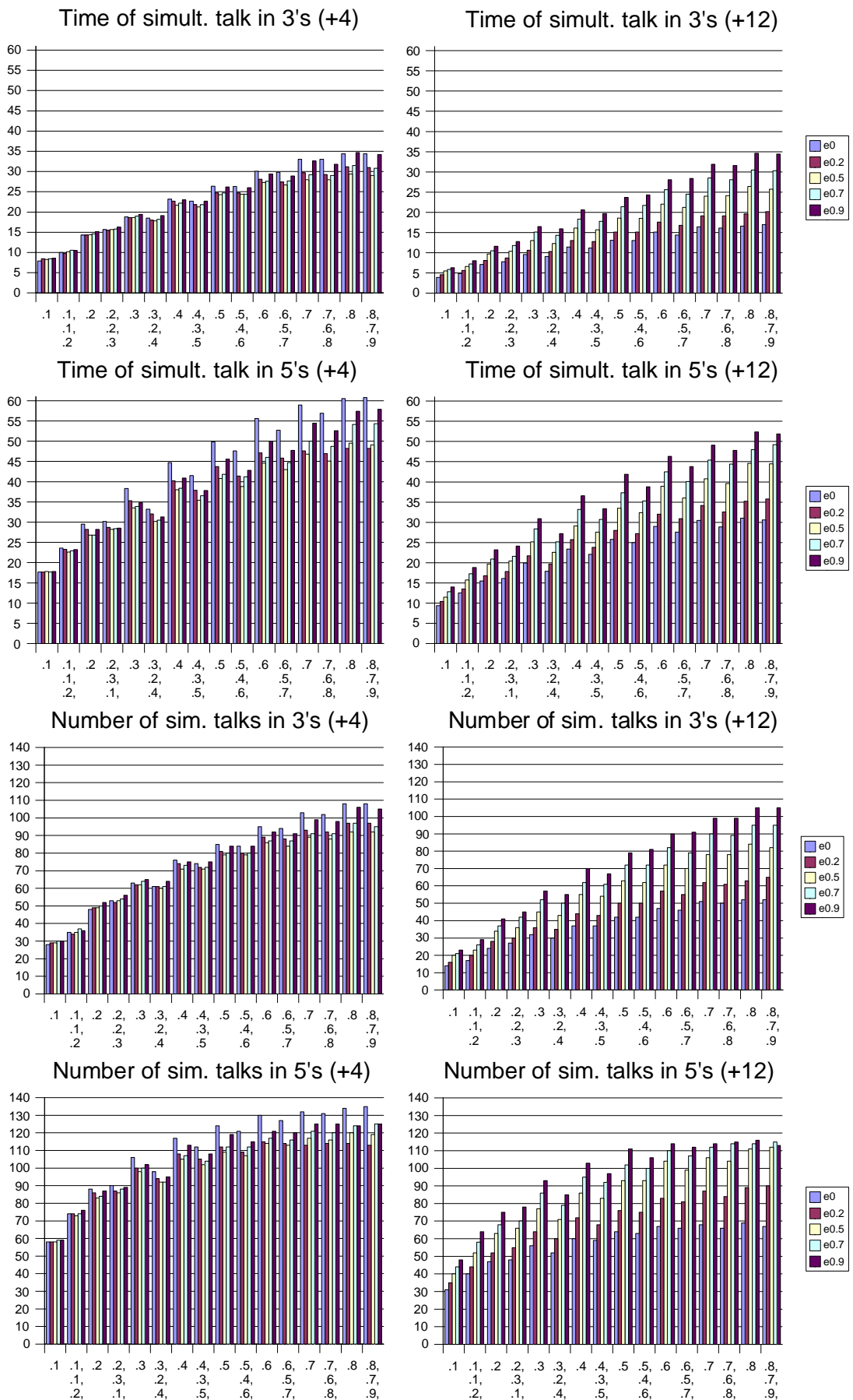
Figure 6.11: Time and number of silent gaps in different maximum utterance lengths
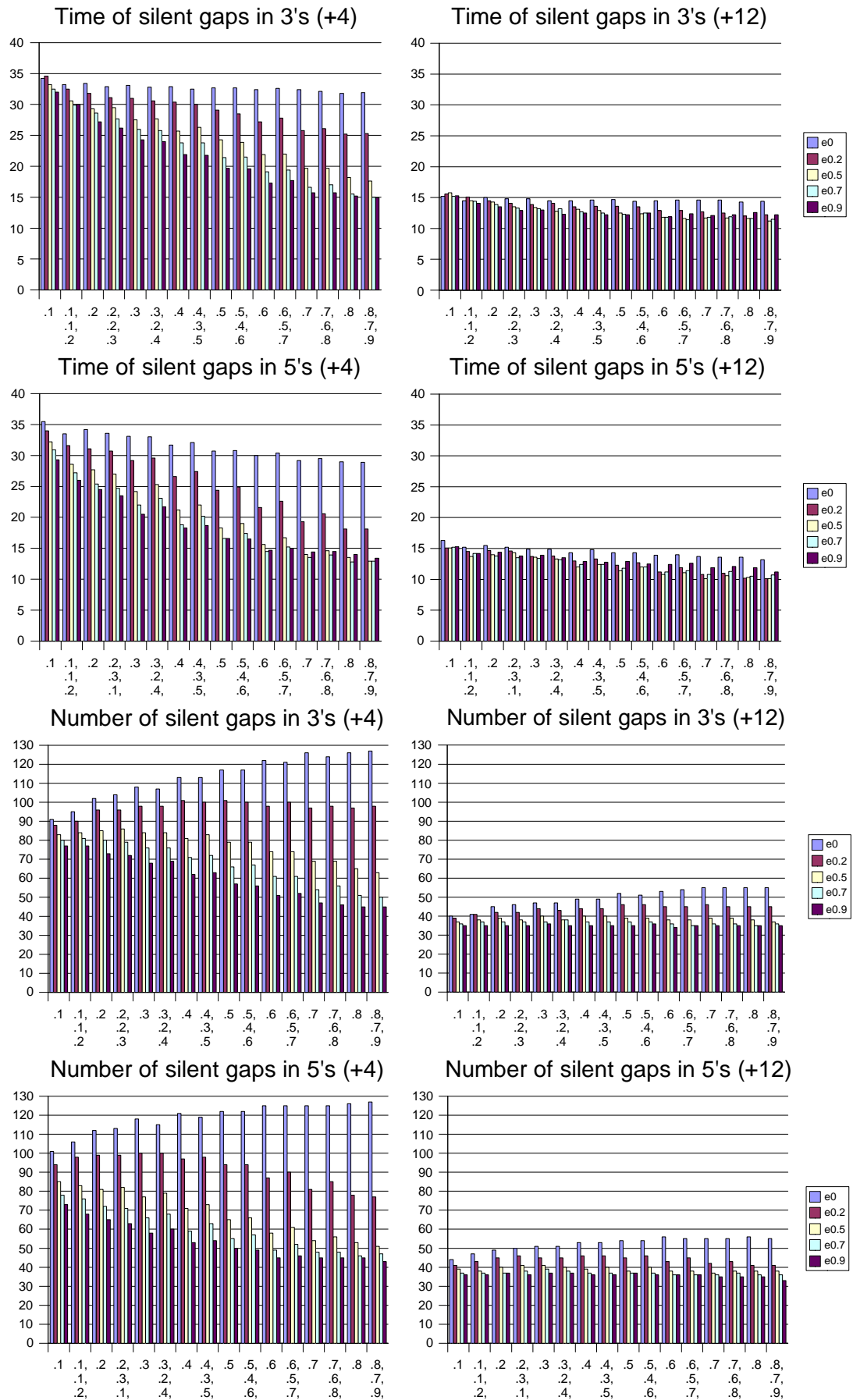
The comparisons for total time and number of silent gaps appear in figure 6.11. The overall increase in utterance lengths reduced these average totals only to a certain minimum, around 12–10s and 35 occurrences, which means the average length was about 0.3s. The mean length calculated by the evaluation program was indeed 0.3s most of the time, with occasional hikes to 0.4s at higher values of EAGERNESS and lower TALKATIVENESS (between 0.1–0.3) for both 3- and 5-agent groups. With 5 agents, however, mean length of silent gaps went down to 0.2s at low values of EAGERNESS (0 and 0.2) and high TALKATIVENESS. The longest silent gaps (and lapses) varied between 3.8–0.6s along the parameter range for 3 agents (it was 4.2–0.8s before), and between 2.6–0.4s for 5's (same as before).

With regards to the comparative results of time and number of silent gaps, the previous soaring totals at low EAGERNESS likelihoods (on the left) are now conspicuously absent with longer utterances. That is because there are much fewer utterances now, therefore much less intervals between them. Only the numbers of silent gaps resemble the previous profile of EAGERNESS variation, in a smaller proportion. Their times show at the highest EAGERNESS an increasing trend that was barely discernible before.

Figure 6.12 shows that the three-time increase in the maximum length of utterances generated by the simulation had as much as three times the (inverse) effect on the number of utterances produced. It is not as much, however: it is nearly three times less utterances with 5 agents, to around 38–50 utterances with peaks of 58. But it is only two-and-a-half times less for 3 agents, around 40 occurrences (whereas in the previous setting the totals were 90 to 110–125). Now, it is only with 5 agents that the greater 'effectiveness' of high EAGERNESS likelihoods at high TALKATIVENESS can be seen modestly, increasing the totals about 20% (at most), by around 10 new utterances. Middle EAGERNESS values, that yielded good results previously, now produce slightly fewer utterances. The reduced total number of utterances, or rather, of opportunities for turn-taking is also reflected in the minimized number of overlaps for 5-agent groups (in figure 6.24, gathered together with overlaps of various evaluations in §6.2.4).

Next we see in figure 6.13 why simultaneous talk now grows steadily almost to previous highs as EAGERNESS increases. The figure shows the comparative totals of middle starts of talk and incomplete utterances (abandoned talk of 1 second or more). Longer utterances now have more TRPs in their middle, which means that the possibilities of eager agents starting to speak forthwith after any TRP, and create 'middle' simultaneous talks, are greater. This in turn increases the frequency with which speakers

Figure 6.12: Number of utterances in different maximum utterance lengths
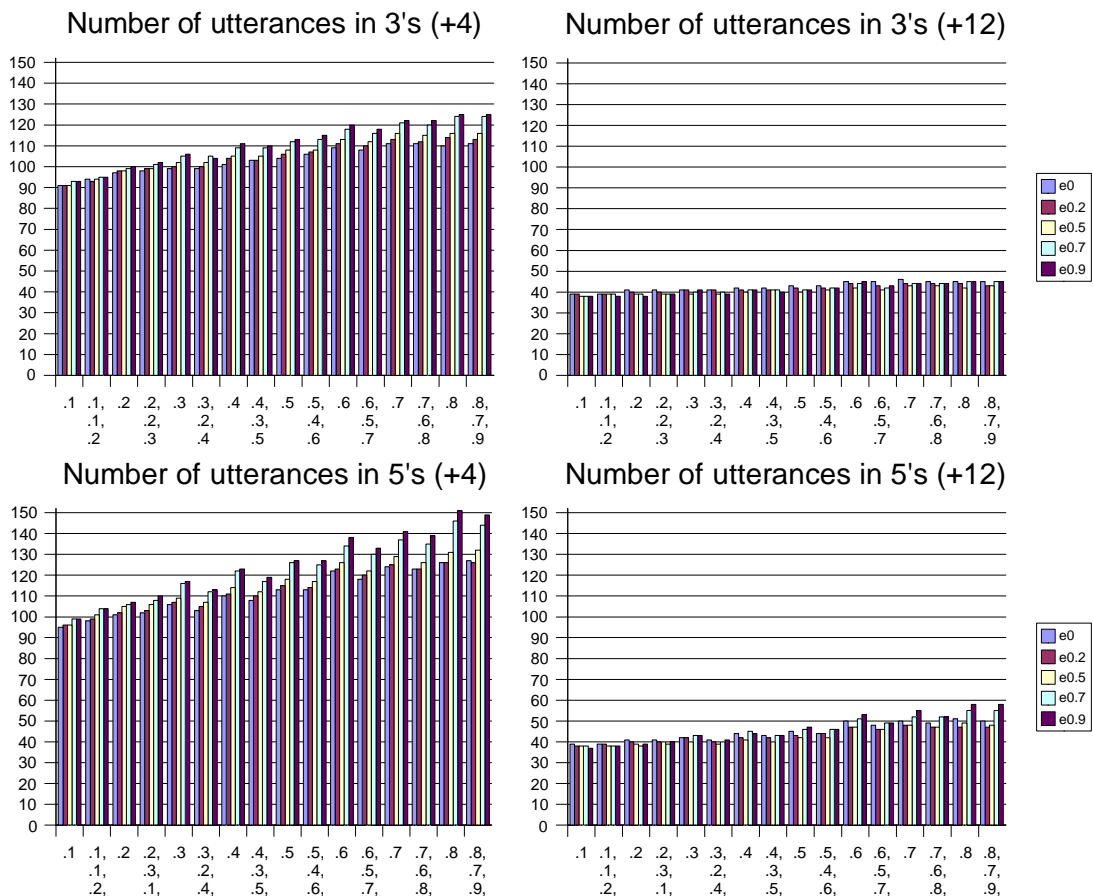
abandon their utterances because of those middle starts.

The charts on the right of this figure show that the three-time increase in the maximum utterance length causes nearly twice as much middle starts of talk and, to a lesser extent, incomplete utterances. There are of course no possible middle starts when EAGERNESS is 0, but they grow steadily with each higher setting, almost twice as much with the longer utterances. The increase ratio is not higher because middle pre-TRPs are generated at substantial intervals along the utterances (around 2s); and then only 50% of the time, following the INTERACTIVITY setting of 0.5 for all agents. A higher likelihood there would mean they are generated *more often* along these fixed intervals. The number of incomplete utterances, on the other hand, remains level with zero EAGERNESS, almost the same as previously, but as this attribute increases the numbers grow relatively more with longer utterances.

Figure 6.13: Middle starts and incomplete utterances in different max. lengths
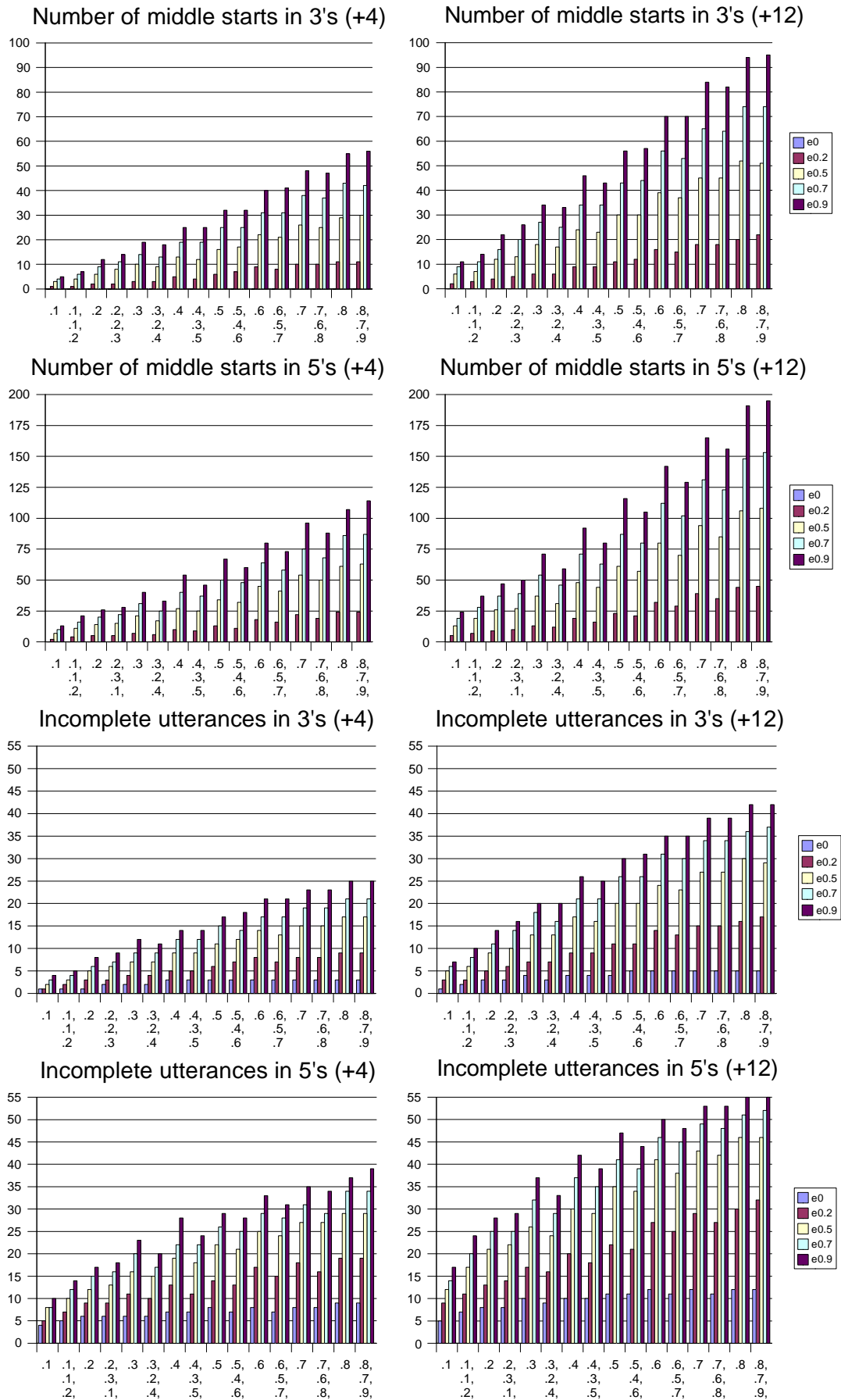
Figure 6.14: Number of individual and 'collective' false-starts in different max. lengths
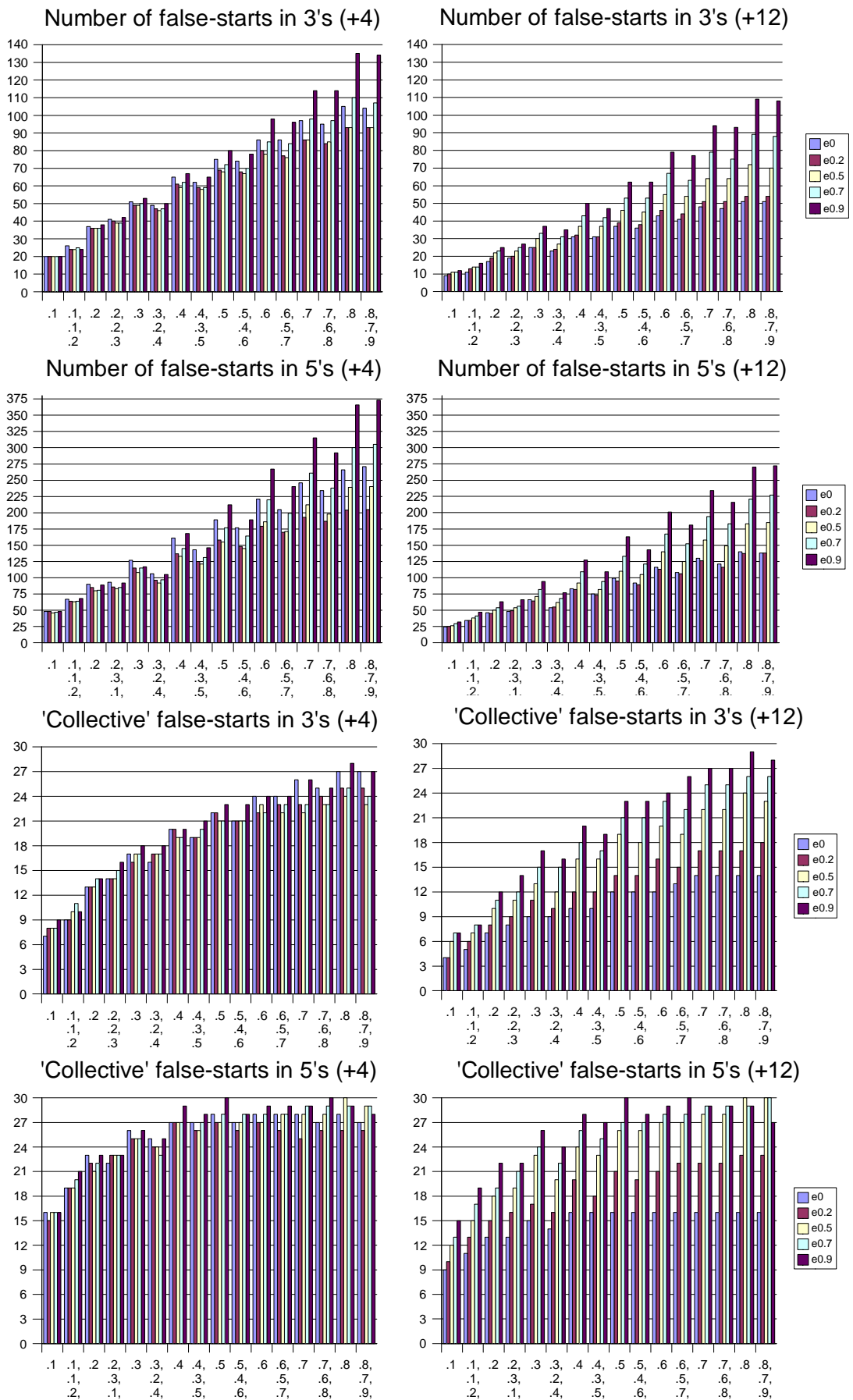
Finally, figure 6.14 presents the totals of individual and collective false-starts for the two group sizes in the two maximum utterance length settings. Both measures show overall reductions with longer utterances, to roughly over half the previous totals, except that reductions are greater for lower values of EAGERNESS. While, previously, variation in this attribute would not cause appreciable differences in the number of false-starts but for slight falls in the low-middle range (for both measures), because of the lesser number of starting clashes that are achieved thus, now with longer utterances there are increases for each higher value of EAGERNESS.

However, number of individual false-starts grow little at low-to-middle values of EA-GERNESS, and only more significantly at higher values along the TALKATIVENESS range. Really, the effect of EAGERNESS with longer utterances is now more on *middle* talk, in the middle of utterances, than in *between* them (which is the locus of false-starts). The reduction at middle EAGERNESS likelihoods from more variation in turn-taking timings that was prevalent with shorter utterances, now only barely begins to appear with 5 agents in 0.2 EAGERNESS.

As for collective false-starts, again they reach equilibrium points as TALKATIVENESS rises, quite visibly in group size 5, less so in 3's. With longer utterances, the various EAGERNESS settings create good increases in their number most of the time, the biggest from 0.2 to 0.5. Whereas with shorter utterances the attribute's influence was minimal: collective false-starts tended to be level along the parameter range.

6.2.3 Confidence variation in 5-agent groups

The next step in the evaluation was to see the effect of another secondary attribute: CONFIDENCE. Leaving EAGERNESS fixed at 0.5 for all agents, evaluations were made along the same range of TALKATIVENESS parameters and in five settings of CONFI-DENCE for all (as with EAGERNESS previously), but just for 5-agent groups now.

The same caveat on the range of 'appropriate' values of TALKATIVENESS can be made for CONFIDENCE. To emulate or at least come closer to real conversant behaviour, likelihoods for this attribute cannot be too high: a value of, say, 0.9 for one agent means that it will almost never abandon its utterance in simultaneous talk, which for *all* agents is certainly pushing the boundaries of any 'reality' too far. They nonetheless should give a profile of the attribute's range of effects in these results.

Figure 6.15: Times of single talk, gaps and multiple talk with varying Confidence
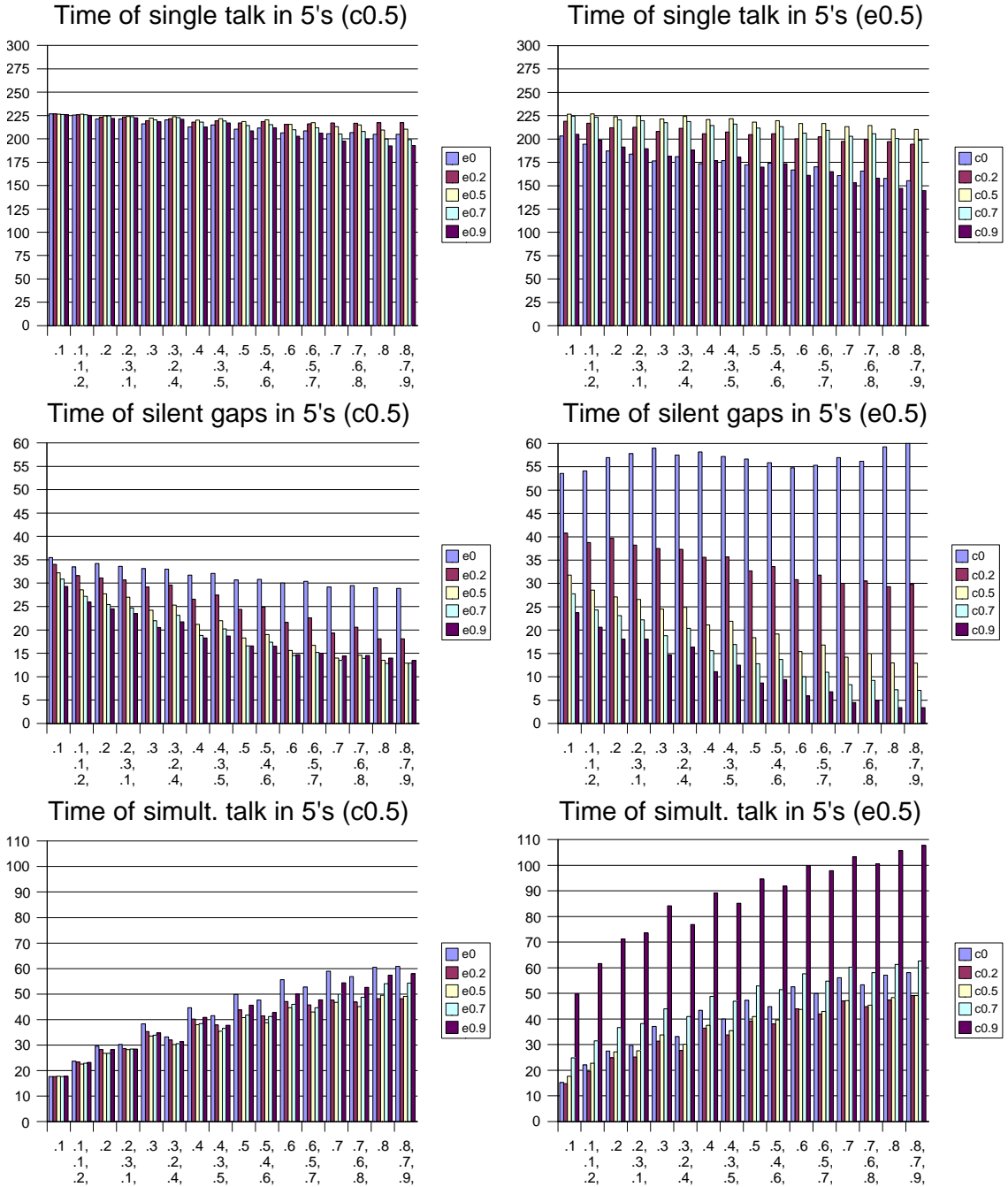
First, figure 6.15 compares the total amounts of single talk, silent gaps and simultaneous talk varying in two secondary attributes. The charts on the left give the (previous) results of EAGERNESS variation with group CONFIDENCE fixed at 0.5. Those on the right show the new totals of CONFIDENCE variation with EAGERNESS fixed at 0.5 (as the legend and chart titles indicate). Clearly then, the middle bars at every TALKATIVENESS setting should be the same on the charts of both sides (for both EAGERNESS *and* CONFIDENCE at 0.5, originally yellow in colour but blank in print).

The totals of single talk varying in CONFIDENCE (right) actually show the same profiles of the EAGERNESS variation (left), except for more pronounced drops at the extremities, 0 and 0.9. The other measures in this figure (6.15) together show where the lost amounts of single talk time went (mind the different chart scales of each measure). When CONFIDENCE is 0 for all agents, they will always stop immediately on simultaneous talk, generating lots of collective false-starts (second right chart of figure 6.18), subsequent simultaneous starts (second right chart of figure 6.17) and, thus, silent gaps inbetween attempts to talk (second right chart of figure 6.16). Hence the time that would be single talk goes to silent gaps (and a little to simultaneous talks too).

On the other side with CONFIDENCE at 0.9, agents would almost never stop in the face of simultaneous talk, clearly producing a lot of it as figure 6.15 shows. This greater amount comes from some simultaneous starts (figure 6.17), but mostly from long simultaneous episodes (bottom right chart of figure 6.18).

With regards to simultaneous talks, we can see that CONFIDENCE likelihoods resulting in lesser amounts are between 0.2 and 0.5—maybe a wider range if agents would have different values. That is again because middle likelihoods for all agents mean *they vary the most* in their decisions, increasing the chances of behaving differently each time so that just one end up continuing, and reducing the time of simultaneous talks without too much false-starts.

The charts of figure 6.16 show the average number of utterances, silent gaps and simultaneous talks for 5 agents in the same variation of EAGERNESS (left) and CONFIDENCE (right). In the number of utterances, the latter attribute has greater influence, with TALKATIVENESS increases becoming more irregular. The explanation is that agents with low CONFIDENCE too easily stop in the face of simultaneous talk, leaving many (individual) false-starts, whereas more confident ones get to finish *many* more utterances.

Figure 6.16: Number of utterances, gaps and multiple talks with varying Confidence
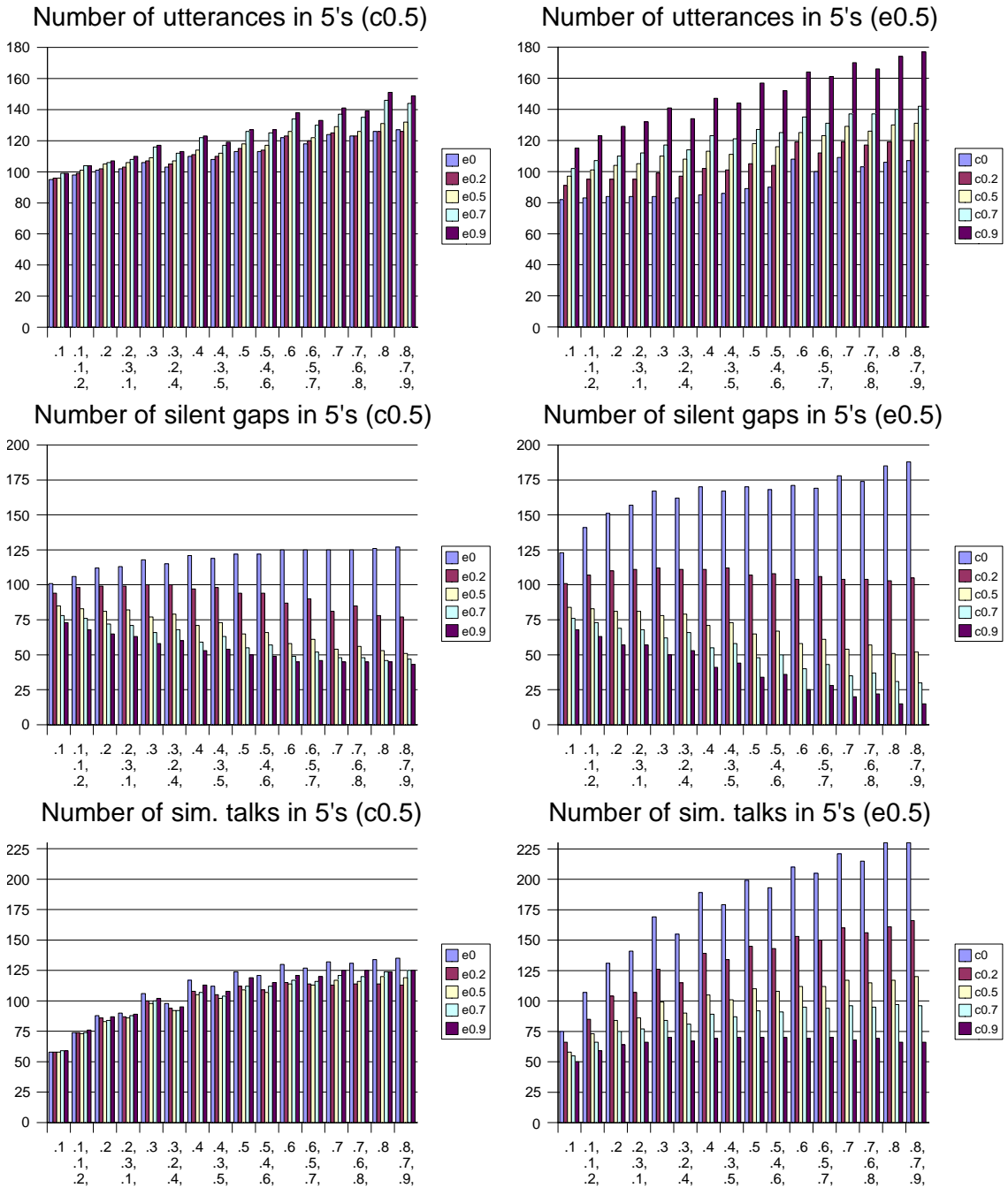
The numbers of silent gaps in figure 6.16 tell more or less the same story as for the times of silent gaps in figure 6.15. But the numbers of simultaneous talk tell a different story altogether. While CONFIDENCE for all agents is low, simultaneous talks will tend to be short and their numbers grow with the TALKATIVENESS of agents. As CONFIDENCE goes around middle likelihoods, the total *times* of simultaneous talk (figure 6.15) decrease a little—because of the 'different behaviour' effect mentioned previously—, then explode as likelihoods *of all agents* go beyond middle likelihoods.

However, the *numbers* of simultaneous talks only decrease as everybody's CONFI-DENCE increases, because agents are less and less prone to stop shortly on simultaneous talk, giving way to other speakers. While numbers decrease, each simultaneous conflict becomes longer and longer as reflected in the soaring numbers of *long* simultaneous talks of figure 6.18 (bottom right chart). Number of (general) simultaneous talks plummets at first, at initial values of CONFIDENCE, then less and less until reaching a minimum of about 66 occurrences.

Accordingly, the mean length of simultaneous talks grow from 0.2s at lower CON-FIDENCE to 0.3–0.4s, 0.5–0.6s and 1–1.6s along the TALKATIVENESS range as the group CONFIDENCE is set at 0.5, 0.7 and 0.9, respectively. Notice the greater increase on the 0.7–0.9 change, and bear in mind that these are *mean* lengths. The longest episodes of simultaneous talks varied along the TALKATIVENESS scale between 1.0–1.6s, 1.2–1.8s, 1.6–2.6s, 3.4–4.2s and 5.0–11.2s in the five settings of CON-FIDENCE, respectively—11.2s was surely a compound of successive simultaneities (stops, restarts and middle starts, without intervening silences), since the maximum utterance length was set back at 4s.

Next, figure 6.17 compares EAGERNESS and CONFIDENCE variation on the average totals of single starts, multiple and middle starts of utterances for 5 agents. As previously, charts on the left show EAGERNESS variation, and those on the right the new CONFIDENCE one. There is not much change or improvement in the case of single starts (one of the weaknesses of the basic model): they are only slightly higher in zero CONFIDENCE than in EAGERNESS, and decrease more or less similarly.

Simultaneous starts, by their turn, tell a similar story to that of number of silent gaps and simultaneous talks in general (figure 6.16). With low CONFIDENCE, agents will frequently stop forthwith in simultaneous talk (collective false-starts in figure 6.18), creating a series of multiple starts, silences and multiple *re*starts, increasingly as

Figure 6.17: Number of single, multiple and middle starts with varying Confidence
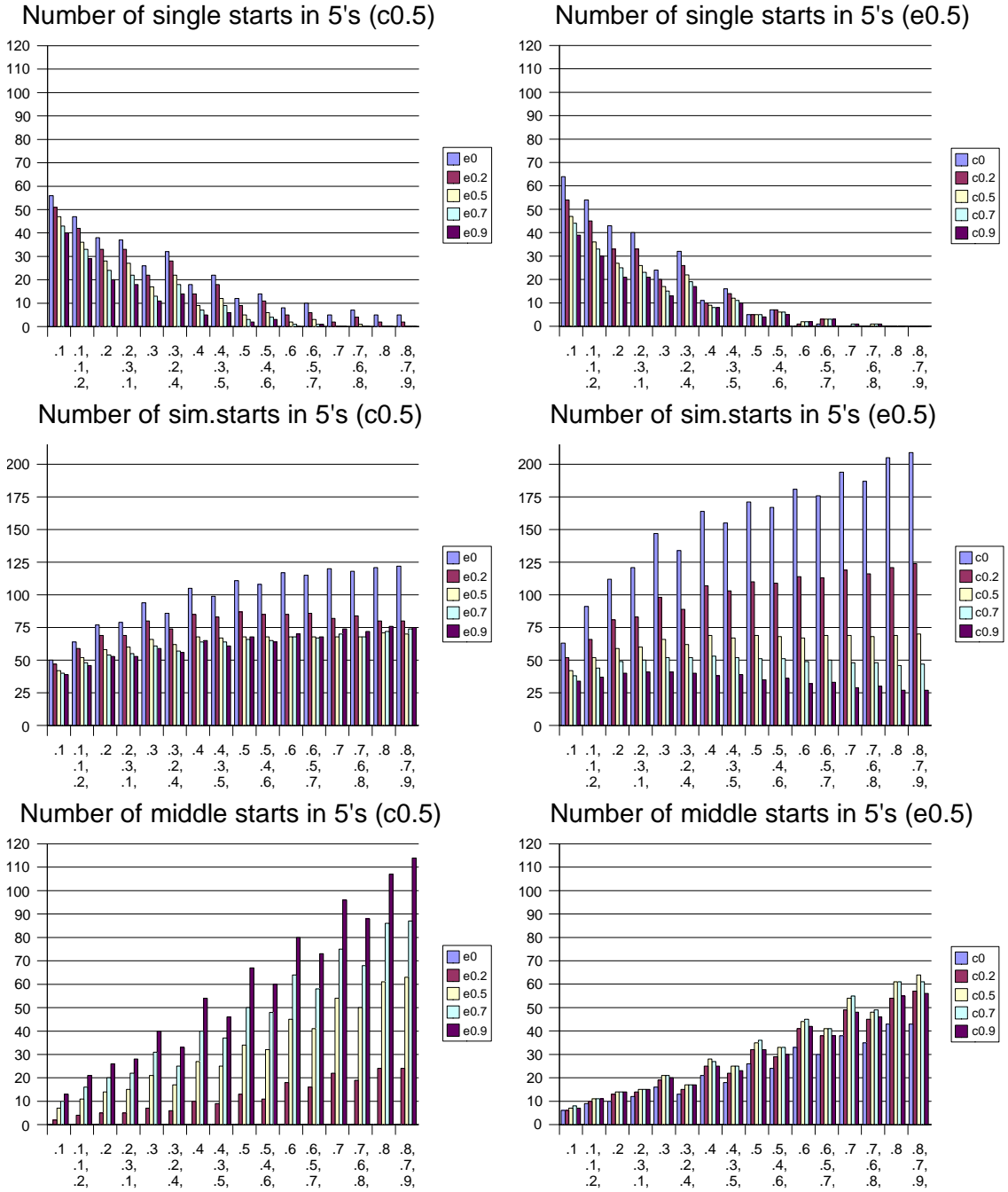
TALKATIVENESS gets higher. Then hardly anyone gets the floor while more than one agent is deciding to speak each time. The result: frequent failures in anyone actually getting a clear shot at talking. Hence the much smaller number of complete utterances (figure 6.16). Then as CONFIDENCE increases the number of simultaneous starts (like that of simultaneous talks in figure 6.16) grows less and less along the TALKATIVE-NESS range until finally beginning to diminish in higher settings, previsibly. And the mean number of multiple starters goes from 2.1 to 3.6 along the parameter range.

As for middle starts, it is EAGERNESS their main influence, clearly, as the charts show. The peaks at middle values of CONFIDENCE (0.5, 0.7) derive from the 'different be-haviour' effect of middle likelihoods, and from a breakdown of the normal single talk that follows turn-taking at the extremities of the attribute, as already described. At zero CONFIDENCE, more false-starts than properly single talk occurs in high TALKATIVE-NESS. At 0.9, most or all agents end up talking at the same time for longer periods, so middle starts of utterances should also decrease a little.

Lastly, figure 6.18 presents the average numbers of individual and collective false-starts, incomplete utterances (of one second or more), and long simultaneous talks (idem). The left charts again show the variation in EAGERNESS, and the right in CON-FIDENCE.

The totals of individual and collective false-starts now change differently through the CONFIDENCE settings than through EAGERNESS. Except for individual false-starts at low TALKATIVENESS, both measures tend to fall in proportional rates as CONFIDENCE increases, because obviously more confident agents persist more in simultaneities, pro-ducing less false-starts.[5]

As for collective false-starts, note that while the number of individual false-starts con-tinually grow along the TALKATIVENESS range in both variations, those of collective false-starts tend to reach level upper limits in both cases too, at every setting of the sec-ond attribute. This different 'profile' derives exactly from their *collective*, as opposed to *individual*, nature. And why their numbers swell at low CONFIDENCE was previ-ously explained: non-confident agents going into cycles of stopping to talk, restarting, then stopping again and so on, more so the higher their TALKATIVENESS. It turns out that the profile of collective false-starts is associated rather directly to that of simulta-neous starts (thus simultaneous talks) and that of silent gaps.

---

[5]The charts are in different scales, so the decreases in collective false-starts are actually very similar to those of individual false-starts (though totals are smaller).

Figure 6.18: Number of false-starts, incompletes & long sim.talks with var. Confidence
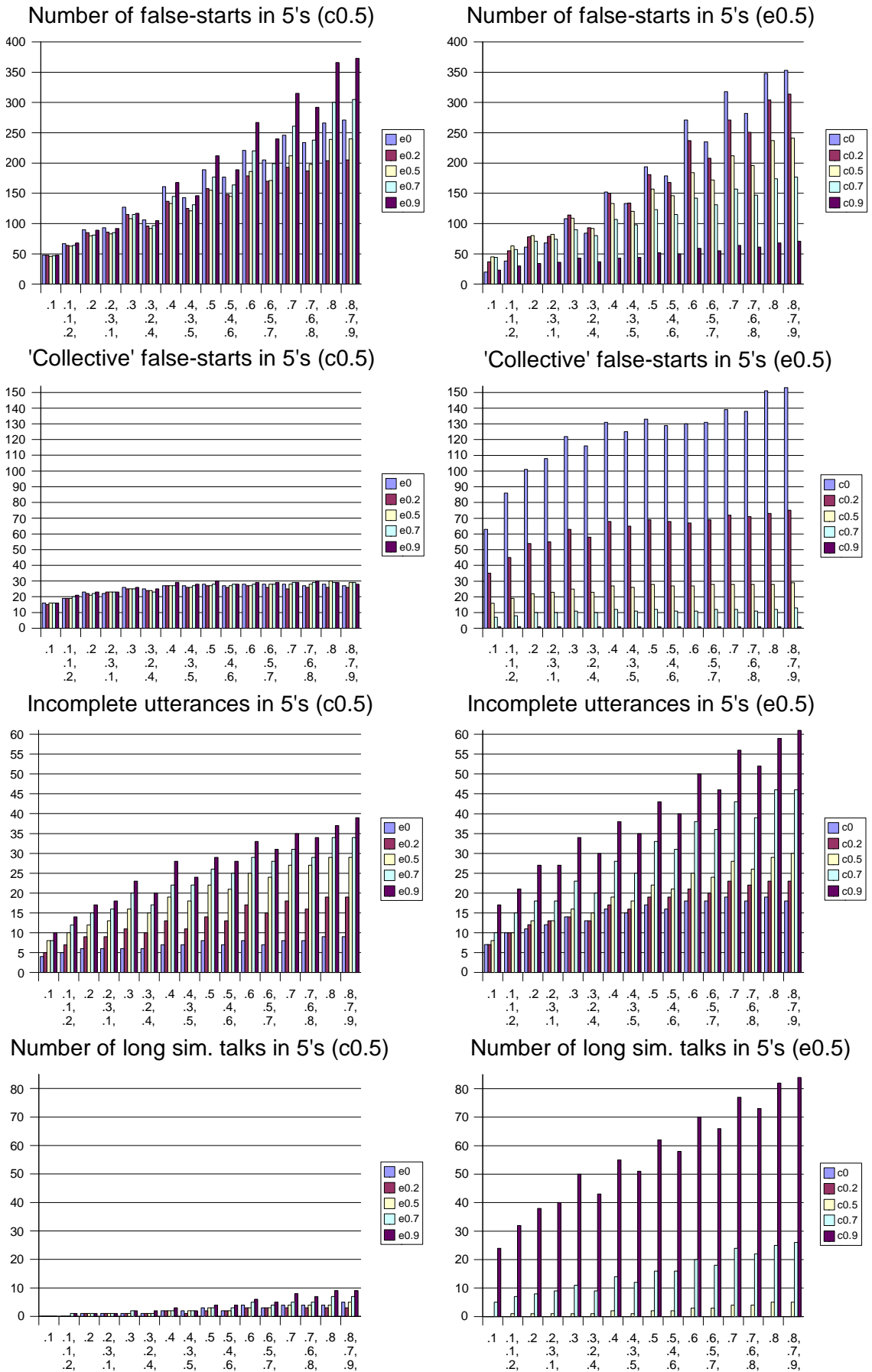
On the other side, the number of incomplete utterances grows slowly in low-middle CONFIDENCE, only rising exponentially at higher settings, as agents become increasingly persistent in talking—though not enough for all of them to continue indefinitely (that would be the case with a 1.0 likelihood)—, causing a lot more abandoned utterances. One or more agents are always going to stop after some (non-short) length of simultaneous talk, turning what would otherwise be false-starts into incomplete utterances. That is why the latter grows inversely (reciprocally) to the former.

Finally, it is no surprise that the number of long simultaneous talks explodes at the highest CONFIDENCE. The interesting thing here is that, with the simulation's system of checking and re-checking the agents' decisions of whether to continue at *every* cycle of simultaneous talk, long episodes (although 1s is not exactly *long*) only end up occurring at middle to high likelihoods in fact. And still at 0.5 CONFIDENCE the number is very small (between 5–27).

In reality, however, even people that would normally concede the floor promptly most of the time, may persist once in a while in occasions wherein they really want to get their message across. Of course this is one of the model's shortcomings, that stems from the lack of contents of talk that could make agents behave and decide differently each time (differently than their 'normal', normalized behaviour), but therein also lies a problem in how simultaneous talk is dealt with.

The decision to stop or persist talking in simultaneity is re-checked every (minimal) cycle of the simulation, in a clockwork fashion. However, while talk is underway it is much the case that, as speakers get less attentive to others and more in formulating their ongoing talk (Levelt 1989), they would take longer to recognize the others' and decide whether to stop. Unlike the 'clockwork' behaviour here, but more according to some other variable (in a psychological sense) independent of the simulation's minimal cycles: some cognitive *response-time*, dependent on the interest and attention one is giving to his or her present talk. 'ATTENTION' was one other attribute that would represent this, and was considered early in the model development but later discarded (still appearing in the code of appendix 1).

6.2.4 Interactivity variation in 5-agent groups

Another evaluation was conducted, this time varying the INTERACTIVITY attribute while keeping all the others at 0.5. The next four figures compare their results in the charts to the right with the previous EAGERNESS variation to the left. The INTER- ACTIVITY attribute just changes the make-up of the generic utterances by increasing or decreasing the frequency (thus number) of intermediary TRPs within them and the 'finishing talks' that allow overlapped transitions. It is therefore of turn-taking interest to observe what happens when utterances allow more or less opportunities for talk (or interruption).

In figure 6.19, the falling time of single talk as INTERACTIVITY (and TALKATIVE- NESS) gets higher comes from the increasing time taken not only by simultaneous talks but also by more frequent overlaps in speaker transitions (shown in appendix 3). INTERACTIVITY contributes directly to both. In the case of simultaneous talks, the increase in intermediary TRPs creates more middle starts of talk but then (as CONFI- DENCE is now back at 0.5) more incomplete utterances (figure 6.22) and simultaneous starts too (figure 6.21).

But the bulk of the time *stolen* from single talk comes in fact from overlapping transi- tions, which although not 'single talk' proper are not exactly *conflictive* simultaneous talk either. Each individual overlap has a (variable) short length of up to 1s correspond- ing to the tail of 'finishing talk' at the end of utterances; so, as the number of overlaps can soar (figure 6.24) to a hundred or more, the total time thus 'taken' from single talk can be significative, 20s or more.

One other measure directly determined by INTERACTIVITY appears for the first time in figure 6.19 (more are in appendix 3): the amount of backchannel vocalizations, that is, feedback in the background of talk. The bottom charts in this figure present the total time they occupy in the two attribute variations. EAGERNESS affects their occurrence only indirectly: in allowing agents to speak immediately when they decide so without waiting for silence, reduces their chance of ever giving any feedback in high TALKATIVENESS, since it only occurs when an agent decides *not* to talk. On the other hand, INTERACTIVITY affects the amount of backchannels directly: more TRPs in the middle of utterances cause not only more middle starts of talk, but also more overlapped feedback in response, when agents decide *not* to talk there.

The profiles of time and number of silent gaps in this variation seems to be a reflection

Figure 6.19: Times of single talk, gaps and multiple talk with varying Interactivity
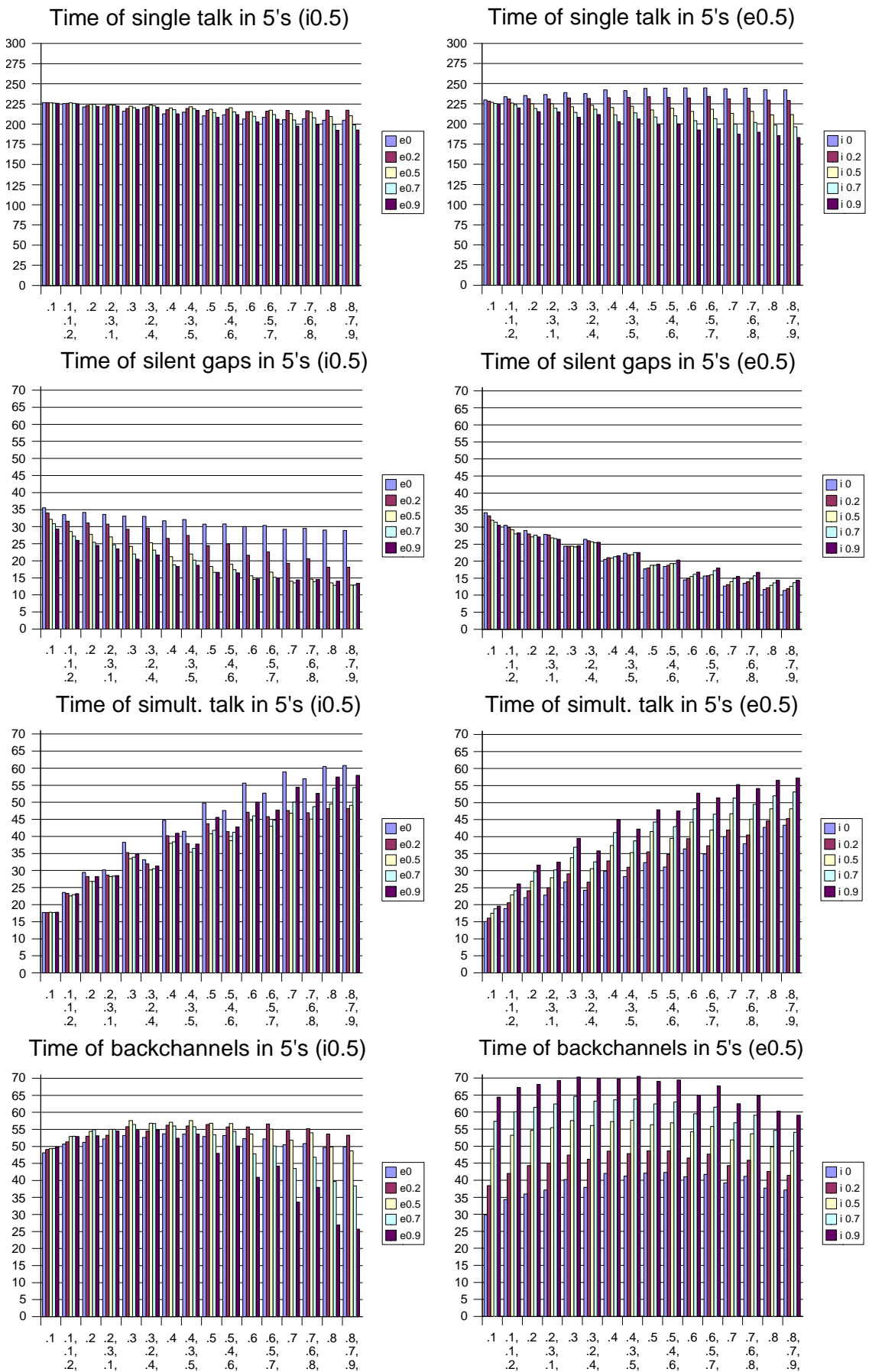
Figure 6.20: Number of utterances, gaps and multiple talks with varying Interactivity
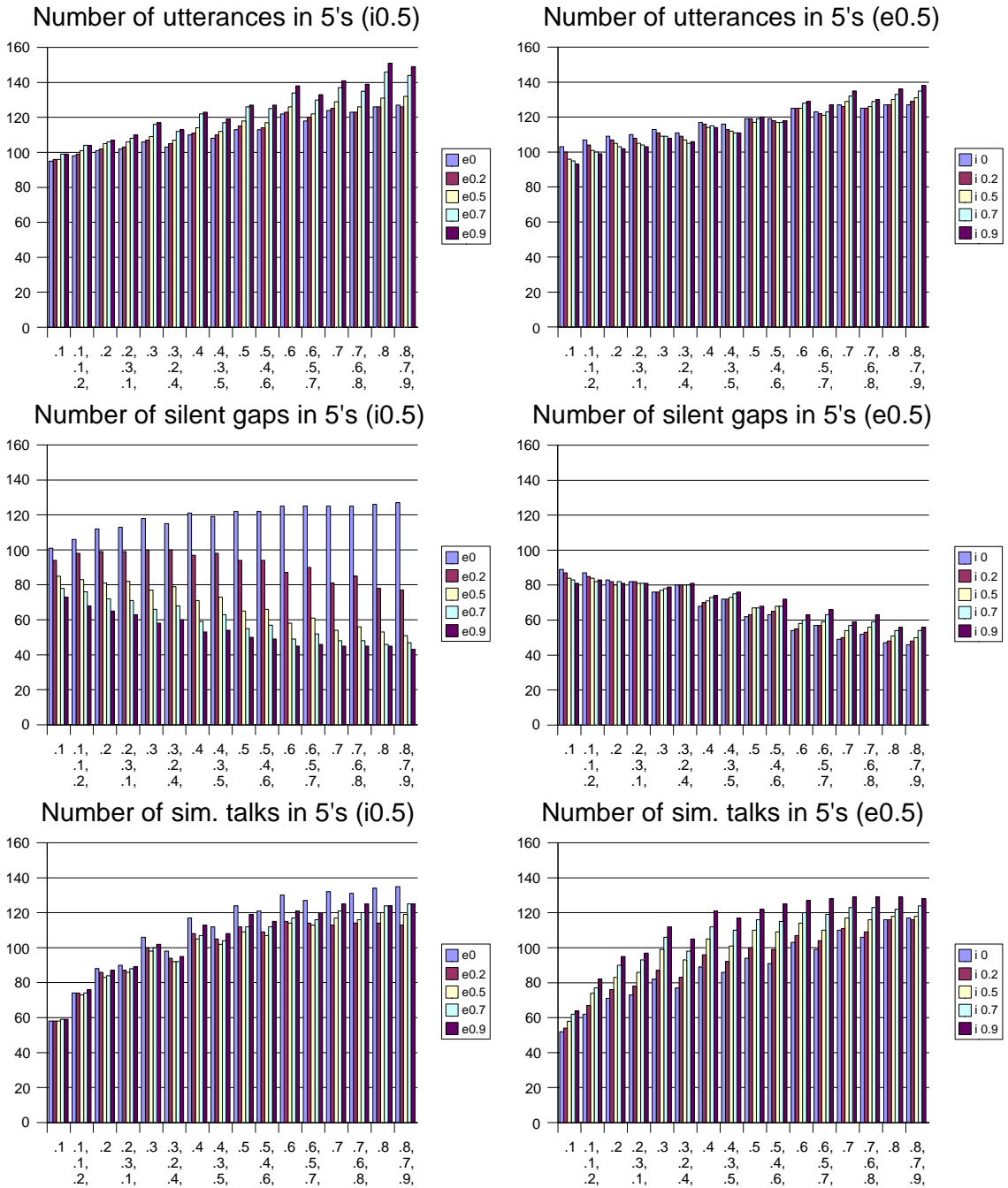
of the resulting number of utterances in figure 6.20, even though they do not appear to match up exactly. As TALKATIVENESS increases, the rising number of intermediary TRPs inverts the initial descending trend in number of utterances, that results from more incomplete (figure 6.22) than complete utterances with more middle starts. In higher TALKATIVENESS, however, *more* agents attempt to start at middle TRPs, creating more complete *and* incomplete utterances, hence more intervening silent gaps.

In figure 6.21, the charts that should get our attention are the bottom ones for middle starts, which appear to be the same. They are not, actually, but very similar. The reason is that varying EAGERNESS or INTERACTIVITY are different, yet related, ways of creating more or less middle starts of utterances. Either through less middle TRPs but with more eager-to-talk agents, or through more middle TRPs but with agents less likely to talk forthwith without waiting for silence.

Finally, figure 6.22 presents the totals of the various 'failure' measures and long simultaneous talks, in the previous EAGERNESS (left) and the INTERACTIVITY variation (right). While the effect of EAGERNESS on individual and collective false-starts hinged on different or similar turn-taking behaviours (respectively, in middle likelihoods or not), now the effect of INTERACTIVITY comes from the varying number of potential middle starts of talk, which, as TALKATIVENESS rises, result in more individual false-starts, incomplete utterances or both (as the group's CONFIDENCE is 0.5).

In addition, the number of collective false-starts also grow initially with rising INTERACTIVITY (bearing in mind their smaller scale) up to the familiar 'collective' limit between 27–30 at high TALKATIVENESS. What happens now is that there are more occurrences of *all* middle starters stopping together with the current speaker. In sum, the influence of INTERACTIVITY on false-starts with the current parameters turns out to be indirect: more middle-starts induce more simultaneous starts, hence more opportunities for both sorts of false-start to occur.

To complete this section, two figures gather together a couple of other measures in the several evaluations seen so far, for comparison. Figure 6.23 presents the utterance distributions of individual agents with different TALKATIVENESS likelihoods in groups of 5, from the previous evaluations: with longer maximum utterance lengths (second charts from top), varying in CONFIDENCE (third charts from top), in INTERACTIVITY (bottom charts), as well as in the initial variation of EAGERNESS (top charts) in three settings for all agents: 0, 0.5 and 0.9. As before, charts to the left give the distribution

Figure 6.21: Number of single, multiple and middle starts with varying Interactivity

Number of single starts in 5's (i0.5)

Number of single starts in 5's (e0.5)

Number of sim.starts in 5's (i0.5)

Number of sim.starts in 5's (e0.5)

Number of middle starts in 5's (i0.5)

Number of middle starts in 5's (e0.5)
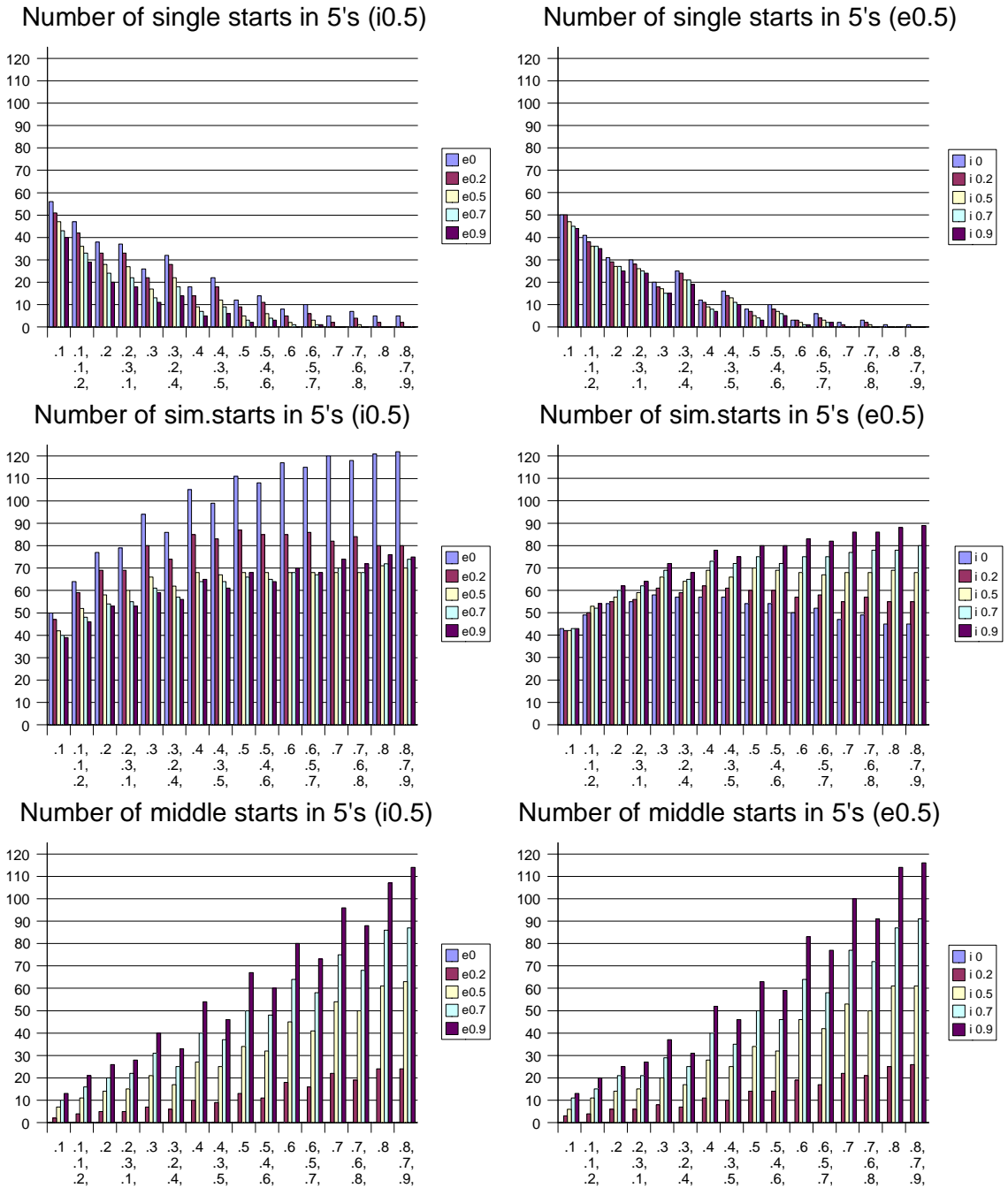
Figure 6.22: Number of false-starts, incompletes & long sim.talks with var. Interactivity

Number of false-starts in 5's (i0.5)

Number of false-starts in 5's (e0.5)

'Collective' false-starts in 5's (i0.5)

'Collective' false-starts in 5's (e0.5)

Incomplete utterances in 5's (i0.5)

Incomplete utterances in 5's (e0.5)

Number of long sim. talks in 5's (i0.5)

Number of long sim. talks in 5's (e0.5)
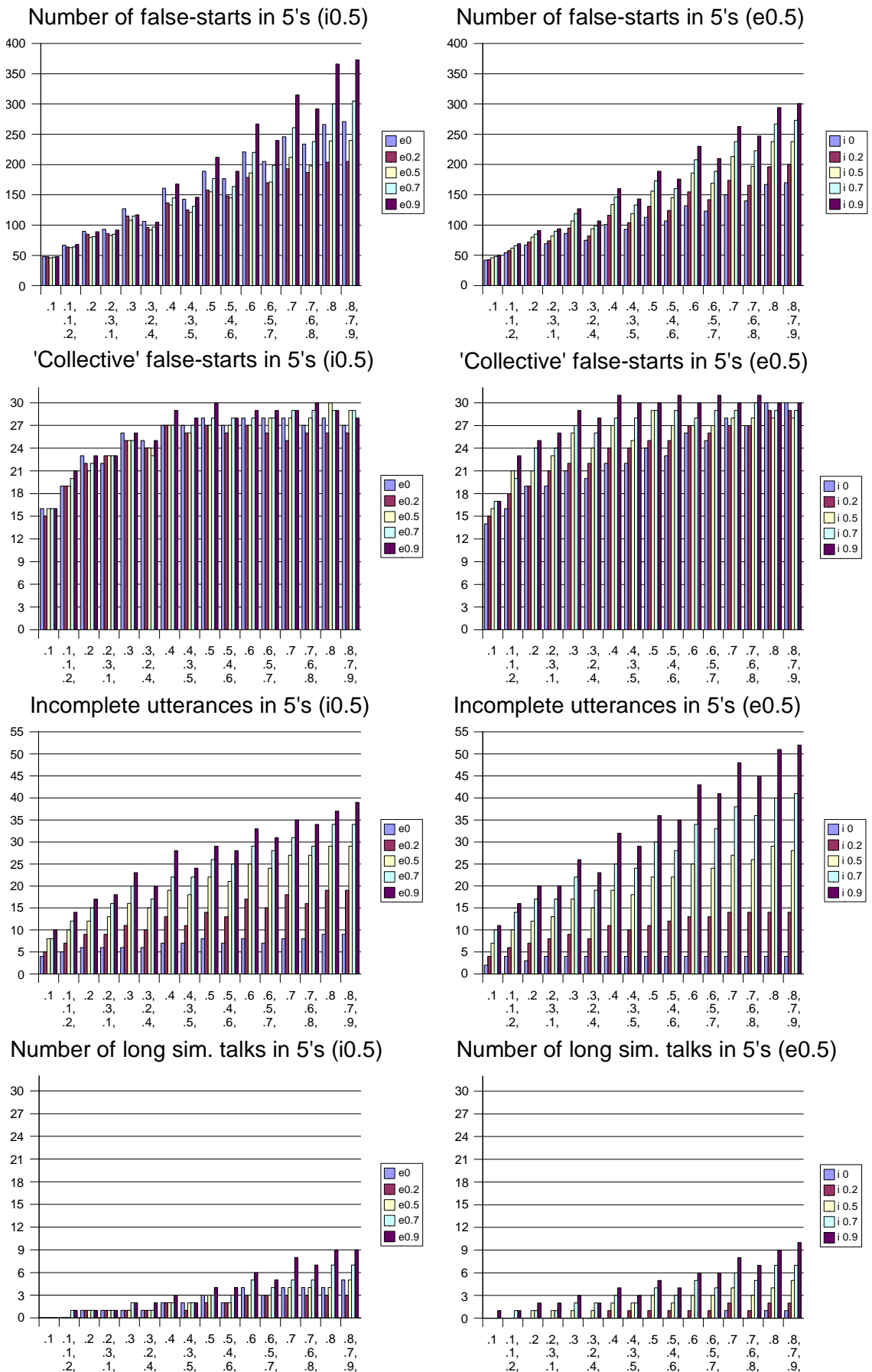
Figure 6.23: Utterance distribution amongst the agents in several variations
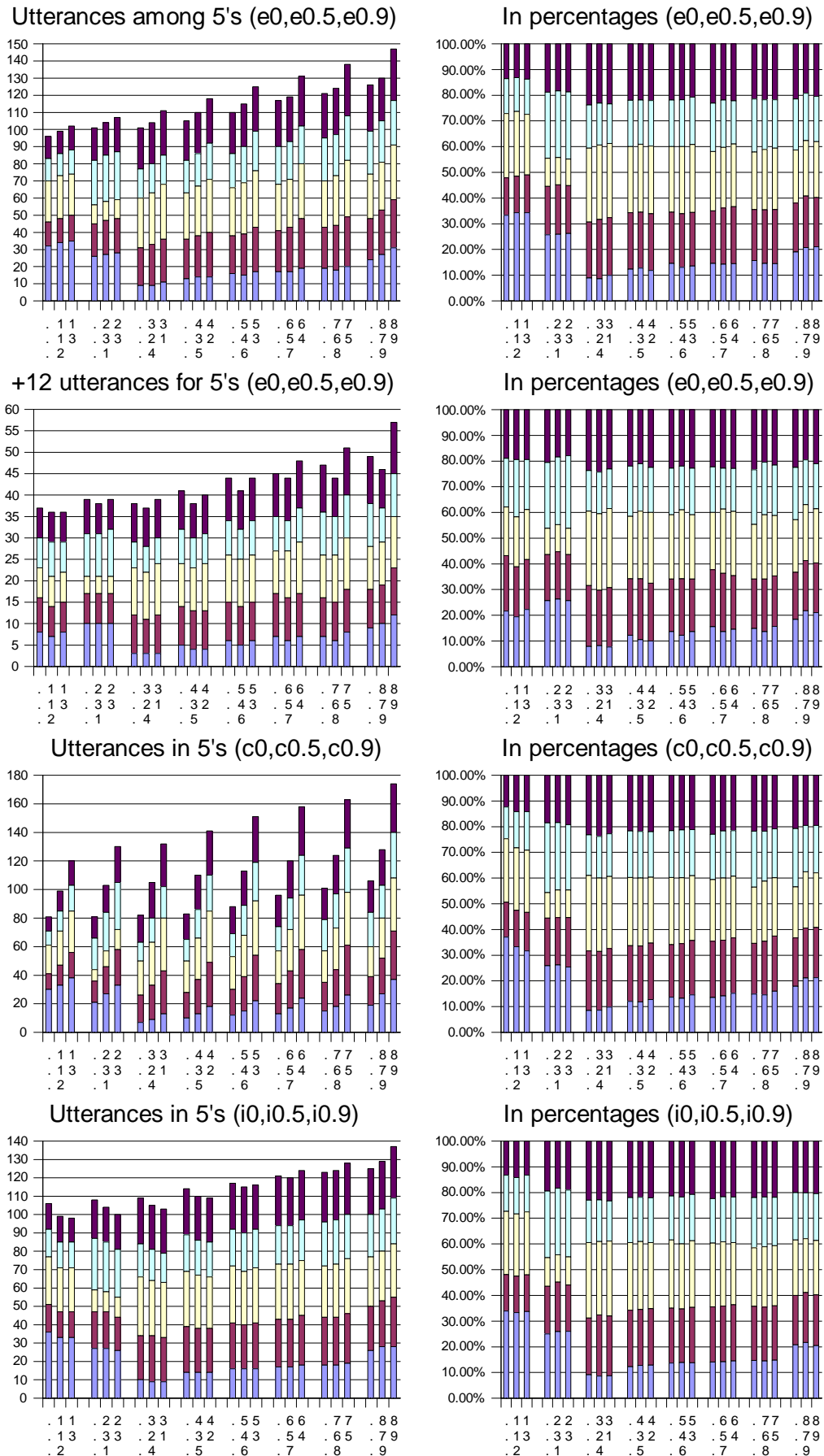
Figure 6.24: Number of overlaps (and latches) in several variations
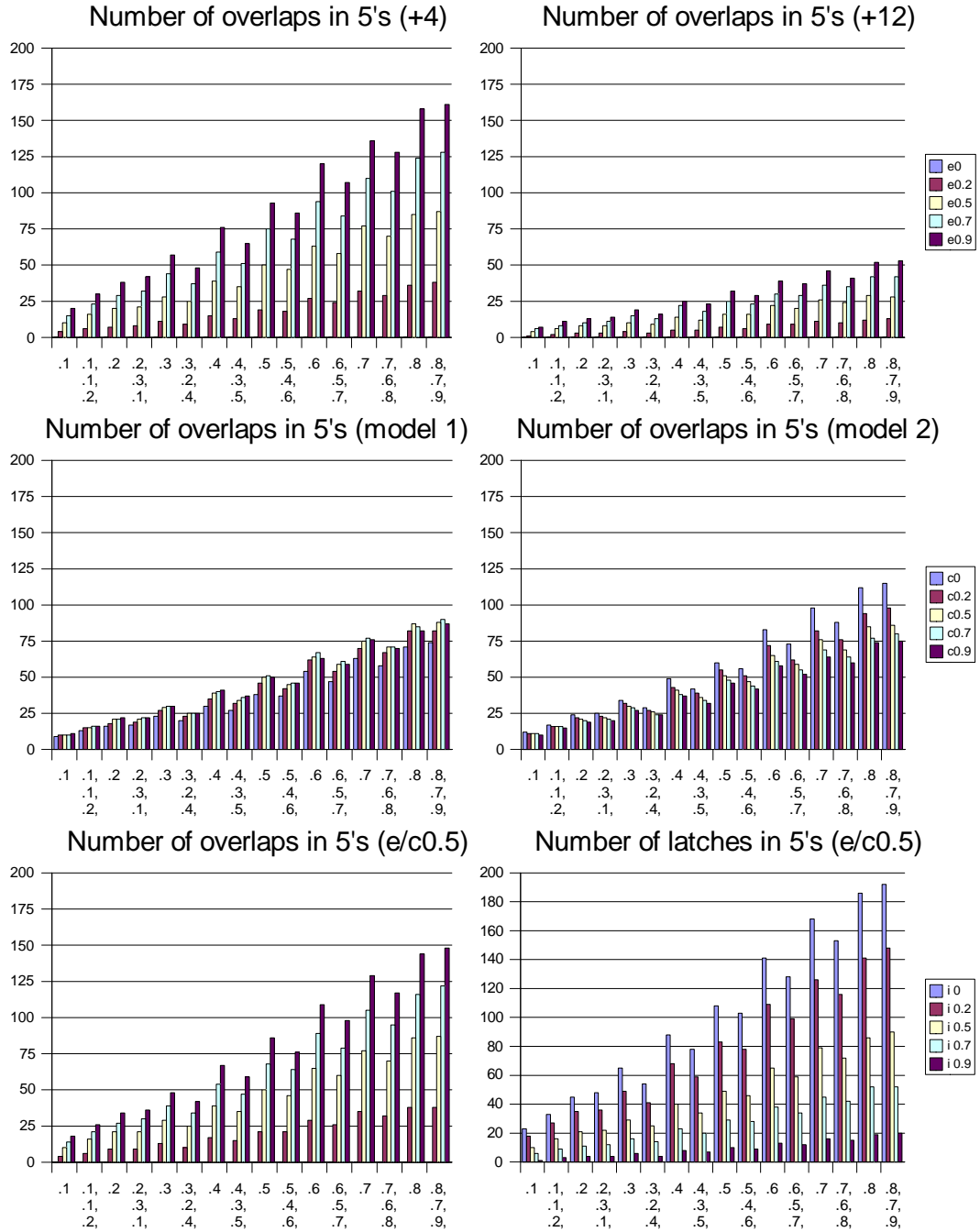
in absolute numbers and those to the right the percentages of the total.

While rising EAGERNESS values for all agents increased utterance numbers both individually for all and the proportions of only the most talkative ones (top charts), the CONFIDENCE variation did not. Increasing this attribute for all agents (third charts) seems to give slightly more percentage shares of the total of utterances to *less* talkative agents at the expense of the more talkative ones. Though of course in absolute numbers, increases were much higher than in any other variation (note the different scales). As for INTERACTIVITY, there does not seem to be much significant change in utterance shares, logically as this attribute is not a turn-taking one; only in absolute numbers is that utterances increased a little.

Lastly, figure 6.24 gathers together the overlap totals of the previous EAGERNESS variation in the two maximum utterance length parameters (top charts), CONFIDENCE variation in the basic and the second model that is evaluated in the next section (middle charts), and the present INTERACTIVITY variation also showing the number of latches that is conversely proportional to that of overlaps (bottom charts).

The last two measures are the interesting ones here to show another effect of INTERACTIVITY in the resulting turn-taking of this simulation. The rising parameter sequence given to this attribute is not symmetric, otherwise overlaps and latches would show perfect inversely proportional numbers. But the parameters are not 0, 0.2, 0.5, 0.8, 1.0, but slightly lower on the last two values (0.7, 0.9): that is why the number of overlaps is smaller therewith than the number of latches with low likelihoods along the TALKATIVENESS range. Nonetheless, overlaps and latches should pair up more or less consistently at 0.5.

6.3   Model 2: procedures of simultaneous talk

The second model (§4.4.4) was evaluated for 5-agent groups in five CONFIDENCE values for all agents, just as in subsection 6.2.3 above, to whose results it will be compared. This attribute is the one relevant here because model 2 only differs from the first in more elaborate procedures for the resolution of simultaneous talk, wherein CONFIDENCE is used. The new procedures make agents less fickle in their decisions to continue speaking, specially when they are near the end of their utterances or these are short, with a second or less to end. This reduces the occurrence of false-starts and

increases the frequency of long simultaneous talks (figure 6.29).

So the charts of the following figures vary in CONFIDENCE as the lateral legends indicate, as well as the in same TALKATIVENESS distribution of the previous section. EAGERNESS and other attributes remain fixed at 0.5 and the maximum utterance length at 4s. The charts on the right give the new totals in the second model, while those on the left the previous ones from the first model for comparison (§6.2.3).

Single talk times in figure 6.25 show that model 2 results in comparatively more of it at low CONFIDENCE values, and less otherwise. Middle likelihoods still create more single talk, but lower values now generate more: like in 0.2, which equalizes and surpasses 0.5 in high TALKATIVENESS. As the other charts show, these changes come from more simultaneous talk and less silences being generated in all settings, with the latter diminishing much more with low CONFIDENCE as it is taken over by more single talk (which seems to indicate a more effective use of time). Totals of single talk go from about 215s at the low side to below 185s in low-to-middle (0–0.5) CONFIDENCE values on the high side, and just over 150s with 0.7. This represents between 72–61% (50% with 0.7) of the total time of the discussions.

In a glance then, the new procedures give this attribute a more *balanced* effect, away from its 'explosive', exponential influence in the basic model. They seem to make the interaction more productive in terms of agents getting their utterances across more often, with less *failures* (figure 6.29) and consequently less intervening silences (figures 6.25 and 6.26) being replaced by more single talk, although with more simultaneous talk too.

The new procedures improve the resolution of simultaneous talk in the initial model by addressing the 'clockwork' way with which it works (as talked about at the end of subsection 6.2.3): realizing it in multiple steps (cycles) of CONFIDENCE decisions, instead of just one and the same at each cycle. This provides agents with a bolder overall behaviour according to the value of this attribute.

The new procedures clearly increase the amount of simultaneous talk, by around 15–35s along CONFIDENCE and TALKATIVENESS parameters: a little in low likelihoods and more in higher. Simultaneous talk now represents about 10–23% of the discussions in up to middle CONFIDENCE likelihoods (and 13–33% with 0.7), whereas they were 5–20% before. More simultaneous talk (with correspondingly less single talk) is not necessarily a bad thing though: it means that people are persisting more with their

Figure 6.25: Times of single talk, silent gaps and multiple talk in models 1 & 2
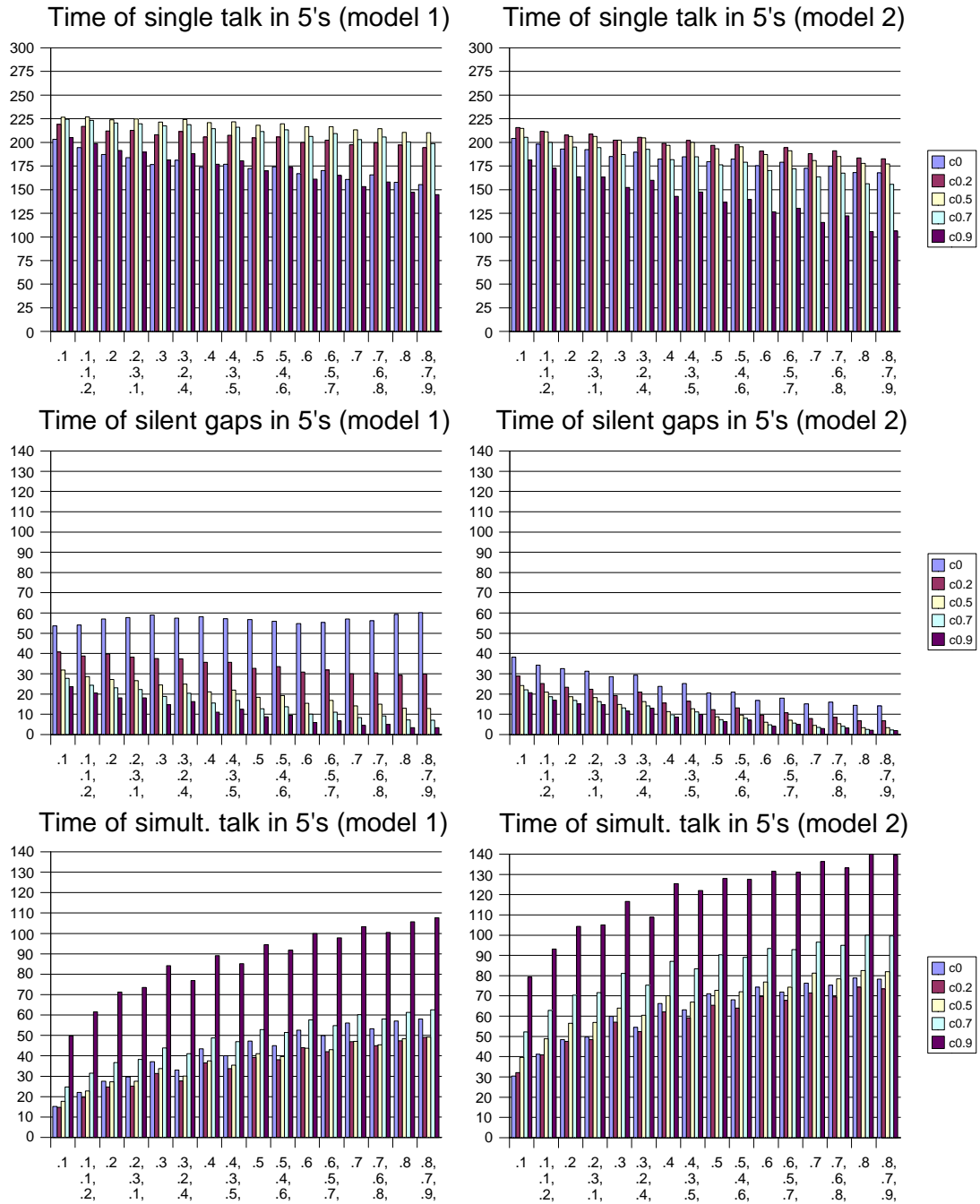
talk, producing more of it in simultaneity with others. Of course, beyond a certain length and frequency the discussions would just become unintelligible to everybody, like drunkards' talk (or perhaps some family discussions): a setting of 0.9 CONFIDENCE for everybody, maybe even 0.7, seem to be past this point.

Judging by how the *number* of simultaneous talks (figure 6.26) is now smaller and more level as TALKATIVENESS increases, their *mean* lengths must increase significantly. Indeed they were calculated as ranging between 0.5–0.6s, 0.6–0.8s, 0.8–1.2s, 1.1–1.7s, and 1.8–3.3s respectively for each CONFIDENCE setting along the TALKATIVENESS range. The *longest* cases of simultaneous talk were between 2.0–5.4s, 3.0–7.0s, 4.0–8.4s, 5.8–13.6s and impressive 8.2–22.4s for the five settings of CONFIDENCE along the TALKATIVENESS range. The shortest lengths remained 0.2s in all cases.

Total time of silent gaps with zero CONFIDENCE is 2–10s and 45s less, respectively at the lowest and highest sides of TALKATIVENESS, as compared to the totals in the basic model. At least in low settings of CONFIDENCE, the 'lost' time must have gone into single talk, since simultaneous talk is always more. As mentioned already, this is then an indication that the new procedures resolve conflicts to one clear 'winner' more often (hence the additional single talk) by reducing the previous recurrent pattern of false-starts and silent gaps.

The resulting numbers of silent gaps and simultaneous talks in figure 6.26 show agents really less fickle in talking simultaneously: even at zero CONFIDENCE they produce similar or smaller numbers than previously at a setting of 0.2. The mean lengths of silent gaps remained between 0.4–0.2s along the parameter range, and the *longest* measured gaps were between 2.8–0.8s more homogeneously in all CONFIDENCE values.

The number of utterances (figure 6.26) increases by 20–60 in model 2, which means there must be more completely simultaneous utterances since single talk does not increase in the same proportion, even decreasing in high CONFIDENCE. Also there is a familiar profile in high TALKATIVENESS: the 'different behaviour' effect of middle likelihoods that results in smaller totals than those at highest or lowest CONFIDENCE. That is, the more frequently *different* decisions of simultaneous speakers about whether to continue talking (or not) makes fewer ones to persist for longer (in the best case, just one), which reduces the number of simultaneous utterances, and thus the totals of utterances (see the corresponding number of incomplete utterances in figure 6.29).

Figure 6.27 charts the distribution of utterances to individual agents with different

Figure 6.26: Number of utterances, silent gaps and multiple talks in models 1 & 2
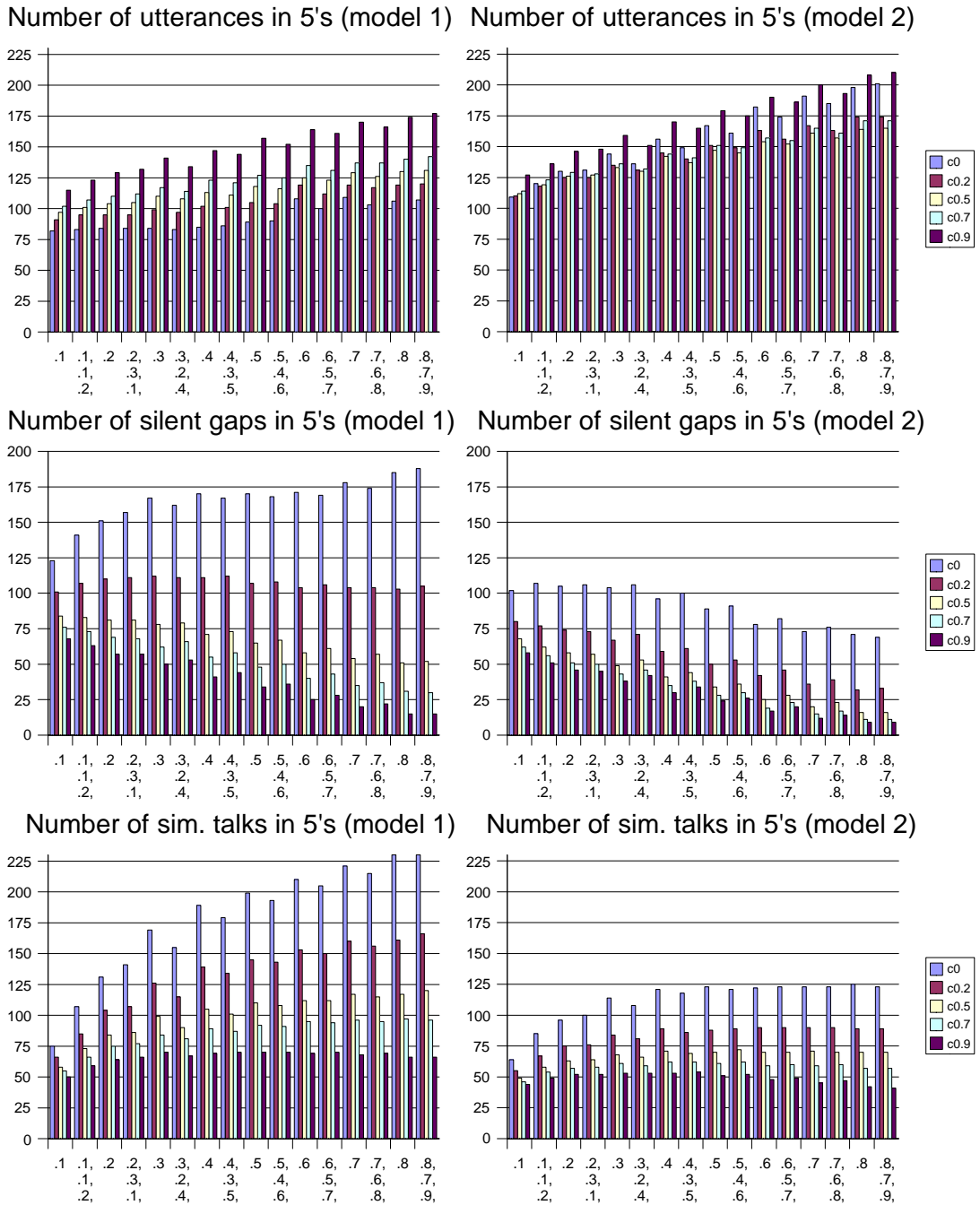
Figure 6.27: Utterance distribution amongst the agents in models 1 & 2
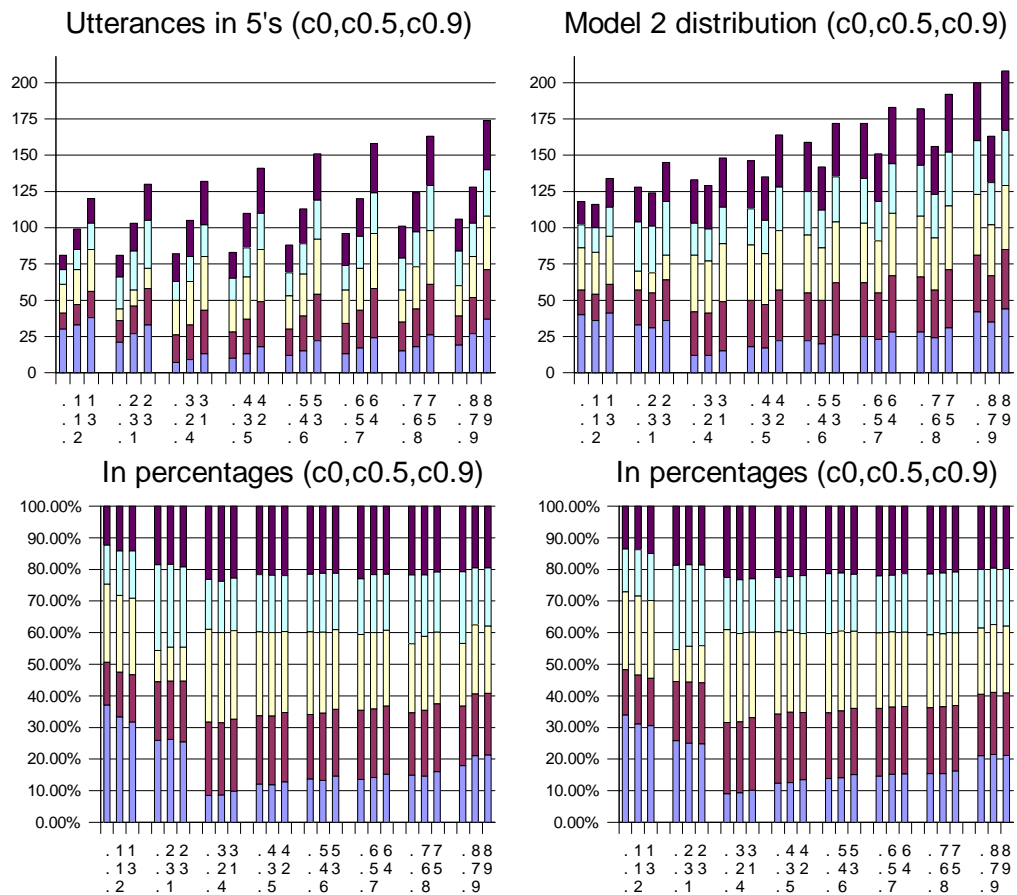
TALKATIVENESS likelihoods in the basic and extended models (left and right charts). The new model generates more utterances in absolute numbers, yet the percentage shares remain very similar to those of the basic model: less talkative agents tend to increase theirs with higher CONFIDENCE by a tinier fraction this time, and more homogeneously, without the 'bumps' visible at high TALKATIVENESS in the basic model. In absolute numbers, the difference is in the character of the various CONFIDENCE settings, as was mentioned just previously for the number of utterances.

Next, the totals of single, multiple and middle starts of talk for the two models appear in figure 6.28. While single starts did not change much in model 2 (though slightly less throughout), the number of episodes of simultaneous starts at low CONFIDENCE likelihoods have dropped significantly, while middle starts have fallen only by a third overall. Though there are less simultaneous starts, they still seem to remain high in comparison with the dwindling numbers of single starts (mind the different scales). The *mean number* of starters in simultaneous starts grows along the parameter range

Figure 6.28: Number of single, multiple and middle starts in models 1 & 2
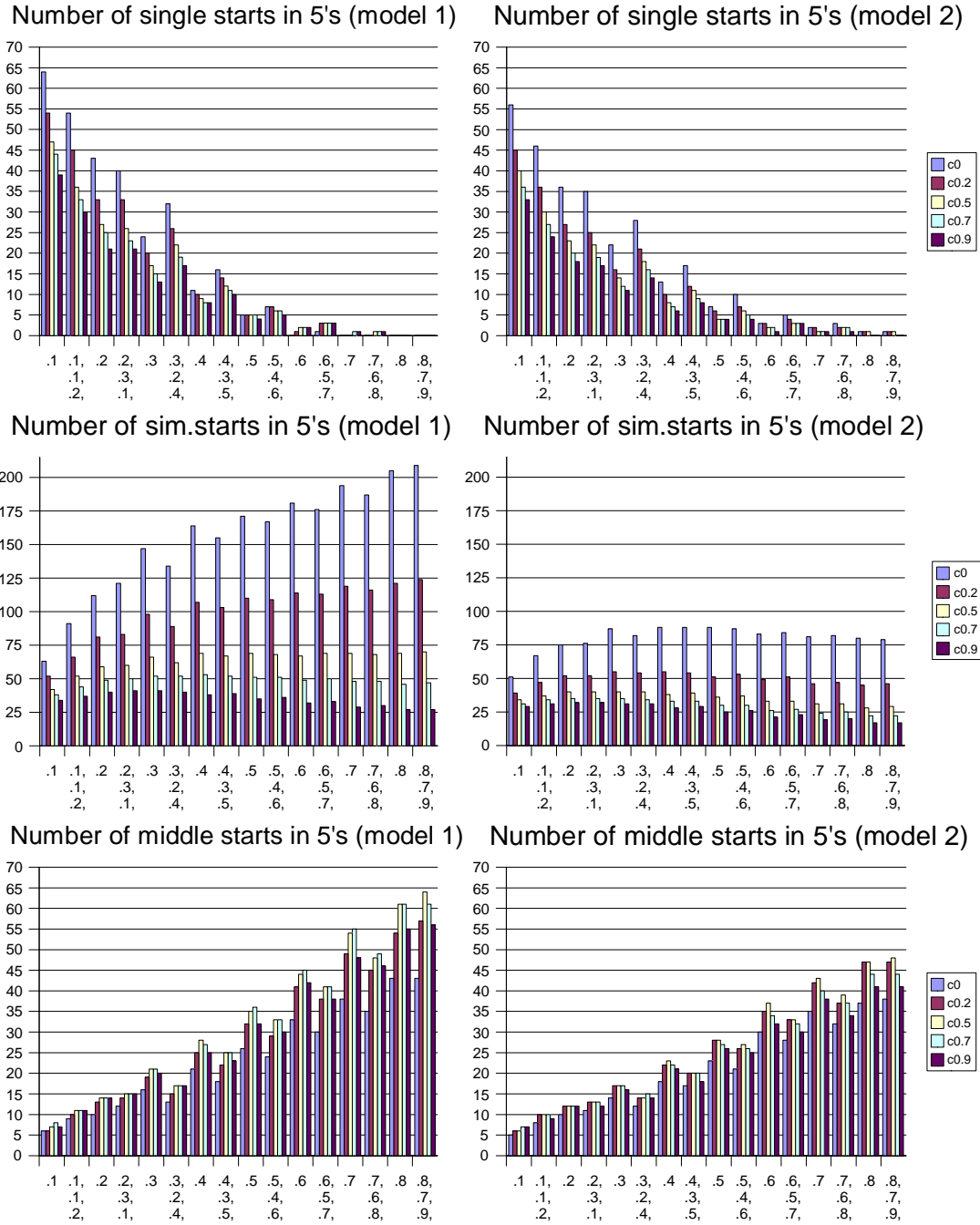
Figure 6.29: Number of false-starts, incompletes and long simult.talks in models 1 & 2

Number of false-starts in 5's (model 1)

Number of false-starts in 5's (model 2)

'Collective' false-starts in 5's (model 1)

'Collective' false-starts in 5's (model 2)

Incomplete utterances in 5's (model 1)

Incomplete utterances in 5's (model 2)

Number of long sim. talks in 5's

Number of long sim.talks in 5's (model 2)
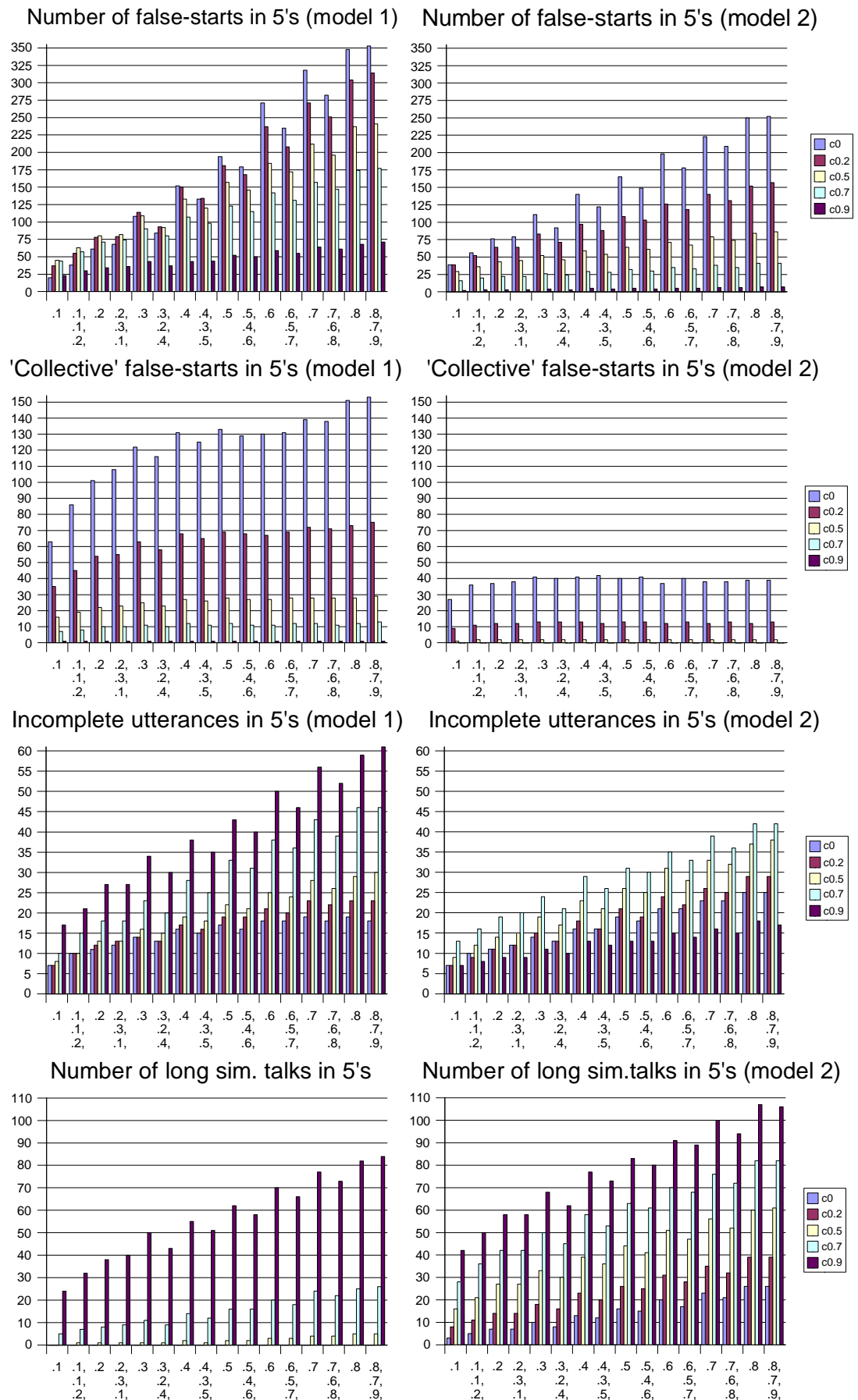
from 2.3 (2.2 at 0.9 CONFIDENCE), to 4.1, 3.8, 3.5, 3.3 and 3.3, respectively for each setting of this attribute at the highest TALKATIVENESS parameters. Higher CONFIDENCE settings drag down the mean number of starters because, logically, more persisting speakers mean less subsequent simultaneous starters.

Finally, the average totals of false-starts, incomplete utterances and long simultaneous talks for the two models appear in figure 6.29. These charts show the more revealing comparisons between the models. With the new procedures, individual false-starts drop down to the levels for 3 or 4-agent groups in the basic model, except that they decrease significantly *more* as CONFIDENCE increases. Collective false-starts also diminish by a factor of 3, more or less. And the number of incomplete utterances partly explains why there are more utterances now, many of them simultaneous: because much fewer utterances end up being abandoned in the new model, even if they are wholly simultaneous with others.

As for long simultaneous talks (the same speakers persisting for one second or more), it is only now that this measure becomes any relevant, since in the basic model there were significant numbers only at high values of CONFIDENCE. With a more complex resolution of simultaneous talk, long episodes tend to happen in possibly more realistic numbers—speaking intuitively, since they clearly occur in discussions with a certain frequency, depending on how talkative are the participants. The *mean* number of speakers persisting simultaneously for a second or more in model 2 varied between 2.0–2.2 at the low side, to 2.1, 2.2, 2.3, 2.4 and 2.6, respectively, for each CONFIDENCE setting at the high TALKATIVENESS side.

6.4   Model 3: different sorts of TRPs

The third model (§4.4.5) adds turn-taking restrictions of different 'sorts' of TRPs in addition to the basic free-for-all type, giving speakers some leeway in controlling oncoming turn-takings. They are thus, in general, able to satisfy immediate interactional goals while talking: whether they have more to say or want somebody else, or anybody, to speak next. With this greater control in turn-taking, less simultaneous starts are expected (which means less simultaneous talks) and, then, less subsequent false-starts. In sum, less unnecessary or unratified attempts at talking depending on the possible 'sorts' of the prior utterances. This extended model was evaluated for the same group

size of 5 but now in a reduced set of TALKATIVENESS parameters with just the *same* likelihoods for all agents, as the horizontal axis in the following charts indicate (one number only meaning it is the same for everybody).

Two attributes are relevant with the new procedures: VERBOSITY and SELECTIVITY, which define respectively how often a speaking agent would produce utterances signalling that the speaker itself has more things to say (and is going to say it) *or* is selecting others to speak next. The model was evaluated in four likelihoods of the first attribute plus two settings of the second for all agents, as indicated by the legends in the right-hand charts of this section. VERBOSITY varies with SELECTIVITY at zero in order to show just the effect of 'More' (more-to-say) utterances in comparison to the *same* attribute variation from the second model (left-hand charts), which does not have them. Save for one setting of 0.7 there, all results are with CONFIDENCE of 0.5 for all in both models (as all other unmentioned attributes).

The two non-zero values of SELECTIVITY for all agents in the evaluation of the new model change the nature of the *free* interaction to a more controlled one where speakers direct the talk to specific others from time to time (as in e.g. individual questions). The model was programmed so that this comes *first* when the agent decides the 'sort' of utterance it will produce: a 0.5 likelihood then really means it will choose select-next utterances 50% of the time. It is *in the rest of the time* that the agent would then decide between free turn-taking utterances (that end in 'pTRP') or 'More' utterances according to its VERBOSITY likelihood; so, continuing to talk is conditional on the levels of (the absence of) next-selection.[6]

VERBOSITY also determines whether speakers want to continue talking without explicitly indicating so. Because of this double function, 'More' is chosen only in *half* the attribute's likelihood: the idea was to emulate that speakers not always (not often, in fact) produce utterances indicating *explicitly* through their contents that they have more to say, even if they may often want to continue talking. Whether this is actually so or not is an empirical question; and it is one issue the *way* that these sorts of utterances are chosen according to a set of parameters (whether probabilities or some other method); another is *what effect* their resulting proportions have for the various measures characterizing the discussion. It is *this* the main focus of the section.

---

[6]Of course it could be that the opposite might make more sense: of only deciding whether selecting someone, or not caring about it, when the speaker is *not* continuing to talk. But that would still leave a problem of deciding between explicitly indicating more-to-come or not: it would separate these more closely related options.

Figure 6.30: Distribution of the generic utterance sorts of model 3

Distribution of utterance sorts in 5's (v0-0.7; s0.2,s0.5)

Utterance sorts in percentages (v0-0.7; s0.2,s0.5)
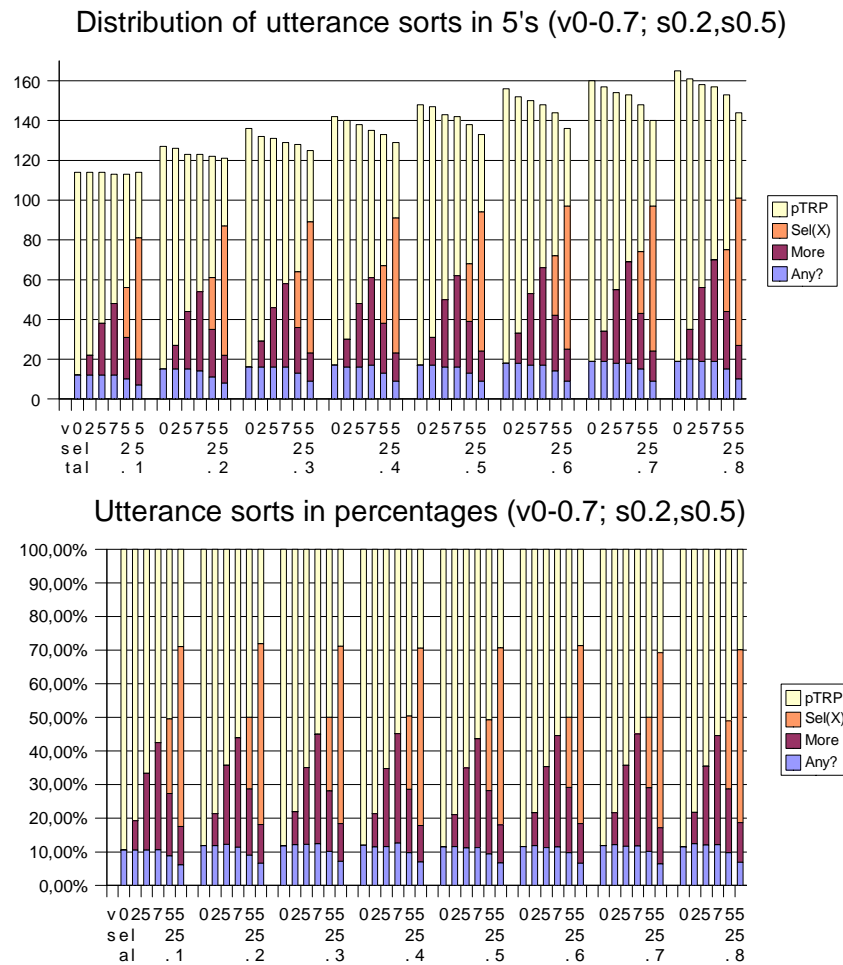
In addition to the two sorts of utterances that restrict turn-taking, a third one (`Any?`) actually stimulates it, encouraging any listener, in fact everybody, to speak next. It should create some extra simultaneous talk, but not much because its frequency was fixed to a default small likelihood for everyone: 10% (of the time no one is selected to speak). The idea again was just to simulate that this type of utterance occurs in a small proportion in many (or some) types of discussions, for example when one or more participants act as informal moderators inviting the group from time to time to talk about one thing or another.

Figure 6.30 charts the model's distribution of the four utterance sorts with the parameters of this evaluation indicated on the horizontal axis: the four VERBOSITY likelihoods at zero SELECTIVITY (increasing the proportions of 'More'), plus the two non-zero SELECTIVITY settings (adding SELECT(X) utterances). These six parameters are

in the usual TALKATIVENESS variation as well, although we see that the differences are only in absolute numbers: the percentage shares remain more or less the same along the range of this attribute.

Figure 6.31 compares the amounts of single talk, silent gaps and simultaneous talk in the two models. The new model generates more single talk by about 2s, 5–7s, 9–11s and 13s in each VERBOSITY likelihood respectively (0, 0.2, 0.5, 0.7), *no matter* how talkative are the agents. The previous model, in comparison, allows only at low TALKATIVENESS (0.1, 0.2) that higher VERBOSITY settings influence (minimally) the amount of single talk, through speakers more oft wanting to continue talking. As TALKATIVENESS increases, however, this is nulled because, there, agents can only succeed in continuing to speak if none other wants to, which gets to happen less and less (soon, never) with a more talkative group. In the new model with 'More' forcefully restricting turn-taking, VERBOSITY now makes a small difference throughout the TALKATIVENESS range.

SELECTIVITY gives higher totals, of about 9s more single talk with 0.2, and 20–28s more with 0.5 for all agents, in comparison to the same model's results without next-selecting utterances. The increase from the previous model, then, without select-next utterances and the same 0.5 VERBOSITY is around 17–20s and 28–37s, respectively for 0.2 and 0.5 SELECTIVITY.

The greater amount of single talk comes from the decreasing amounts of simultaneous talk that can be seen in the bottom right chart of figure 6.31. Simultaneous talk levels are only similar to those of the previous model at zero VERBOSITY, dropping with each higher setting of this attribute that increases the occurrence of 'More' utterances. This happens because (and when) listener agents comply to the restrictions and avoid speaking, therefore generating more single starts instead of simultaneous ones (figure 6.33), and consequently less simultaneous talk. However, they only comply up to a certain point: when they are talkative and confident enough, they may decide to *interrupt* even when the speaker signals more-to-say.

Simultaneous talk is less in the new model by around 2–16s, regularly along the TALKATIVENESS range and more as VERBOSITY rises. Likewise, the presence of select-next utterances in non-zero SELECTIVITY controls turn-taking even more to induce less simultaneous starts and simultaneous talks. The reductions in these cases are between 11–17s and 17–28s, respectively for 0.2 and 0.5 likelihoods, in comparison

Figure 6.31: Times of single talk, silent gaps and multiple talk for models 2 & 3
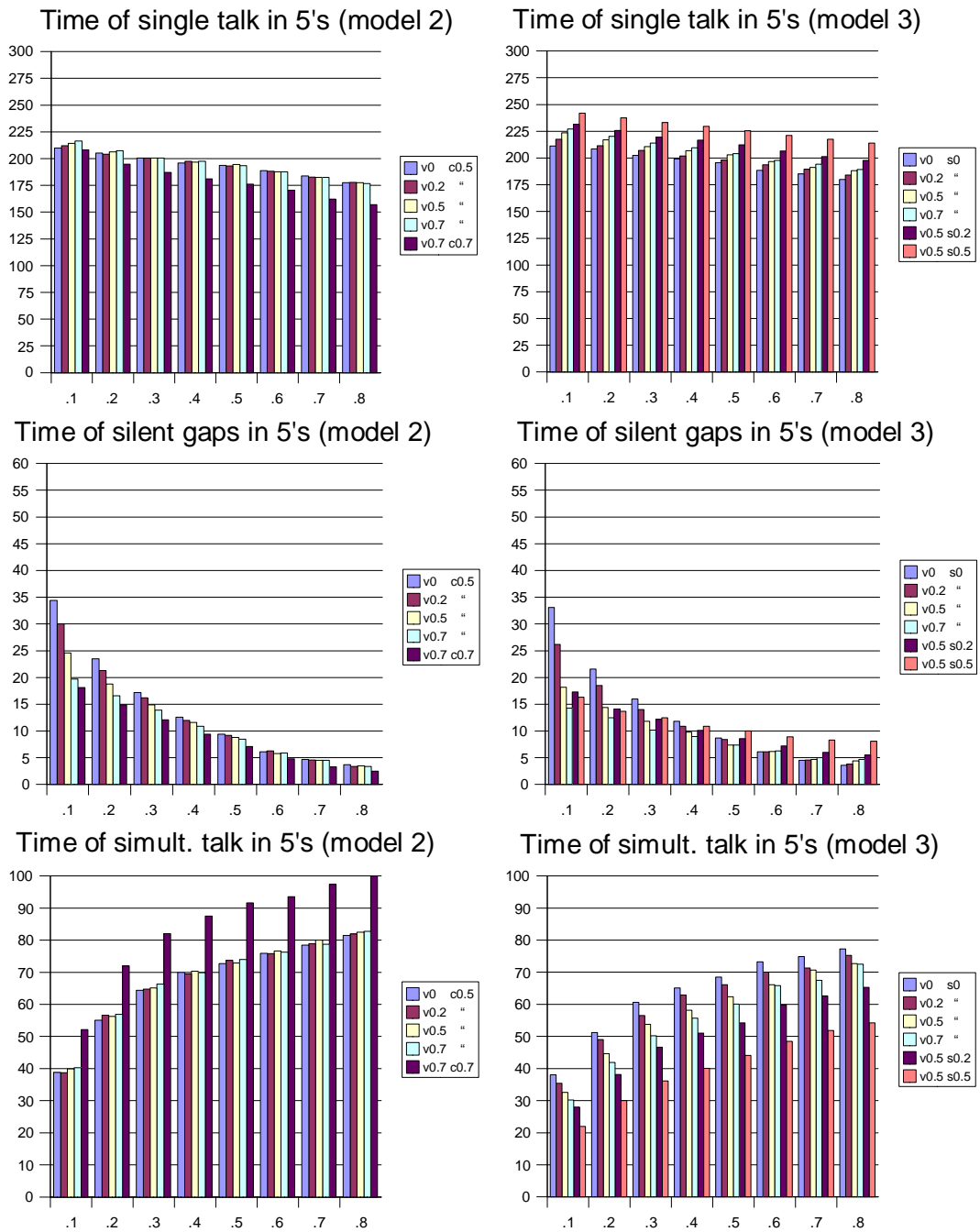
to the second model's results in the same VERBOSITY of 0.5 (and, of course, zero SELECTIVITY).

What distinguishes the time of silent gaps in the new model from the previous is that it falls more at low TALKATIVENESS for each higher VERBOSITY setting. That is the (increasing) time that turns into *pauses* of continuing utterances with the help of 'More' (see the time of 'no talk' for model 3 in appendix 3). But as TALKATIVENESS for everybody increases, those utterances get to be interrupted by (more talkative) agents that decide to speak there nonetheless, causing new silent gaps and matching the same previous levels at high TALKATIVENESS.

Non-zero SELECTIVITY also results in a little more silent gap time by forcing clean speaker transitions through the next-selecting utterances, which (without hesitations that would allow interruption) forces the selected agent alone to take a turn in the usual way designed here: half the time speaking forthwith and overlapping or latching the selecting utterance, or waiting for a minimal silence (one cycle, 0.2s) before starting, in the other half of the time, since EAGERNESS is at 0.5.

Next, figure 6.32 compares the occurrences of all utterances (including continuing and simultaneous), just continuing utterances, silent gaps and simultaneous talks, in the two models. As regularly as single talk time increases, the number of utterances now *decrease* with rising VERBOSITY and SELECTIVITY parameters. This is more pronounced as TALKATIVENESS is higher because the new turn-taking restrictions then prevent an ever increasing number of utterances that would have been produced simultaneously by more talkative agents. Part of these would end up being long simultaneous talks, and figure 6.34 indeed shows a regular (and similar) reduction of their number in the new model too. VERBOSITY at zero results in 1–6 *more* utterances in the new model, but the totals drop as this attribute rises, to some 10 utterances *less* than the previous model with 0.7 VERBOSITY and around 20 less with 0.5 SELECTIVITY.

With 'More', continuing utterances occur even in high TALKATIVENESS now, yet still in low numbers. Their increase is small both because of the low frequency with which the new utterance sort has been programmed to appear (figure 6.30)—that is, sparsely: in half VERBOSITY and in the leftover of SELECTIVITY—, but also by the fact that listeners may still interrupt the speaker if their TALKATIVENESS and CONFIDENCE are high. Because of this, the number of continuing utterances still falls as TALKATIVENESS increases.

Figure 6.32: Number of utterances, silent gaps and multiple talks for models 2 & 3

Number of utterances in 5's (model 2)
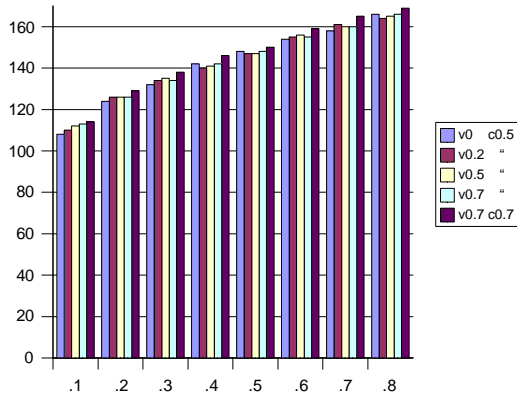
Number of utterances in 5's (model 3)
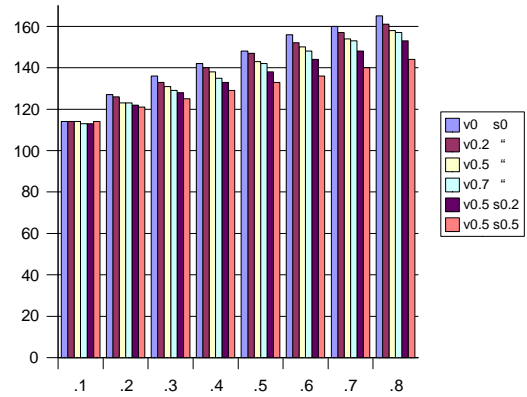
Continuing utterances in 5's (model 2)
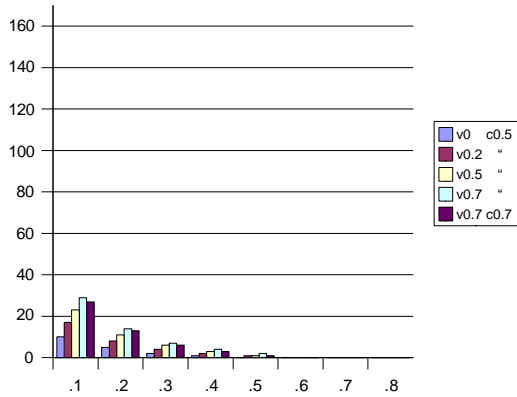
Continuing utterances in 5's (model 3)
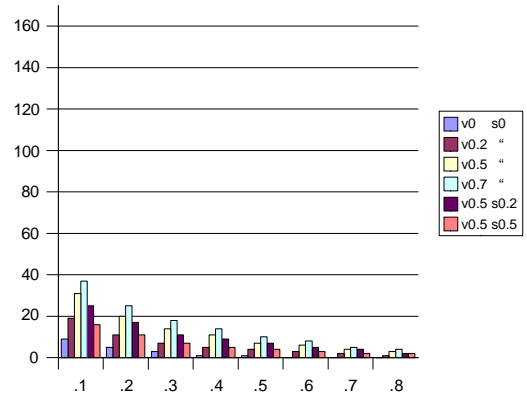
Number of silent gaps in 5's (model 2)
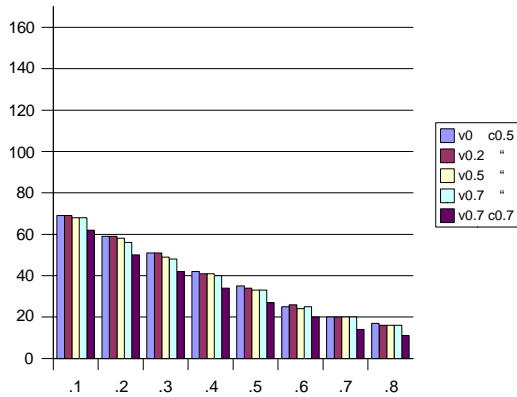
Number of silent gaps in 5's (model 3)
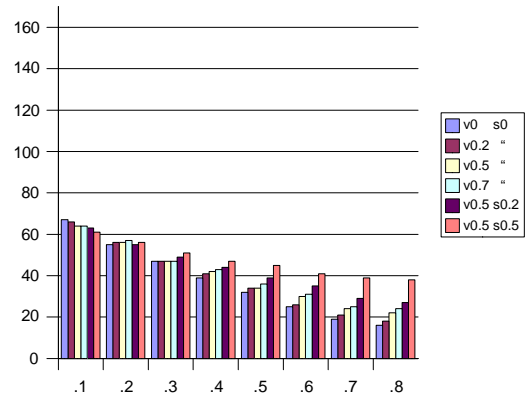
Number of sim.talks in 5's (model 2)
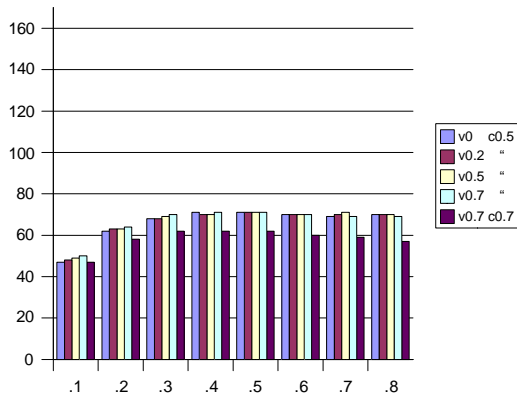
Number of sim. talks in 5's (model 3)
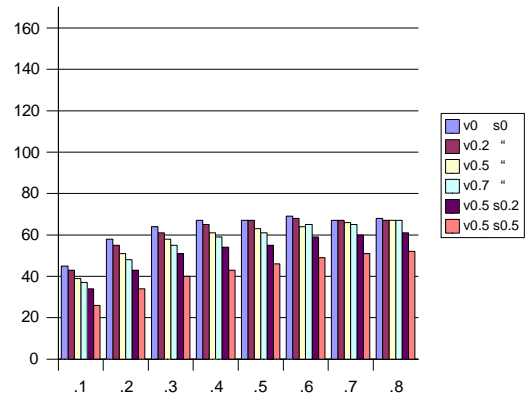
The number of silent gaps in the new model follows the previous descending profile only in low TALKATIVENESS. As it rises, silent gaps get more frequent than in the previous model by up to 10 occurrences more, and even more with positive SELECTIVITY. Mean length of silent gaps is 0.5–0.4s only initially in the lowest TALKATIVENESS and VERBOSITY likelihoods, but soon it is 0.3–0.2s and then just 0.2s when TALKATIVENESS is 0.5 and up. The longest silent gaps created in zero SELECTIVITY were between 2.2–0.8s along the parameter range, and 2.8–0.8s with positive SELECTIVITY. Except for the latter's longer results, mean and longest lengths were more or less the same as the previous model.

Lastly, the number of simultaneous talks in figure 6.32 falls with higher settings of VERBOSITY and SELECTIVITY, but less and less as TALKATIVENESS rises (more interruption). The *mean* length of simultaneous talks increases slowly along these parameters, between 0.8–1.1s. And the *longest* found lengths of continuous simultaneous talks were between 4.2–8.8s, and less with positive SELECTIVITY: up to 6.6s.

In figure 6.33 we can see the biggest improvement of the new procedures: the increase in the number of single starts with more-to-say and select-next utterances. What is clear, too, is that the effectiveness of more-to-say gradually decreases as TALKATIVENESS increases: single starts in the four VERBOSITY settings fall more and more in comparison to the two SELECTIVITY settings.

Nevertheless, the rise in single starts is the most clear evidence of an improvement in the flow of the interaction, through less turn-taking conflicts. It is the fulcrum of all the other improvements measured in this model: in single talk time, simultaneous talk, false-starts, etc. The simulation now starts to take into account some of the mechanisms to restrict turn-taking that are built into the utterances, in their contents, and allow participants to accomplish some immediate interactional goals in talking and in taking turns of talk.

Multiple starts seem to fall (less and less) along the parameter range in converse rate as single starts diminish, since these are complementary measures. They level up at the high TALKATIVENESS side around 23–30 occurrences. The mean number of starters in these simultaneous starts varied between 2.3–3.2 along the parameter range.

As for false-starts, incomplete utterances and long simultaneous talks in figure 6.34, there are improvements too, with numbers falling to levels more or less what one could expect of real discussions. As middle starts remained more or less the same (fig-

Figure 6.33: Number of single, multiple and middle starts for models 2 & 3

Figure 6.34: Number of false-starts, incompletes and long simult.talks for models 2 & 3
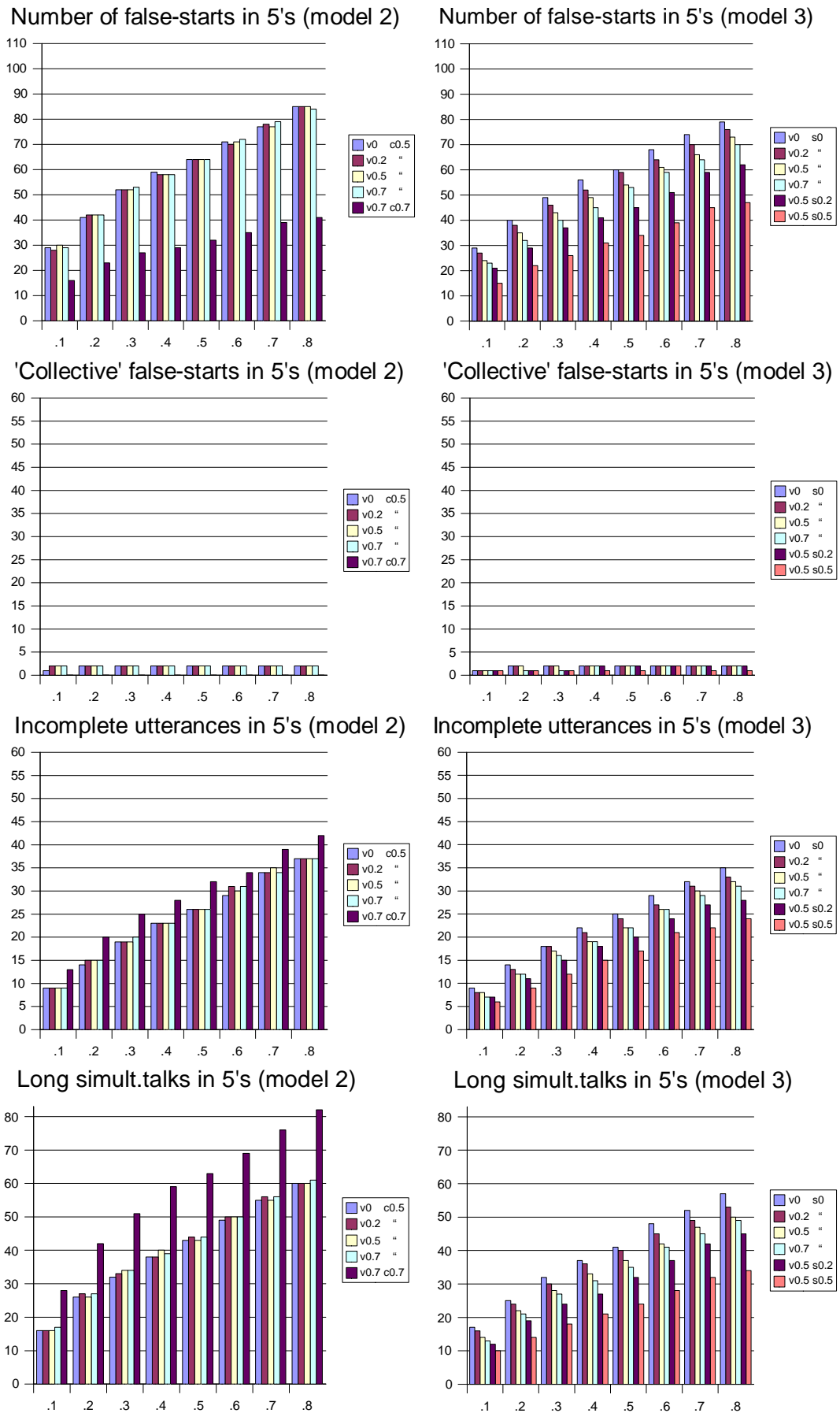
ure 6.33), the reduction of incomplete utterances must be from the smaller numbers of simultaneous starts (relative to the growing single starts) causing less simultaneous talks. Thus, less individual false-starts and long simultaneous talks. The mean number of speakers in long simultaneous talks remained between 2.1–2.2 along the parameter range: so, the vast majority of long simultaneous talks were between two speakers, which matches observations from the literature (Schegloff 2000, see §2.5.1).

6.5   Model 4: hesitations and interruption

The last extension to the basic model (§4.4.6) adds the hesitations and speech discontinuities that are ubiquitous in real talk. Hesitations effectively increase the total length of the utterances, but also introduce the possibility of interruption. This can happen when the speaker is hesitating continuously for a certain time (defined as 0.8s here): then a more talkative and confident listener may (just may) decide to interrupt it by starting to speak (just like after 'More').

Interrupting in the middle of a hesitating speaker's utterance has the same effects of a middle-start, either resulting in an incomplete utterance or simultaneous talk that is subsequently resolved as usual by whoever persists longer through their CONFIDENCE likelihoods. In many cases, however, the hesitating agent simply does not speak anymore and yields the floor, having been thus interrupted by the new starter.

Silent hesitations in the middle of utterances were *not* counted as talk or any type of silence in the evaluation, though non-silent hesitations (`ta- `) *were* counted as talk (perhaps incoherently). A small reduction in the amounts of talk, *single* talk at least, is therefore expected in the results (figure 6.35) without any increase in silence times. Then the amounts of talk and silence now do not add up to the total, because a portion (silent hesitations) is not counted as anything. Anyway, in reality clearly *less* meaningful talk gets to be produced in the same period of time with silent hesitations then.

Hesitating (both silently or not) can also occur *before* someone starts to speak, when the floor is granted to that one by virtue of having been selected to speak *or* when the speaker (already) goes on to begin a new utterance after a pause. In these cases the silent interval just becomes longer with hesitation. Silence then *is* counted as 'no talk'. As this model was evaluated without next-selection (SELECTIVITY zero) for comparison to the previous results in the same four VERBOSITY settings, only the hesitation

while pausing will be relevant here. Its effect should be just that: some increase in the time of pausing, or 'no talk' (appendix 3), *as long as* TALKATIVENESS is low. Because, otherwise, the hesitating agent is simply preempted by others starting to speak as normal. In any case, there shall not be many visible changes in the aggregate results.

As indicated by the legends in the charts to follow, the model was evaluated in two settings of HESITATION for all agents: 0.2 and 0.5. These are the likelihoods that they hesitate *per cycle* of simulation, combined with the four VERBOSITY settings of the prior section and along the same TALKATIVENESS range. It is as if the previous model's results correspond to zero HESITATION. As always, other attributes were fixed at 0.5, *except* SELECTIVITY at 0—so, when comparing the results with those of the previous model, ignore their two non-zero SELECTIVITY settings.

Figure 6.35 presents the times of single talk, silent gaps and simultaneous talk in the new model (right) in comparison to the previous one (left). Single talk is indeed a little less, falling by 11–16s in 0.2 HESITATION (more as VERBOSITY rises) and 17–27s with 0.5. The reductions are more or less the same across the TALKATIVENESS range.

Amount of silent gap time is just a tad less on the low side with 0.5 HESITATION, but it grows along the TALKATIVENESS range amounting to a couple of seconds more than the previous model on the high side. This seems to be, then, the small effect of interruptions causing subsequent false-starts, including a couple more collective ones (figure 6.38), and in turn a few more short intervals (figure 6.36) before someone speaks again.

Similar is the difference in simultaneous talks to the previous model. Total times are slightly smaller on the low side but grow more along the TALKATIVENESS range, and even more with higher HESITATION, to end up being about 5–15s more at the high side. Again the increase comes from simultaneous talk caused by interruption attempts while the speaker hesitates, as the higher numbers of individual false-starts and incomplete utterances attest (figure 6.38).

Number of utterances in the right-hand charts of figure 6.36 show a similar profile of less occurrences at low TALKATIVENESS rising to nearly the same levels of the previous model at the high side. This is even more visible with a higher HESITATION of 0.5. As utterances become longer with more hesitation in their middle, clearly fewer ones get to be realized in the same time, but the increase in interruptions as TALKATIVENESS rises means that more and more utterances are cut short and replaced

Figure 6.35: Times of single talk, silent gaps and multiple talk for models 3 & 4
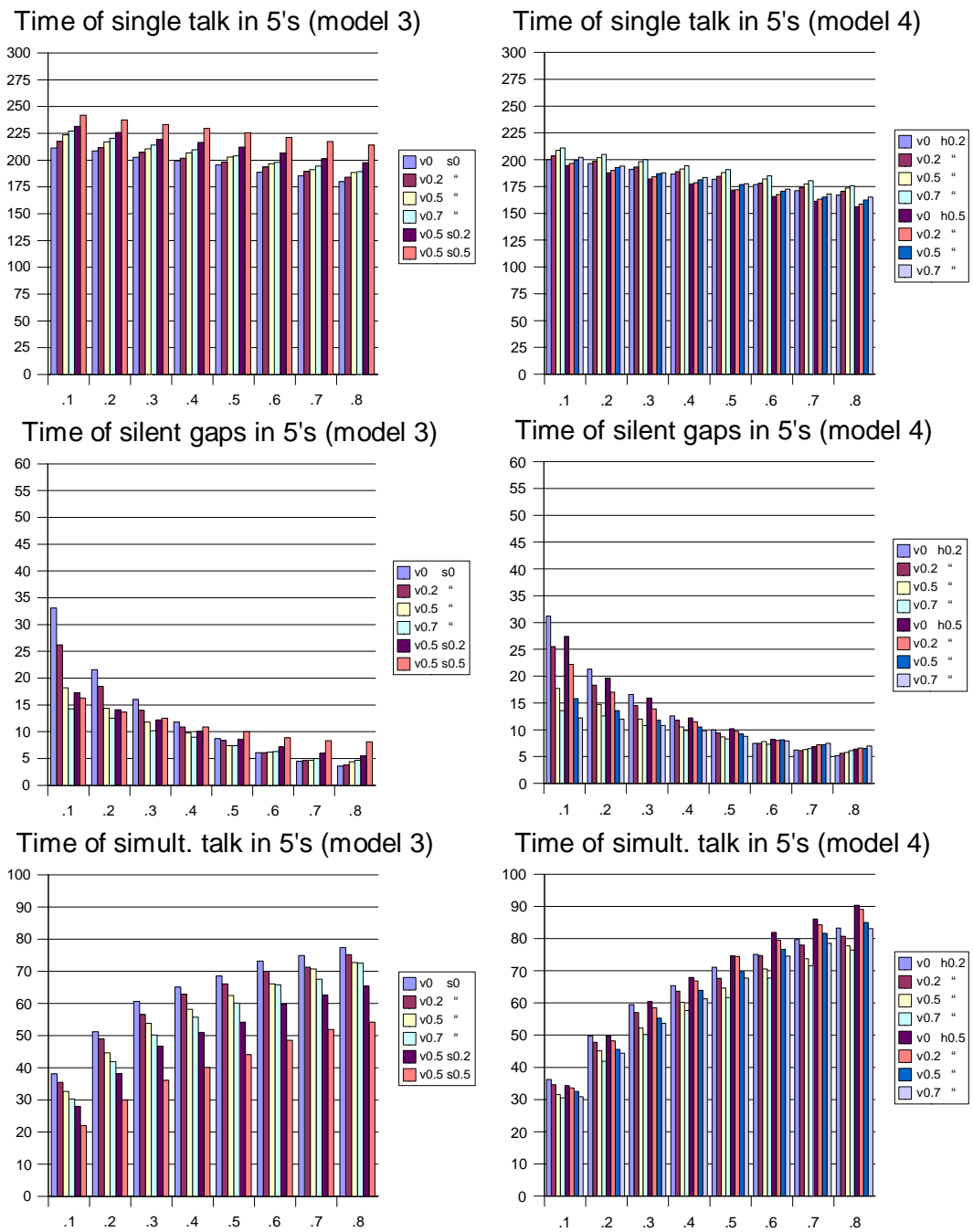
Figure 6.36: Number of utterances, silent gaps and multiple talks for models 3 & 4
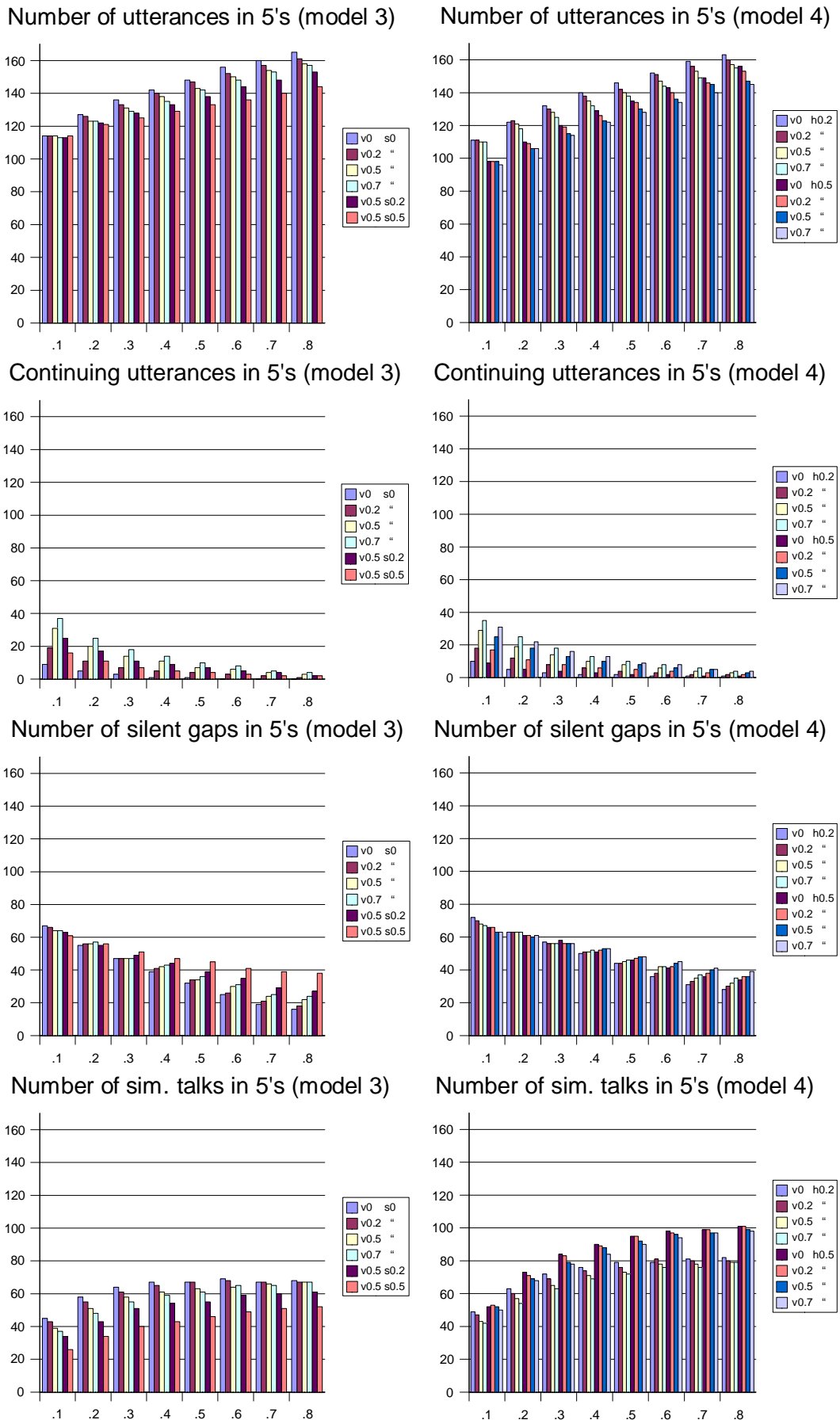
Figure 6.37: Number of single, multiple and middle starts for models 3 & 4

Number of single starts in 5's (model 3)

Number of single starts in 5's (model 4)

Number of sim.starts in 5's (model 3)

Number of sim.starts in 5's (model 4)

Number of middle starts in 5's (model 3)

Number of middle starts in 5's (model 4)
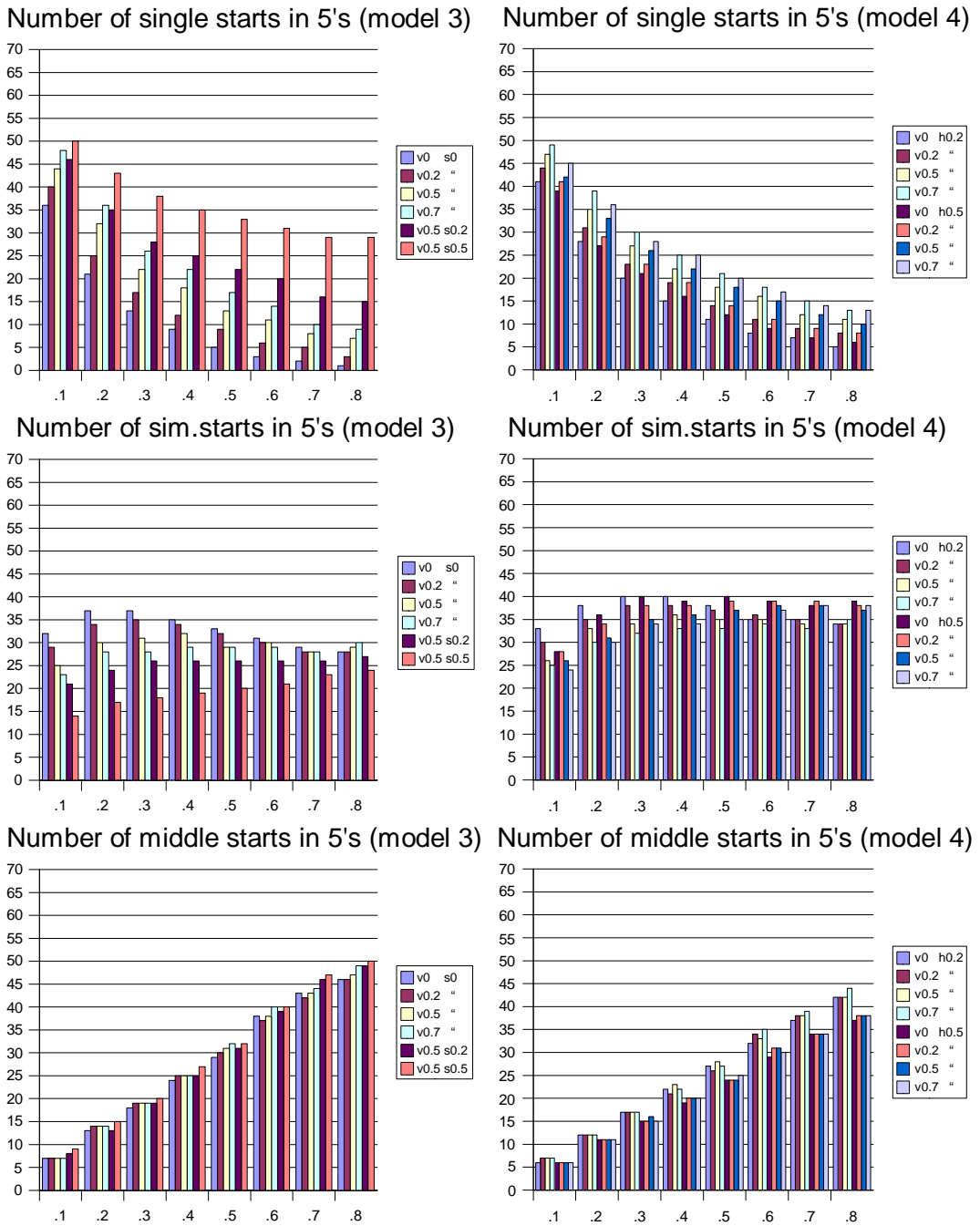
by new ones, reducing their lengths by force—what impatient speakers achieve when they interrupt others, and in a certain sense what was interpreted in the literature as 'clipping redundancy' (Oreström 1983).

The number of silent gaps increases uniformly in the new model by 5–18 occurrences. Their mean lengths got significantly shorter then: mostly 0.2s, only 0.3s at the lower likelihoods of VERBOSITY (0, 0.2) and TALKATIVENESS (0.1–0.3). The *longest* silent gaps that were found ranged more or less in the same intervals as previously: 2.4–0.8s along the TALKATIVENESS range.

The number of simultaneous talks increased just like their equivalent total times, but much more in high TALKATIVENESS; in fact a lot more at 0.5 HESITATION. Mean lengths are smaller than before: 0.7–1s in 0.2 HESITATION and 0.6–0.9s in 0.5 along the parameter range, while being 0.8–1.1s before. The longest simultaneous talks found were measured as more or less the same: between 3.8–8.4s.

Totals of single starts charted in figure 6.37 also follow the same pattern: pretty much the same numbers at low TALKATIVENESS but falling slightly *less* as the attribute increases in the new model with hesitations, so that they are about 5 occurrences more at the high side for the same VERBOSITY settings. Simultaneous starts also increase by as much now in high TALKATIVENESS, while middle starts of utterances actually decreased: that is because what is counted as middle starts are those immediately after pre-TRPs followed by talk (i.e. not at the end of utterances).

Starts of talk during *silent* hesitations are not counted as middle starts but as *single* starts. This means that the number of *actual* (non-interrupting) single starts is smaller than shown: which makes sense since utterances now are longer with hesitations.

Finally, we see in figure 6.38 that hesitations make the number of individual false-starts to increase again, but not to the high levels of model 1. The increase in incomplete utterances (a drop with 0.2 and a hike with 0.5 HESITATION) is perfectly reasonable: the more hesitations the more chances that the utterance is going to be abandoned by someone interrupting any of them.

Lastly, long simultaneous talks amount to a little less than before because two (or more) speakers will only continue talking simultaneously as long as none of them hesitates: whoever happens to do it first will abandon speakership to the other(s) who have not hesitated. Thenceforth, when just one agent remains talking it could hesitate as normal without losing the floor then, unless of course someone decides to interrupt.

Figure 6.38: Number of false-starts, incompletes and long sim.talks for models 3 & 4
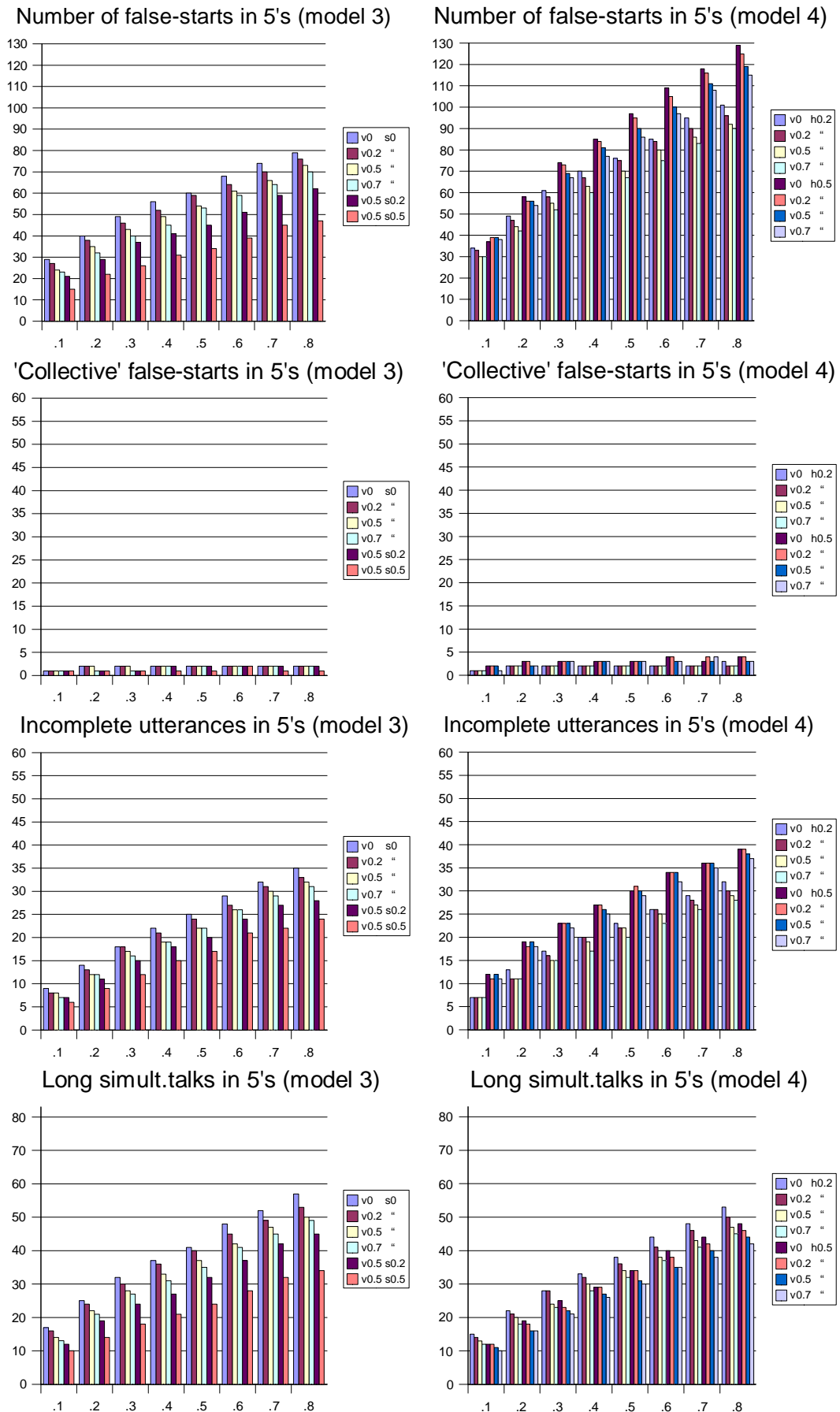
6.6   Summary and discussion

The evaluation of the basic model in some of its parameter variations showed that:

First, as simultaneous talk is concentrated at speaker transitions, the total amount of single talk in a discussion shall depend more directly on the overall lengths of the utterances produced therein—or their maximum length, in the case of the random lengths of this simulation (cf. figure 6.9).[7] So, given a judgement of the maximum length people are expected to talk at a time, the averages resulting from these simulations could provide a rough estimate about the amount of single talk in a discussion, no matter people's actual levels of participation (TALKATIVENESS)—because, as we saw given a reasonable CONFIDENCE in persisting on simultaneous talk, the resulting total averages remained surprisingly similar throughout (cf. figure 6.1).

Of course, single talk time *is* reduced by the amount of simultaneous talk in a discussion, mainly in bigger groups and with higher levels of TALKATIVENESS. In these cases it occurs more because of the parties' EAGERNESS and CONFIDENCE behaviours being more *similar* than different: in the case of this simulation, with non-middle likelihoods (cf. figures 6.1 and 6.15). This gives agents less propensity to vary their turn-taking timings (either speak immediately or wait for silence) and simultaneous talk behaviour (in the basic model, either continue speaking or quit). So the more variantly people behave in a discussion the more single talk they end up producing, and the 'better' for the interchange of talk.

Indeed variation in the timing of turn-takings and in resolving talk conflicts diminished the occurrences of simultaneous starts especially at high TALKATIVENESS levels (cf. figures 6.7 and 6.17), thus reducing simultaneous talks (figures 6.2 and 6.15). In these high levels, lower EAGERNESS and CONFIDENCE likelihoods (between 0.2–0.5) for all agents was best, as it increased the chances of always just one standing out (cf. figures 6.1 and 6.15). But conversely, in lower levels of TALKATIVENESS, middle-to-high likelihoods for all (between 0.5–0.7) was better, generating less simultaneous talk and a little more single talk as it increased the chances of one standing out every time.

Moreover, the evaluation of different TALKATIVENESS parameters for the agents in a group demonstrated the *differing* trend again: groups with various likelihoods to talk generated somewhat *less* simultaneous starts (thus, talks), less overlapped transi-

---

[7]Since utterance lengths are random here, the relationship is not as trivial as if they were, say, in a normal distribution with a mean and standard deviation: then the totals would be derivable directly.

tions and less false-starts—hence a little more single starts, silent gaps and single talk time—than groups that were given the same nearest TALKATIVENESS for all agents (i.e. comparing a group with [0.4, 0.3, 0.5, 0.4] to one with all 0.4 parameters). This was more salient in the middle range of likelihoods between 0.3–0.6 and in bigger groups. So the *more* different are the parties' decisions and their frequency, the better is for the flow of the interaction, quantitatively at least. Individual utterance totals also showed that the higher the difference in TALKATIVENESS values, the (slightly) more that more talkative agents would take from the share of participation of less talkative ones, even those with high EAGERNESS (cf. figure 6.6).

The average amounts of simultaneous talk were measured between 2.7–7.3% of the total discussion time at the lowest TALKATIVENESS levels, and between 11.7–23% at the highest, in any EAGERNESS and in middle CONFIDENCE likelihoods. The mean length of simultaneous talks was around 0.3–0.5s, and the longest simultaneities were found between 1.4–3.8s along the range of TALKATIVENESS parameters. This would conform with observations by Schegloff (2000) (§2.5.1) that most occurrences of simultaneous talk are over very quickly, by the second or third 'beat' that the parties recognise the simultaneity (though this could amount to slightly more than 0.3–0.5s), while some other episodes of simultaneous talk may persist to considerable length.

The mean lengths of silent gaps (i.e. the intervals of turn-taking excluding latched and overlapped transitions) were between 0.4–0.2s: 0.4s with low-to-middle levels of TALKATIVENESS, and 0.3–0.2s at the higher levels and group sizes (0.2s is the minimum from the simulation's granularity, not necessarily the *real* minimum). These averages are smaller than the 0.5s of the mean timing of turn-takings observed by Bull and Aylett (1998), but conform to the observation of Wennerstrom and Siegel (2003) that the probability of turn-taking is highest on average until the first 0.5s after the end of an utterance (whereas afterwards, in silences longer than 0.5s in their data, the probability was higher that the *same* speaker would continue).

The results of model 2, with more elaborate procedures for simultaneous talk, pointed out to an improvement in the smaller average number of simultaneous starts and false-starts (cf. figures 6.28 and 6.29), resulting from bolder behaviours especially in low CONFIDENCE likelihoods. This would contribute to a higher number of utterances (figure 6.26), although less single talk and more simultaneous talk as well (figure 6.25): that is, more complete, short simultaneous utterances produced. It was interesting (and unpredicted), in particular, that the more confident behaviour of the second model

would contribute to a smaller number of simultaneous starts and middle starts (figure 6.28) along the range of the TALKATIVENESS levels, as well as less false-starts (individual and collective, in figure 6.29). This did remedy one of the unsatisfactory profiles of the basic model: the inordinate amount of false-starts created by the high occurrence of simultaneous starts that would come from simple decisions and a 'fickle' behaviour in the presence of simultaneous talk.

The results of model 3, with different sorts of TRPs derived from a few discrete 'types' of utterances, showed how the addition of even a minimal form of contents of the talk would improve nearly all measures of effectiveness of the discussion: more single and less simultaneous talk (figure 6.31) from more single and less multiple starts (6.33), together with a smaller number of individual false-starts, incomplete utterances and 'long' simultaneous utterances (6.34) throughout—at the same time, with a smaller number of utterances in total (6.32). The explicit directing of talk to a specific other party (by select-next utterances) caused the greater impact on the aggregate results; and talk explicitly indicating that the current speaker is going to continue (from more-to-say utterances) also had some improvement in the number of successful continuing utterances (figure 6.32). No doubt the addition of more types of (or actual) contents of talk and more visual and other nonverbal information about the agents' intentions in the turn-taking would improve the interactional profiles even more.

Lastly, the introduction in model 4 of low and middle likelihoods of speaking hesitations for all agents in the group (and their respective interruption, though in somewhat extreme cases of continuous hesitation), by its turn, predictably 'worsened' some of the average measures of the interaction, causing a little more simultaneous starts and, thus, simultaneous talks (cf. figures 6.37 and 6.35) and more individual false-starts, though slightly less incomplete utterances and long simultaneous ones (figure 6.38). However, these increases occurred in middle-to-high levels of TALKATIVENESS; in low levels, they were either minimal or reversed: actually *reducing* the average measures. And a change from a HESITATION of 0.2 to 0.5 for all (meaning that agents would then hesitate half the time they speak) affected the results only minimally.

The conclusion then is that hesitations, while variably present in real talk, interfere very little in the average flow of the interaction. What seems to make the big difference in real conversations is the interpretation of contents and types of contents of talk (not to mention the context and other information) that is much of what makes parties continue to listen or start to speak, in a complex inter-relationship of decisions.

Chapter 7

Conclusion and future possibilities

This is basic research. The simulation is, in a nutshell, a working demonstration of some *issues* regarding turn-taking in a free verbal interaction (small group discussion). To my knowledge, it is the first distributed-agent simulation of group discussion, and the first implementation of a synthesis of the overall descriptive outline of turn-taking that the empirical literature has produced so far. The main use of this work may then be experimental: in discovering turn-taking mechanics by an approximative representration of its operation, and by manipulating and adjusting its fundamental aspects: simultaneous talk, constraints to turn-taking, hesitation and interruption, etc.

So the major contribution of this work, I think, will be in the *ideas* contained in it: the design of the simulation which could be used for similar cognitive models, and the control procedures for speaking or listening with the various 'sorts' of utterances and hesitation. Of course, the configuration of abstract behaviours here would need some adapting to more concrete developments involving actual speech and language generation. Nevertheless, other investigators could use part of these procedures to the control module of conversational agents and talking heads. They would then comprise part of the central control of the talking agent, determining when to start the processes of talk and planning dialogue, and when to interpret spoken talk in listening.

Also, the models and their associated evaluation program could be usable as a simulational 'test-suite' of general conditions of small group discussions or other types of interaction. Maximum utterance lengths, likelihoods to talk and how eager people speak in a real conversation or discussion could be measured up and calculated, and then fed to the simulation to estimate their probable average totals in a large number

of interactions. And the procedures could be further modified and/or extended to suit other similar genres of verbal interaction or complexities of small group discussions unforeseen here. And by modifying the associated evaluation program one would have the framework for quantitative evaluation of the resulting modified interaction too.

As for the current models, the quantitative evaluation has showed a few interesting results. First, utterance lengths are indeed the major determiners of total amount of single talk, with levels of participation (TALKATIVENESS) having a smaller effect. Notwithstanding, simultaneous clashes do decrease the totals of single talk, in bigger groups and higher TALKATIVENESS, and as parties behave more similarly than differently throughout. Bolder behaviours in simultaneous talk (model 2) reduced the amount of false-starts and increased the number of (partly simultaneous) utterances, so that agents talked more simultaneously—not necessarily a bad thing, as sometimes people care more to externalize their thoughts than ensuring they are actually being listened. The addition of even a few discrete 'contents' of talk yielding different 'sorts' of TRPs (model 3), substantially improved the flow of the interaction by directing talk to appropriate listeners, reducing simultaneous starts and talks overall. And, lastly, hesitations that are variably present in most instances of conversation did not depreciated that improvement significantly.

The evaluation also uncovered average totals of single talk, simultaneous talk and of turn-taking timings (silent gaps) that approach the characteristics of conversational interaction described in the literature: overwhelmingly more single talk than multiple talk, with simultaneous clashes being resolved quickly to a single speaker most of the time, but also resulting in false-starts and incomplete utterances. This first attempt to emulate a complex verbal interaction through a multi-agent, probabilistic simulation with abstract behaviours without contents of talk showed a reasonable similarity to the profile of actual (in this case) small group discussions, demonstrating that it is a feasible approach to investigate the minutiae of the turn-taking process quantitatively, through comparisons to real data measurements.

But the models in this thesis are a long way from being satisfactory, of course. They are but a first step. As the evaluation demonstrated, abstract behaviours without actual talk and probabilistic decisions create a lot of multiple starts in detriment to single starts; agents appear to behave too 'swiftly', generating what could be viewed as unrealistic behaviours. The agents would then be considered unnaturally sharp speakers, 'mindless' non-pensive parties (as indeed they are).

They do not have any attention variation or interest variation, nor any provision for longer response-times, except what is simulated for feedback responses. Agents' response-times are always one-cycle long, so it is equated to the length of the cycle, the resolution of the simulation. Agents also do not change their likelihoods throughout the discussion, varying for example the *interest* they might have in the talk; or their turn-taking strategies facing repeated situations when they are not able to talk. Also, the symbolic behaviours with explicit pre-TRP indications are a long way from the reality of the complex combination of linguistic and prosodic cues and visual nonverbal behaviours that together indicate the transition-relevance places, and encourage or discourage turn-taking.

Finally, these textual behaviours were designed for viewing results on-screen, not for connecting the simulation to a speech and language generator for *real* testing of what the discussion would look like. These various obvious limitations of this initial model show us obvious continuing extensions as future work.

7.1 Assynchronous framework: other response-times

The central loop of the simulation generates a *synchronous* simulation: each 'moment' has its behaviours synchronized outwith the agents. This is twinned with behaviours having a certain length that is assumed of each moment, representing the minimal simulated unit of time: the cycle. It is the simplest design and probably the most feasible for other similar cognitive models and applications of this model.

But another possible configuration that could represent exactly *how* talk occurs in groups of people is an assynchronous simulation. An assyncronous multi-agent configuration would be a much more complex one: agents would live as independent processes (or threads), but they would still have to have a channel of communication, the 'environment' (now assynchronous). The environment, however, would have to have a broadcasting architecture forwarding each 'message' (or packet of behaviours) to all the agents assynchronously. This channel distribution would work in a higher priority than the agents themselves, so they receive the messages (talk and other perceptual behaviours) more or less at the same time.

Agents in such a system would have to have a fundamental feature, that the present model does not have: *buffering of the input*. As the communication is assynchronous,

they would have to decide whether to ignore or 'process' (interpret) the various messages they receive and react to them assynchronously, generally sometime later (emulating the response-times humans have in responding to other actions). This in all likelihood would have to mean various cascading processes (also threads?) of interpretation and reaction, as in some cognitive psychological models of speech and language interpretation, e.g. Levelt (1989).

Another aspect such a system would have to represent, which the present simple model also does not do, is a certain (or variable) response-time to the behaviours. In the present model, response-times are mostly one-cycle, precisely because agents do not have any input buffering. They are simple automata that react immediately to the behaviours of each cycle since these will disappear in the next moment. Consequently, agents behave 'fast' and sharply like the automata they are, scaled to the resolution of the simulation defined by the *cycle-time* parameter.

In sum, what such a system should incorporate for a better reproduction of human cognitive processes is a simulation of response-times: reaction times that would be more than 'one cycle' of simulation. They should fluctuate with the attention and interest of the conversants in the talk and with any *conflict* of simultaneous talk and behaviours: which cause, for example, the hitches that generally indicate the recognition of a conflict in simultaneous talk.

7.2 Variable attribute values

Another limitation of this simulation regarding the variability of patterns of group discussions is that the agent attributes remain constant throughout the talk. In actual talk, the likelihoods of making the various decisions described in §4.2 would vary a lot, depending on several factors: so the likelihoods of taking the various turn-taking decisions in the simulation should *fluctuate* throughout the discussion.[1]

For example, TALKATIVENESS could vary based on aspects of the talk and other participants involved, if these things are modelled too: as mentioned earlier, the topic being talked about and the participant's knowledge or an associated relevant informa-

---

[1]Of course, this assumes no contents of talk are being generated and interpreted; if they are, then, decisions to talk would naturally relate to them. But this does not mean other parameters of this simulation could not be used anymore. A *gradual* inclusion of contents of talk replacing probabilistic decisions like TALKATIVENESS first, is probably the best course of a future development of this sort.

tion about it, his or her interest, attention, emotional states, acquaintanceship, gender and age of the interlocutors at each interaction, and so forth. It could as well fluctuate in the group as a whole (say, in 'collective' patterns like in flocks) in the course of a discussion to simulate how 'heated' it becomes, or how much everyone is (perceived to be) willing to talk.

This in turn would influence everyone's EAGERNESS and CONFIDENCE according to this perception of the 'competition to talk'. If overall talkativeness is low, if there are long pauses and hesitations with no interruption, and long gaps without talk, then both parameters would remain on a 'careful' or 'polite' low level. If more participants appear to be wanting to talk every time, or they talk more often and give more feedback, not giving in to others—in sum, they participate more—, eagerness and confidence would have to increase to what is needed if one wants to get a turn of talk, eventually.

People in general would start to speak earlier and louder or would not give up easily (and thus create more simultaneous talk) when a discussion gets accalarated. This change may also be directed towards specific others: for example, when someone seems always to speak earlier and interrupt others, or always 'wins' the turns, preventing others from talking.

Participants, therefore, *in order to suit their various purposes in conversation* need to change and adapt their strategies of turn-taking along the conversation in view of other participants, the talk, and the *history* of the interaction so far—like when they repeatedly fail to get a turn at talk. These various outcomes could feed back to the turn-taking parameters, which in turn would change the nature of the interaction, then feed back again and so on. It would be good if agents are able *learn* turn-taking parameters (like EAGERNESS and CONFIDENCE) in their interactional situation and group, taking into account the others' apparent needs to talk as well.

CONFIDENCE could also vary with the perceived appropriateness to talk at certain moments: if someone was eager to speak in the middle of another's utterance, he or she should be less likely to continue than the established speaker—unless it is something important, more than what the speaker was so far saying. The same would apply for the speaker, who should have varying degrees of confidence in continuing to talk.

On a purely statistical view of the patterns of interaction, TALKATIVENESS could seem to depend on how recently one has last talked: the more recently, the more likely to talk again soon (Stasser and Taylor 1991). The same seems to hold for feedback, and

interchangeably; that is, recent feedback would seem to affect likelihood to talk and vice-versa. If someone has been quiet for a while then starts giving feedback, it is a sign of more interested in the current discussion, so it seems likely he or she may want to speak in the short run.

Lastly, it is evident when analysing the amount of talk and feedback statistically in small group discussions that less talkative participants give more feedback, and vice-versa. As if they would feel compelled to signal periodically that they are paying attention, even though (or because) they are not speaking much. The present model partly incorporates such a disposition in the order of decisions: only decide about feedback when *not* wanting to talk; so the more one talks, the less feedback it gives. But on the other hand FEEDBACK would still have to vary to support this need to 'show attentivity', to avoid 'awkward silences'.

Thus we begin to see how all these parameters are interconnected, and how this whole 'systematics' of their fluctuation would be complex. The various outcomes of the interaction, including what participants are talking about, would feed back to change the attributes. This fluctuation would be a sort of 'second-order' model on top of the current one. It would bear less on modelling of turn-taking and more on the *dynamics* of the interaction in small group discussions, on the simulation of the group discussion *per se*, which is outside the scope of the thesis. The agent modelling becomes more complex when turn-taking decisions (or in this model, their likelihoods) have to be changed and adapted as a consequence of what is being talked about, how the other participants behave and are perceived to behave, and the history of the interaction.

7.3   Cues of turn-taking and nonverbal behaviours

As discussed in chapter 4, the pre-TRPs represent a significant simplification of reality. They are explicit indications of possible completion of the utterance, given by the speaker, rather than coming from the listeners' interpretation of the various auditory, linguistic and visual cues: syntax-semantics, pragmatics, discourse, prosody, loudness and rhythm (tempo, etc), and nonverbal behaviours (gestures, gaze, etc). Furthermore, the pre-TRPs have all *the same force* in the simulation, indistinguishably—precisely because there is no representation of talk and its contents.

When I introduced the 'sorts' of pre-TRPs in §4.1.7, I said that there may be a wider

and finer range of TRPs with regards to their restrictions to turn-taking, but that the four sorts would represent reasonable *discrete* generalizations. And indeed, with actual talk, each of the four 'sorts' of TRP probably have a continuously varying degree with which they restrict, encourage or discourage turn-taking. There may be varying degrees with which utterances indicate select-next, more-to-say or anybody-to-talk next, inducing varying likelihoods with which agents (or specific ones) would decide to talk then— inducing their varying TALKATIVENESS, to use parameters of this simulation.

So an obvious possible expansion of the present model would be to change the behaviours to incorporate representative (yet still symbolic) cues of possible completion of the utterances, and also turn-yielding cues. For example, the 'pTRP' behaviour could be replaced by various degrees of 'coming' to a possible completion, as in Padilha and Carletta (2003): `tal3 tal2 tal1 tal0`. These various degrees could be a continuous repetition to different certainties: e.g. `talk tal2 tal1 talk tal3 tal2 tal2 tal1 tal2 tal1 tal0`; the smaller the number, the more certain an utterance completion would seem to be. Agents with different EAGERNESS parameters could decide to speak at various of these points.

A better way to represent the various levels of speech cues and the various modalities of nonverbal behaviours would be with agents returning more than one behaviour each time: one for each modality forming a 'packet' of behaviours. For example, talk could be replaced (or complemented) by `syn2 syn1 syn0` for the various levels of syntactic completeness; then `int2 int1 int0` for the various degrees with which intonation indicates possible utterance completion; then `sem2 sem1 sem0` for various levels of semantic or pragmatic completeness (which relates to 'More').

Nonverbal behaviours such as gaze, posture shifts and arm and head gestures could all be represented by additional behaviours in different modalities, as simulated in Padilha and Carletta (2002). For example: `G->AgtA POST GEST` could be a 'packet' of nonverbal behaviours indicating gazing at agent A, changing body posture and arm gesture, which could indicate that the agent is continuing or starting to speak.

Among other things, this would eliminate the need for the overlapping 'tal_' behaviour. Now the specific pre-TRPs—which actually embody a summary of the *contents* of each ending utterance with regards to turn-taking restrictions—would not have to be 'pre-TRPs' anymore, because turn-yielding or holding indications would be given off by actual linguistic and paralinguistic cues. They could come at any point in the

utterance then, representing that that kind of information (of turn-taking restriction, or encouragement, etc) could come at any point within the unfolding utterance, even near the beginning: e.g. "John, what do you. . . ", "Anybody wants to. . . ", "First of all, etc. . . ". Listeners would thus be able to predict what one is going to say and perhaps interrupt it in the middle.

7.4   Plugging speech

As the simulation is symbolic, it is difficult to assess how realistic are the patterns and general behaviours it generates. One possibility to evaluate them qualitatively would be to connect the simulation to a speech-generation output. As long as an agent's action is 'talk', the output could be a (repetitive) "blah blah" recording, with different voices for the various agents. Other short clips that could be played back for the various simple behaviours of this simulation would be "uhuh", "huh?", and "bla-" (interrupted) for 'ta- ', and so on. Of course, this would result in very strange artificial discussions. But it would provide a more direct way to assess naturality on the implemented practices of turn-taking, in a 'cheap' way. A sort of 'wind tunnel' for tweaking and improving the interactional control of conversational agents.

Other more elaborated schemes could be thought, maybe even more worthwhile at the end than this one. Instead of simple short 'blah's for each 'talk', there could be a pool of pre-recorded utterances with various voices for the different agents. Each time an agent would start a new utterance it would choose (randomly?) one pre-recorded utterance and play it seamlessly through as long as the agent has the speakership and is sending 'talk' behaviours. Hesitations could be the same short pre-recorded "bla-" or a "erm" recording. The point in the pre-recorded utterance that the speaker agent is playing back could be saved in the *length* variable (§4.4.2), to be played thenceforth at the next 'talk' of the subsequent cycle.

One problem with this approach is that the resulting discussion with the conjunction of various arbitrary pre-recorded utterances would not make much sense. This would make it difficult for lay subjects to judge: they would have to rate timings of turn-taking having to overlook the *sense* of what is being said in a highly artificial discussion. But if we ignore what the agents are saying and focus only on the micro-level decisions, it could provide a decent way to assess and adjust new models of turn-taking.

Appendix A

Code listings

The simulation in its various models and extended procedures were implemented in simple Java programs. The first file, `model.java`, constains the classes:

- `Talker`, defining the agent, its attributes and 'program of operation' with the procedures of the basic model of figure 4.3, section 4.4.2;

- `Group`, implementing the multi-agent group and the blackboard framework;

- and `model`, the main class with the execution and output cycle-by-cycle.

The other files just contain different versions of the `Talker` class with extended procedures of agent operation corresponding to the same extended models described in chapter 4. File `model2.java` has the simultaneous talk practices of figure 4.5, §4.4.4; file `model3.java` the sorts of turn-taking constraints of figure 4.6, §4.4.5; and file `model4.java` has the procedures for hesitation of figure 4.7, §4.4.6. These extended procedures must be compiled after `model.java` since only this one has the framework and the front-end classes of the simulation.

The listings here are for convenient browsing. If you would like to obtain the files instead of typing it all, you will probably be able to download them from my site about this work, wherever it may be in the future (undefined at the time of writing).

```
[model.java]

/******************************************************************************
   Turn-taking model for a simulation of small group discussion
   by Emiliano G. Padilha           previous versions 21/Mar/02, May/03, Feb/04
                                     4th version: begun Jul/04
********************************************************************************
Behaviours:  -  , talk, TALK, ta- ,  pTRP, More, AgtX, Any?,  uhuh, huh?

Basic model: no AgtX/Any?/More(only pTRP), no hesitation, nonverbal behaviours
             (hence, attributes SELECTIVITY, HESITATION, NONVERBAL ain't used yet)

Compile with:  $ javac model.java
Run with:      $ java model          (a page with instructions will appear)
********************************************************************************/
import java.io.*;

class Talker         //the agents representing conversants in the discussion
{
  final int myself;    //1,2,3,...: simple identification of self
  double TALKATIVENESS, //likelihood of wanting to talk (with no one or someone)
         EAGERNESS,     //likelihood to speak forthwith or wait for a pause
         FEEDBACK,      //likelihood to give feedback (either uhuh or huh?)
         CONFIDENCE,    //likelihood to persist in talking simultaneously
         ATTENTION,     //(unused) mean time to perceive that others are talking
         VERBOSITY,     //likelihood to continue to talk starting new utterances
         TRPAUSING,     //mean length of TRP pauses (after utterances) in secs.
         INTERACTIVITY, //likelihood of mid-utterance TRPs (pTRP, More) and tal_
         HESITATION,    //likelihood of hesitations or disfluencies in mid-talk
         SELECTIVITY,   //likelihood of selecting a next speaker at a TRP
         NONVERBAL;     //(unused) likelihood of GEStures, POSTure shifts, NODS

 Talker(int id)  //initialize the agents with default values
 {
   myself = id;           FEEDBACK  = 0.3;  VERBOSITY  = 0.5;  INTERACTIVITY= 0.5;
   TALKATIVENESS= 0.2;  CONFIDENCE= 0.6;  TRPAUSING  = 0.5;  HESITATION   = 0.3;
   EAGERNESS   = 0.4;  ATTENTION = 0.3;  SELECTIVITY= 0.1;  NONVERBAL    = 0.3;
 }

 void attribute(char code, double value)
 {                                       //command-line setting of the attributes
   switch(code)
   {
    case't': TALKATIVENESS= value; break;  case'v': VERBOSITY    = value; break;
    case'e': EAGERNESS    = value; break;  case'p': TRPAUSING    = value; break;
    case'f': FEEDBACK     = value; break;  case'i': INTERACTIVITY= value; break;
    case'c': CONFIDENCE   = value; break;  case'h': HESITATION   = value; break;
    case'a': ATTENTION    = value; break;  case's': SELECTIVITY  = value; break;
                                           case'n': NONVERBAL    = value; break;
    default: System.out.println("*** Wrong attribute code: "+code+"\n");
   }
   return;
 }
             //agent names, for printing and SELECT(X)
static String Names[]={"????","AgtA","AgtB","AgtC","AgtD","AgtE","AgtF","AgtG"};
```

```
 //parameters global to the simulation, instead of individual to each agent
 static double cycleTime = 0.2;    //clock-cycle is 0.2s (5 cycles per second)
 static double MaxUtterances = 4;  //maximum length of any utterance (4 seconds)
 static double NotUnderstand = 0.1;//frequency of negative feedback 'huh?' (10%)

 //agents' memory: states/timers (make the model 'almost' finite-state?)
 int speaker, length, tail, pause, midTRPs, feedback, huhed, huhfor;
 boolean wannaTalk;

String  cycle(String was[]) //cycle activation: the agents' program of operation
{
 int talked=0;     //check input at every cycle to see who's talking (besides me)
 for (int i=1; i< was.length ;i++) if (i!= myself) //preferably any NEW speakers
 {
   if (isTalk(was[i]) && (talked==0 || talked==speaker ||
                                       i!=speaker && Test(0.5)))  talked= i;
   if (was[i]=="huh?") huhed= i; //any negative feedback?
 }
 if (huhed> 0 && was[huhed]!="huh?") //respond when the 'huh?' vocalization ends
    if (huhfor==myself)
       return was[myself].startsWith("tal")? " - ":StartTalk();  else huhed= 0;
 if (was[speaker]=="pTRP") huhfor= speaker;  //to whom exactly is the 'huh?' for

 if (speaker==myself) //I'm the speaker
 {
   if (--length> 0) //utterance: talk talk talk...
   {
     if (talked> 0 && was[talked]!="tal_" && !Test(CONFIDENCE))
        { speaker= talked;  return " .. "; }               //simultaneous talk
     if (length% midTRPs==0 && Test(INTERACTIVITY)) return "pTRP";
     return "talk";                                       //mid-utterance pTRPs
   }
   if (length==0) //utterance ending: decide on continuing, signal pre-TRP
   {              //and add a 'tail' of trailing-off talk between 0.2-0.6s
     wannaTalk= Test(VERBOSITY);    tail= -1-(Test(INTERACTIVITY)?around(0.4):0);
     pause= tail-around(TRPAUSING);   return "pTRP";
   }
   if (length> tail) return "tal_";          //finish the utterance, then pause
   if (length> pause && talked==0) return " .  "; //"different" silence: a pause
 }

 if (--feedback> 0) return was[myself];  //say the rest of the feedback vocaliz.
 if (feedback==0 && was[myself].startsWith(" "))return GiveFeedback(); //delay'd

 if (talked==0 && wannaTalk || speaker==0 && Test(TALKATIVENESS))
    return StartTalk();     //anyone who wannaTalk goes first, then 2nd starters
 speaker= talked;          //update speaker
 if (was[speaker]=="pTRP")
 {
   wannaTalk= Test(TALKATIVENESS); //decide whether to talk (forget previous), &
   if (wannaTalk && Test(EAGERNESS)) return StartTalk();  //whether to start now
   if (Test(FEEDBACK)) if (wannaTalk || Test(FEEDBACK))
                          return GiveFeedback();  else feedback= around(0.5);
 }                                             //delay feedback ~~ 0.2-0.8s
 return " -  ";
}
```

```
                              //for a real normal distribution: Math.nextGaussian()
  int around(double mean){ return (int)Math.round(
                                   (mean*Group.random()+mean/2.0) /cycleTime); }

  boolean  Test(double attr){ return Group.random()< attr; } //problst. decisions
  boolean isTalk(String did){ return did.startsWith("ta") || did=="pTRP"; }

  String StartTalk()  //start utterance immediately, set length, midTRP intervals
  {
    length= 1+(int) Math.round(Group.random()*MaxUtterances/cycleTime); //random#
    midTRPs= around(2.0);    //possible mid-utterance pre-TRPs roughly at each 2s
    speaker= myself;         wannaTalk= false;       huhed= 0;
    return "talk";
  }
  String GiveFeedback() { feedback= around(0.4);    //feedback length ~~ 0.2-0.6s
                          return Test(NotUnderstand)? "huh?":"uhuh"; }
}


class Group
{
 static int size;        // group size
 Talker participant[];   // the participants of the group discussion
 String environment[],   // the behaviours of the current cycle (next blackboard)
        blackboard[];     // and behaviours of the previous one

 Group( int groupsize)
 {
   size = groupsize;
   participant= new Talker[size+1];    //indexing is [1..size], not [0..size-1]
   environment= new String[size+1];
   blackboard = new String[size+1];    //initialize the blackboards
   for (int i=1;  i<= size  ;i++)  participant[i]= new Talker(i);
   for (int i=0;  i<= size  ;i++)  environment[i]= blackboard[i]= " - ";
 }
 String[] cycle() //run one cycle of the simulated discussion
 {
   for(int i=1; i<= size ;i++) environment[i]= participant[i].cycle(blackboard);
   String[] t=environment;     environment= blackboard;  //swap arrays
   return blackboard=t;        //returned behaviours are next cycle's blackboard
 }
 static double rand;     //uses 8 digits of a generated random number
 static double random() { if (rand<0.00000001)  return rand=Math.random();
                           rand= rand*10.0 - Math.floor(rand*10.0);  return rand;
                         }
}
/* Group.random() above replaces Math.random() in an attempt to mitigate some
  "wasted randomicity", but it probably doesn't make any difference.
  After much testing it appears that Math.random() is anything but random, often
  resulting in all-or-nothing decisions like either nobody wanting to talk or to
  give feedback or everybody deciding to talk or give feedback at the same time.
  The problem may lie in the fact that each random call generates a real number
  (0.xxxxx..), and the simulation only uses the first 1 or 2 digits each time.
  So Group.random() encapsulates Math.random() to use all of its random digits.
  Alternative: (new Random()).nextDouble(); (include java.util.Random;)
  This seeds the randomizer each time of use: good when waiting for keypress. */
```

```java
public class model   //vertical output (one cycle per line): with ENTER pressing
{                    //SPACE+ENTER for continuous output timed by the cycleTime
 static Group group;
 static final int maxSize= 7; //maximum size of 'small' groups

 static public void  main(String[] args)
 {
  System.out.println(
    "\nTurn-taking model in a simulation of small group discussion v.22/07/04");
  int size; //first argument is the group size
  if (args.length < 1 || (size= Integer.parseInt(args[0]))< 2 || size> maxSize)
  {
    group= new Group(1); //getting the default attribute values
    System.out.println(
     "\nProvide the group size as the first argument (max. "+ maxSize +").\n"+
     "Then optionally in any order, the cycle-time defining the simulation's\n"+
     "resolution or granularity (-X), the maximum length of any utterances \n"+
     "in the discussion (+X), the frequency of negative feedback (?X), and the"+
     "\nattributes (cXYZ..), where X (Y,Z,..) in all these parameters is a dig"+
     "it\ncorresponding to their one-tenth probabilities (0.X, 0.Y, 0.Z) or se"+
     "conds\nfor the maximum utterance length (+X), and c is the attribute:\n"+
     "\nt - Talkativeness: want to talk at TRPs or when no one is talking?\n"+
     "e - Eagerness: start immediately or wait for the speaker to stop?\n"+
     "f - Feedback: give feedback (uhuh or huh?) when NOT wanting to talk?\n"+
     "c - Confidence: persist in talking simultaneously with others?\n"+
     "a - Attention: time in secs. of perceiving others talking (unused)\n"+
     "v - Verbosity: continue to talk (by starting new utterances)?\n"+
     "p - Pausing: mean length of pauses before starting new utterances\n"+
     "i - Interactivity: pre-TRPs in mid-utterance and trailing-off talk\n"+
     "h - Hesitation: pausing or disfluencies in mid-utterance\n"+
     "s - Selectivity: select the next speaker to talk?\n"+
     "n - Nonverbal: frequency of non-verbal behaviour.\n\n"+
     "If there is more than one digit (cXYZ..), one is for each agent in the\n"+
     "group; if there is only one instead (cX), the same value is for all.\n"+
     "\nFor example:\t\tjava model 4 -5 ?0 t3456 v5 e6\n"+
     "starts a 4-group discussion with a cycle-time of 0.5s, no negative\n"+
     "feedback, agents with talkativeness 0.3, 0.4, 0.5, 0.6 respectively,\n"+
     "and all with verbosity 0.5 and eagerness 0.6.\n\n"+
     "Defaults (when no parameter is provided):\n"+
     "- cycle time (-X): "+Talker.cycleTime+"s\n"+
     "- maximum length of utterances (+X): "+Talker.MaxUtterances+"s\n"+
     "- frequency of negative feedback (?X): "+Talker.NotUnderstand+"\n\n"+
     "Attribute defaults (same value for all participants):\n"+
     "- Talkativeness (tXXXX): "+group.participant[1].TALKATIVENESS+"\n"+
     "- Eagerness (eXXXX): "+group.participant[1].EAGERNESS+"\n"+
     "- Feedback (fXXXX): "+group.participant[1].FEEDBACK+"\n"+
     "- Confidence (cXXXX): "+group.participant[1].CONFIDENCE+"\n"+
     "- Attention (aXXXX): "+group.participant[1].ATTENTION+"\n"+
     "- Verbosity (vXXXX): "+group.participant[1].VERBOSITY+"\n"+
     "- Pausing (pXXXX): "+group.participant[1].TRPAUSING+"s\n"+
     "- Interactivity (iXXXX): "+group.participant[1].INTERACTIVITY+"\n"+
     "- Hesitation (hXXXX): "+group.participant[1].HESITATION+"\n"+
     "- Selectivity (sXXXX): "+group.participant[1].SELECTIVITY+"\n"+
     "- Nonverbal (nXXXX): "+group.participant[1].NONVERBAL+"\n");
    return;
  }
```

```
   group= new Group( size ); //create the group

   for (int i=1;  i< args.length  ;i++)
     if (args[i].charAt(0)=='-')  Talker.cycleTime= (args[i].charAt(1)-'0')/10.0;
     else if (args[i].charAt(0)=='+')
                   Talker.MaxUtterances= Integer.parseInt(args[i].substring(1));
     else if (args[i].charAt(0)=='?')
                              Talker.NotUnderstand= (args[i].charAt(1)-'0')/10.0;
     else if (args[i].length()==2)  for (int k=1;  k<= Group.size  ;k++)
             group.participant[k].attribute( args[i].charAt(0),
                                              (args[i].charAt(1)-'0')/10.0 );
     else for (int k=1;  k<= Group.size && k< args[i].length()  ;k++)
             group.participant[k].attribute( args[i].charAt(0),
                                              (args[i].charAt(k)-'0')/10.0 );

   System.out.print("Press ENTER for next or SPACE+ENTER for nonstop "+
                   "simulation.\n\nTime     ");
   for (int i=1;  i<= size  ;i++)  System.out.print(Talker.Names[i]+"|");
   System.out.print("\n         ");
   for (int i=1;  i<= size  ;i++)  System.out.print("-----");

   int time= 0;            //count the time in the discussion
   boolean timed= false;   //simulation starts with pressing ENTER for each cycle
   long timing= (long)Math.round(Talker.cycleTime* 1000.0);  //timing-> cycleTime
   Thread main= Thread.currentThread();
   while (true)
   {
    if (timed)  try{ main.sleep(timing); } catch(InterruptedException _){return;}
    else try { InputStreamReader x= new InputStreamReader(System.in);
              int key;  if ((key= x.read())==' ') timed=true;
                else if (key!='\n') return;
            }catch (IOException _) {return;}

    String[] blackboard= group.cycle();        //run each cycle of the simulation

    System.out.print(time/10 +"."+ time%10);
    System.out.print(time<100? "     ":time<1000? "    ":"   ");
    for (int i=1;  i<= size  ;i++)  System.out.print(blackboard[i]+ " ");
    if (timed) System.out.print("\n");   time+= (int)10*Talker.cycleTime;
   }                               //clock up the time
  }
}
```

```
[model2.java]

/*****************************************************************************
   Turn-taking model for a simulation of small group discussion
   by Emiliano G. Padilha              previous versions 21/Mar/02, May/03, Feb/04
                                       4th version: begun Jul/04
 *****************************************************************************
Behaviours:  -  , talk, TALK, ta- ,  pTRP, More, AgtX, Any?,  uhuh, huh?

Basic model extended with simultaneous talk resolution practices, but:
      no AgtX/Any?/More(only pTRP), no hesitation, nonverbal behaviours
      (hence, attributes SELECTIVITY, HESITATION, NONVERBAL ain't used yet)

Compile with:  $ javac model2.java       (this file only overrides class Talker)
                                          (model.java must be compiled first)
Run with:       $ java model              (a page with instructions will appear)
 *****************************************************************************/
import java.io.*;

class Talker         //the agents representing conversants in the discussion
{
  final int myself;      //1,2..: simple identification of self
  double TALKATIVENESS, //likelihood of wanting to talk (with no one or someone)
         EAGERNESS,      //likelihood to speak forthwith or wait for a pause
         FEEDBACK,       //likelihood to give feedback (either uhuh or huh?)
         CONFIDENCE,     //likelihood to persist in talking simultaneously
         ATTENTION,      //(unused) mean time to perceive that others are talking
         VERBOSITY,      //likelihood to continue to talk starting new utterances
         TRPAUSING,      //mean length of TRP pauses (after utterances) in secs.
         INTERACTIVITY, //likelihood of mid-utterance TRPs (pTRP, More) and tal_
         HESITATION,     //likelihood of hesitations or disfluencies in mid-talk
         SELECTIVITY,    //likelihood of selecting a next speaker at a TRP
         NONVERBAL;      //(unused) likelihood of GESTures, POSTure shifts, NODS

 Talker(int id)  //initialize the agents with default values
 {
   myself = id;            FEEDBACK  = 0.3;  VERBOSITY  = 0.5;  INTERACTIVITY= 0.5;
   TALKATIVENESS= 0.2;  CONFIDENCE= 0.6;  TRPAUSING  = 0.5;  HESITATION   = 0.3;
   EAGERNESS    = 0.4;  ATTENTION = 0.3;  SELECTIVITY= 0.1;  NONVERBAL    = 0.3;
 }
 void attribute(char code, double value)
 {                                       //command-line setting of the attributes
   switch(code)
   {
    case't': TALKATIVENESS= value; break;  case'v': VERBOSITY    = value; break;
    case'e': EAGERNESS    = value; break;  case'p': TRPAUSING    = value; break;
    case'f': FEEDBACK     = value; break;  case'i': INTERACTIVITY= value; break;
    case'c': CONFIDENCE   = value; break;  case'h': HESITATION   = value; break;
    case'a': ATTENTION    = value; break;  case's': SELECTIVITY  = value; break;
                                           case'n': NONVERBAL    = value; break;
    default: System.out.println("*** Wrong attribute code: "+code+"\n");
   }
   return;
 }
             //agent names, for printing and SELECT(X)
static String Names[]={"????","AgtA","AgtB","AgtC","AgtD","AgtE","AgtF","AgtG"};
```

```java
 //parameters global to the simulation, instead of individual to each agent
 static double cycleTime = 0.2;    //clock-cycle is 0.2s (5 cycles per second)
 static double MaxUtterances = 4;  //maximum length of any utterance (4 seconds)
 static double NotUnderstand = 0.1;//frequency of negative feedback 'huh?' (10%)

 //agents' memory: states/timers (make the model 'almost' finite-state?)
 int speaker, length, tail, pause, midTRPs, feedback, huhed, huhfor, simultalk;
 boolean wannaTalk;

String  cycle(String was[]) //cycle activation: the agents' program of operation
{
 int talked=0;     //check input at every cycle to see who's talking (besides me)
 for (int i=1; i< was.length ;i++) if (i!= myself) //preferably any NEW speakers
 {
   if (isTalk(was[i]) && (talked==0 || talked==speaker ||
                                       i!=speaker && Test(0.5)))  talked= i;
   if (was[i]=="huh?") huhed= i; //any negative feedback?
 }
 if (huhed> 0 && was[huhed]!="huh?") //respond when the 'huh?' vocalization ends
    if (huhfor==myself)
       return was[myself].startsWith("tal")? " - ":StartTalk();  else huhed= 0;
 if (was[speaker]=="pTRP") huhfor= speaker;  //to whom exactly is the 'huh?' for

 if (speaker==myself) //I'm the speaker
 {
   if (--length> 0) //utterance: talk talk talk...
   {
     int ShortLength= (int) Math.round(1.0/cycleTime); //'short' utterances: <1s
     if (talked> 0 && was[talked]!="tal_")  //simultaneous talk
     {
       ++simultalk; //count simultaneous talk           //(below) if I hesitated
       if (was[myself]=="ta- " && (was[talked]=="TALK" || !Test(CONFIDENCE)))
          { wannaTalk= Test(TALKATIVENESS);   speaker= talked;  return " .. "; }
       if (was[myself]=="TALK")
          return  was[talked]!="TALK" || Test(CONFIDENCE)? "TALK":"ta- ";
       if (length> ShortLength)
          if (!Test(CONFIDENCE)) { length++;  return "ta- "; } //do a hitch
          else if (!Test(CONFIDENCE)) return "TALK";          //or step-up talk
     }                          //or else continue as normal (the strongest stance)
     if (talked==0)
     {  if (simultalk> ShortLength) //repeat talk if simultaneity was long (>1s)
          { simultalk= 0;  length+= ShortLength;  return "ta/ "; }
        simultalk= 0;
     }
     if (length% midTRPs==0 && Test(INTERACTIVITY)) return "pTRP";
     return "talk";                                  //mid-utterance pTRPs
   }
   if (length==0) //utterance ending: decide on continuing, signal pre-TRP
   {              //and add a 'tail' of trailing-off talk between 0.2-0.6s
     wannaTalk= Test(VERBOSITY);   tail= -1-(Test(INTERACTIVITY)?around(0.4):0);
     pause= tail-around(TRPAUSING);  return "pTRP";
   }
   if (length> tail) return "tal_";         //finish the utterance, then pause
   if (length> pause && talked==0) return " .  "; //"different" silence: a pause
 }
```

```
if (--feedback> 0) return was[myself];  //say the rest of the feedback vocaliz.
if (feedback==0 && was[myself].startsWith(" "))return GiveFeedback(); //delay'd

if (talked==0 && wannaTalk || speaker==0 && Test(TALKATIVENESS))
   return StartTalk();      //anyone who wannaTalk goes first, then 2nd starters
speaker= talked;            //update speaker
if (was[speaker]=="pTRP")
{
  wannaTalk= Test(TALKATIVENESS); //decide whether to talk (forget previous), &
  if (wannaTalk && Test(EAGERNESS)) return StartTalk();  //whether to start now
  if (Test(FEEDBACK)) if (wannaTalk || Test(FEEDBACK))
                         return GiveFeedback();  else feedback= around(0.5);
}                                                   //delay feedback ~~ 0.2-0.8s
return " - ";
}
                       //for a real normal distribution: Math.nextGaussian()
int around(double mean){ return (int)Math.round(
                              (mean*Group.random()+mean/2.0) /cycleTime); }

boolean  Test(double attr){ return Group.random()< attr; } //problst. decisions
boolean isTalk(String did){ return did.startsWith("ta") || did=="pTRP"
                                                        || did=="TALK"; }

String StartTalk()  //start utterance immediately, set length, midTRP intervals
{
  length= 1+(int) Math.round(Group.random()*MaxUtterances/cycleTime); //random#
  midTRPs= around(2.0);    //possible mid-utterance pre-TRPs roughly at each 2s
  speaker= myself;         wannaTalk= false;      huhed=simultalk= 0;
  return "talk";
}
String GiveFeedback() { feedback= around(0.4);    //feedback length ~~ 0.2-0.6s
                     return Test(NotUnderstand)? "huh?":"uhuh"; }
}
```

```
[model3.java]

/****************************************************************************
   Turn-taking model for a simulation of small group discussion
   by Emiliano G. Padilha         previous versions 21/Mar/02, May/03, Feb/04
                                          4th version: begun Jul/04
 ****************************************************************************
 Behaviours:  -  , talk, TALK, ta- ,  pTRP, More, AgtX, Any?,  uhuh, huh?

 Model extended with simultaneous talk practices and sorts of utterances
         (AgtX/Any?/More besides pTRP), but no hesitation, nonverbal behaviours
         (hence, attributes HESITATION, NONVERBAL ain't used yet)

 Compile with:  $ javac model3.java (this file only overrides class Talker)
                                        (model.java must be compiled first)
 Run with:       $ java model              (a page with instructions will appear)
 ****************************************************************************/
import java.io.*;

class Talker          //the agents representing conversants in the discussion
{
  final int myself;      //1,2..: simple identification of self
  double TALKATIVENESS, //likelihood of wanting to talk (with no one or someone)
         EAGERNESS,       //likelihood to speak forthwith or wait for a pause
         FEEDBACK,        //likelihood to give feedback (either uhuh or huh?)
         CONFIDENCE,      //likelihood to persist in talking simultaneously
         ATTENTION,       //(unused) mean time to perceive that others are talking
         VERBOSITY,       //likelihood to continue to talk starting new utterances
         TRPAUSING,       //mean length of TRP pauses (after utterances) in secs.
         INTERACTIVITY, //likelihood of mid-utterance TRPs (pTRP, More) and tal_
         HESITATION,      //likelihood of hesitations or disfluencies in mid-talk
         SELECTIVITY,     //likelihood of selecting a next speaker at a TRP
         NONVERBAL;       //(unused) likelihood of GESTures, POSTure shifts, NODS

 Talker(int id)  //initialize the agents with default values
 {
   myself = id;           FEEDBACK  = 0.3;  VERBOSITY  = 0.5;  INTERACTIVITY= 0.5;
   TALKATIVENESS= 0.2;  CONFIDENCE= 0.6;  TRPAUSING  = 0.5;  HESITATION   = 0.3;
   EAGERNESS    = 0.4;  ATTENTION = 0.3;  SELECTIVITY= 0.1;  NONVERBAL    = 0.3;
 }
 void attribute(char code, double value)
 {                                    //command-line setting of the attributes
   switch(code)
   {
    case't': TALKATIVENESS= value; break;  case'v': VERBOSITY    = value; break;
    case'e': EAGERNESS    = value; break;  case'p': TRPAUSING    = value; break;
    case'f': FEEDBACK     = value; break;  case'i': INTERACTIVITY= value; break;
    case'c': CONFIDENCE   = value; break;  case'h': HESITATION   = value; break;
    case'a': ATTENTION    = value; break;  case's': SELECTIVITY  = value; break;
                                           case'n': NONVERBAL    = value; break;
    default: System.out.println("*** Wrong attribute code: "+code+"\n");
   }
   return;
 }
             //agent names, for printing and SELECT(X)
static String Names[]={"????","AgtA","AgtB","AgtC","AgtD","AgtE","AgtF","AgtG"};
```

```
//parameters global to the simulation, instead of individual to each agent
static double cycleTime = 0.2;     //clock-cycle is 0.2s (5 cycles per second)
static double MaxUtterances = 4;  //maximum length of any utterance (4 seconds)
static double NotUnderstand = 0.1;//frequency of negative feedback 'huh?' (10%)
static double AskAnybody = 0.1;  //frequency of Any? utterances (Has anybody..)

//agents' memory: states/timers (make the model 'almost' finite-state?)
int speaker, length, tail, pause, midTRPs, feedback, huhed, huhfor, simultalk;
boolean wannaTalk;  String  Sort;

String  cycle(String was[]) //cycle activation: the agents' program of operation
{
 int talked=0;     //check input at every cycle to see who's talking (besides me)
 for (int i=1; i< was.length ;i++) if (i!= myself) //preferably any NEW speakers
 {
   if (isTalk(was[i]) && (talked==0 || talked==speaker ||
                                     i!=speaker && Test(0.5)))  talked= i;
   if (was[i]=="huh?") huhed= i; //any negative feedback?
 }
 if (huhed> 0 && was[huhed]!="huh?") //respond when the 'huh?' vocalization ends
    if (huhfor==myself)
       return was[myself].startsWith("tal")? " - ":StartTalk();  else huhed= 0;
 if (preTRP(was[speaker])) huhfor= speaker;  //to whom exactly is the 'huh?' for

 if (speaker==myself) //I'm the speaker
 {
   if (--length> 0) //utterance: talk talk talk...
   {
     int ShortLength= (int) Math.round(1.0/cycleTime); //'short' utterances: <1s
     if (talked> 0 && was[talked]!="tal_")  //simultaneous talk
     {
       ++simultalk; //count simultaneous talk          //(below) if I hesitated
       if (was[myself]=="ta- " && (was[talked]=="TALK" || !Test(CONFIDENCE)))
          { wannaTalk= Test(TALKATIVENESS);   speaker= talked;  return " .. "; }
       if (was[myself]=="TALK")
          return  was[talked]!="TALK" || Test(CONFIDENCE)? "TALK":"ta- ";
       if (length> ShortLength)
          if (!Test(CONFIDENCE)) { length++;  return "ta- "; } //do a hitch
          else if (!Test(CONFIDENCE)) return "TALK";          //or step-up talk
     }                        //or else continue as normal (the strongest stance)
     if (talked==0)
     {  if (simultalk> ShortLength) //repeat talk if simultaneity was long (>1s)
           { simultalk= 0;  length+= ShortLength;  return "ta/ "; }
        simultalk= 0;
     }
     if (length% midTRPs==0 && Test(INTERACTIVITY))
        if (Sort!="pTRP" && Sort!="More" && Test(0.2)) return Sort;
        else return "pTRP";                    //mid-utterance pTRPs or AgtX|Any?
     return "talk";
   }
   if (length==0) //utterance ending: decide on continuing, signal pre-TRP
   {              //and add a 'tail' of trailing-off talk between 0.2-0.6s
     tail= -1-( Test(INTERACTIVITY)? around(0.4):0 );
     pause= tail-around(TRPAUSING);
     wannaTalk= Sort=="pTRP"? Test(VERBOSITY): Sort=="More"? true: false;
     return Sort; //AgtX,Any?,More,pTRP
```

```
   }
   if (length> tail) return "tal_";            //finish the utterance, then pause
   if (length> pause && talked==0) return " . "; //"different" silence: a pause
 }

 if (--feedback> 0) return was[myself];  //say the rest of the feedback vocaliz.
 if (feedback==0 && was[myself].startsWith(" "))return GiveFeedback(); //delay'd

 if (talked==0 && wannaTalk || speaker==0 && Test(TALKATIVENESS))
    return StartTalk();      //anyone who wannaTalk goes first, then 2nd starters
 if (talked!=0 || Sort=="pTRP" || Sort=="Any?")  speaker= talked;
 if (preTRP(was[speaker]))                     //^^not when the turn is not free
 {
   Sort= was[speaker]; //keep last pre-TRP (is turn free for ttaking: pTRP|Any?)
   if (Sort=="pTRP") wannaTalk= Test(TALKATIVENESS); else //forget prior pre-TRP
   if (Sort=="More") wannaTalk= Test(TALKATIVENESS*EAGERNESS); else //discourage
   if (Sort=="Any?") wannaTalk= Test(TALKATIVENESS+EAGERNESS); else //encourage
                     wannaTalk= Sort==Names[myself];    //next-speaker selected
   if (wannaTalk && Test(EAGERNESS)) return StartTalk();  //whether to start now
   if (Sort=="pTRP" || Sort=="More")      //feedback, but not after AgtX or Any?
      if (Test(FEEDBACK)) if (wannaTalk || Test(FEEDBACK))
                             return GiveFeedback();  else feedback= around(0.5);
 }                                                   //delay feedback ~~ 0.2-0.8s
 return " - ";
}
                        //for a real normal distribution: Math.nextGaussian()
 int around(double mean){ return (int)Math.round(
                                   (mean*Group.random()+mean/2.0) /cycleTime); }

 boolean  Test(double attr){ return Group.random()< attr; } //problst. decisions
 boolean isTalk(String did){ return did.startsWith("ta") || did=="TALK"
                                                        || preTRP(did); }
 boolean preTRP(String did){ for (int i=1;  i<= Group.size  ;i++)
                               if (did==Names[i]) return true;
                             return did=="pTRP"|| did=="More"|| did=="Any?"; }

 String StartTalk()  //start utterance immediately, set length, midTRP intervals
 {
   if(Test(SELECTIVITY)){ int x; do x=(int)Math.ceil(Group.random()*Group.size);
                    while (x==myself || x> Group.size);  Sort= Names[x]; }
   else Sort= Group.size> 2 && Test(AskAnybody)?"Any?":
                          Test(VERBOSITY*0.5)?"More": "pTRP";
   length= 1+(int) Math.round(Group.random()*MaxUtterances/cycleTime); //random#
   midTRPs= around(2.0);    //possible mid-utterance pre-TRPs roughly at each 2s
   speaker= myself;         wannaTalk= false;       huhed=simultalk= 0;
   return "talk";
 }
 String GiveFeedback() { feedback= around(0.4);   //feedback length ~~ 0.2-0.6s
                    return Test(NotUnderstand)? "huh?":"uhuh"; }
}
```

```
[model4.java]

/*****************************************************************************
  Turn-taking model for a simulation of small group discussion
  by Emiliano G. Padilha           previous versions 21/Mar/02, May/03, Feb/04
                                    4th version: begun Jul/04
*****************************************************************************
Behaviours:  - , talk, TALK, ta- , pTRP, More, AgtX, Any?,  uhuh, huh?

Model extended with simultaneous talk practices, sorts of utterances and
     with hesitations; but no nonverbal behaviours (hence, NONVERBAL not used)

Compile with:  $ javac model4.java (this file only overrides class Talker)
                                    (model.java must be compiled first)
Run with:      $ java model         (a page with instructions will appear)
*****************************************************************************/
import java.io.*;

class Talker        //the agents representing conversants in the discussion
{
  final int myself;    //1,2..: simple identification of self
  double TALKATIVENESS, //likelihood of wanting to talk (with no one or someone)
         EAGERNESS,      //likelihood to speak forthwith or wait for a pause
         FEEDBACK,       //likelihood to give feedback (either uhuh or huh?)
         CONFIDENCE,     //likelihood to persist in talking simultaneously
         ATTENTION,      //(unused) mean time to perceive that others are talking
         VERBOSITY,      //likelihood to continue to talk starting new utterances
         TRPAUSING,      //mean length of TRP pauses (after utterances) in secs.
         INTERACTIVITY, //likelihood of mid-utterance TRPs (pTRP, More) and tal_
         HESITATION,     //likelihood of hesitations or disfluencies in mid-talk
         SELECTIVITY,    //likelihood of selecting a next speaker at a TRP
         NONVERBAL;      //(unused) likelihood of GESTures, POSTure shifts, NODS

 Talker(int id)  //initialize the agents with default values
 {
   myself = id;           FEEDBACK  = 0.3;  VERBOSITY  = 0.5;  INTERACTIVITY= 0.5;
   TALKATIVENESS= 0.2;  CONFIDENCE= 0.6;  TRPAUSING  = 0.5;  HESITATION   = 0.3;
   EAGERNESS    = 0.4;  ATTENTION = 0.3;  SELECTIVITY= 0.1;  NONVERBAL    = 0.3;
 }
 void attribute(char code, double value)
 {                                        //command-line setting of the attributes
   switch(code)
   {
    case't': TALKATIVENESS= value; break;  case'v': VERBOSITY    = value; break;
    case'e': EAGERNESS    = value; break;  case'p': TRPAUSING    = value; break;
    case'f': FEEDBACK     = value; break;  case'i': INTERACTIVITY= value; break;
    case'c': CONFIDENCE   = value; break;  case'h': HESITATION   = value; break;
    case'a': ATTENTION    = value; break;  case's': SELECTIVITY  = value; break;
                                           case'n': NONVERBAL    = value; break;
    default: System.out.println("*** Wrong attribute code: "+code+"\n");
   }
   return;
 }
             //agent names, for printing and SELECT(X)
static String Names[]={"????","AgtA","AgtB","AgtC","AgtD","AgtE","AgtF","AgtG"};
```

```
 //parameters global to the simulation, instead of individual to each agent
 static double cycleTime = 0.2;     //clock-cycle is 0.2s (5 cycles per second)
 static double MaxUtterances = 4;  //maximum length of any utterance (4 seconds)
 static double NotUnderstand = 0.1;//frequency of negative feedback 'huh?' (10%)
 static double AskAnybody = 0.1;  //frequency of Any? utterances (Has anybody..)

 //agents' memory: states/timers (make the model 'almost' finite-state?)
 int speaker, length, tail, pause, midTRPs, feedback, huhed, huhfor, simultalk;
 boolean wannaTalk;  String  Sort;  int hesitate;


String  cycle(String was[]) //cycle activation: the agents' program of operation
{
 int talked=0;     //check input at every cycle to see who's talking (besides me)
 for (int i=1; i< was.length ;i++) if (i!= myself) //preferably any NEW speakers
 {
   if (isTalk(was[i]) && (talked==0 || talked==speaker ||
                                      i!=speaker && Test(0.5)))  talked= i;
   if (was[i]=="huh?") huhed= i; //any negative feedback?
 }
 if (huhed> 0 && was[huhed]!="huh?") //respond when the 'huh?' vocalization ends
    if (huhfor==myself)
       return was[myself].startsWith("tal")? " - ":StartTalk();  else huhed= 0;
 if (preTRP(was[speaker])) huhfor= speaker;  //to whom exactly is the 'huh?' for

 int ShortLength= (int) Math.round(1.0/cycleTime); //'short' utterances: <1s
 if (speaker==myself) //I'm the speaker
 {
   if (--length> 0) //utterance: talk talk talk...
   {
     if (talked> 0 && was[talked]!="tal_")  //simultaneous talk
     {
       ++simultalk; //count simultaneous talk          //(below) if I hesitated
       if (Hesita(was[myself]) && (was[talked]=="TALK" || !Test(CONFIDENCE)))
          { wannaTalk= Test(TALKATIVENESS);   speaker= talked;  return " .. "; }
       if (was[myself]=="TALK")
          return  was[talked]!="TALK" || Test(CONFIDENCE)? "TALK":"ta- ";
       if (length> ShortLength)
          if (!Test(CONFIDENCE)) { length++;  return "ta- "; } //do a hitch
          else if (!Test(CONFIDENCE)) return "TALK";          //or step-up talk
     }                          //or else continue as normal (the strongest stance)
     if (talked==0)
     {  if (simultalk> ShortLength) //repeat talk if simultaneity was long (>1s)
           { simultalk= 0;  length+= ShortLength;  return "ta/ "; }
        simultalk= 0;
     }
     if (length% midTRPs==0 && Test(INTERACTIVITY))
        if (Sort!="pTRP" && Sort!="More" && Test(0.2)) return Sort;
        else return "pTRP";                   //mid-utterance pTRPs or AgtX|Any?
     if (--hesitate> 0 || length> ShortLength && Test(HESITATION) && ++length>0)
        return Test(HESITATION)? "ta- ":" -  "; //hesitation: disfluency,silence
     return "talk";
   }
```

```
    if (length==0) //utterance ending: decide on continuing, signal pre-TRP
    {               //and add a 'tail' of trailing-off talk between 0.2-0.6s
      tail= -1-( Test(INTERACTIVITY)? around(0.4):0 );
      pause= tail-around(TRPAUSING);
      wannaTalk= Sort=="pTRP"? Test(VERBOSITY): Sort=="More"? true: false;
      return Sort; //AgtX,Any?,More,pTRP
    }
    if (length> tail) return "tal_";          //finish the utterance, then pause
    if (length> pause && talked==0) return " .  "; //"different" silence: a pause
  }
  if (--feedback> 0) return was[myself];  //say the rest of the feedback vocaliz.
  if (feedback==0 && was[myself].startsWith(" "))return GiveFeedback(); //delay'd

  if (talked==0 && wannaTalk && Sort!="" || speaker==0 && Test(TALKATIVENESS))
      return StartTalk();       //anyone who wannaTalk goes first, then 2nd starters
                              //(below)don't change speaker if the turn isn't free
  if (talked!=0 || Sort=="pTRP" || Sort=="Any?")  speaker= talked;
  else for (int i=1;  i<= Group.size  ;i++)  if (Sort==Names[i]) speaker= i;

  if (was[speaker]=="talk") { Sort= ""; hesitate= 1; }//no turn-taking after talk
  else if (Hesita(was[speaker]) && ++hesitate>= ShortLength &&    //interrupt the
          Test(TALKATIVENESS*EAGERNESS))  return StartTalk();        //hesitation

  if (preTRP(was[speaker]))
  {
    Sort= was[speaker]; //keep last pre-TRP (is turn free for ttaking: pTRP|Any?)
    if (Sort=="pTRP") wannaTalk= Test(TALKATIVENESS); else //forget prior pre-TRP
    if (Sort=="More") wannaTalk= Test(TALKATIVENESS*EAGERNESS); else //discourage
    if (Sort=="Any?") wannaTalk= Test(TALKATIVENESS+EAGERNESS); else //encourage
                      wannaTalk= Sort==Names[myself];      //next-speaker selected
    if (wannaTalk && Test(EAGERNESS)) return StartTalk();  //whether to start now
    if (Sort=="pTRP" || Sort=="More")      //feedback, but not after AgtX or Any?
        if (Test(FEEDBACK)) if (wannaTalk || Test(FEEDBACK))
                            return GiveFeedback();  else feedback= around(0.5);
  }                                                //delay feedback ~~ 0.2-0.8s
  return " - ";
}
                        //for a real normal distribution: Math.nextGaussian()
int around(double mean){ return (int)Math.round(
                              (mean*Group.random()+mean/2.0) /cycleTime); }
boolean  Test(double attr){ return Group.random()< attr; } //problst. decisions
boolean isTalk(String did){ return did.startsWith("ta") || did=="TALK"
                                                || preTRP(did); }
boolean preTRP(String did){ for (int i=1;  i<= Group.size  ;i++)
                              if (did==Names[i]) return true;
                            return did=="pTRP"|| did=="More"|| did=="Any?"; }
boolean Hesita(String did){ return did=="ta- " || did.startsWith(" "); }

String StartTalk()  //start utterance immediately, set length, midTRP intervals
{
  //if I have the floor I can hesitate inbetween utterances
  boolean IHaveTheFloor= Sort==Names[myself] || speaker==myself;
  if(Test(SELECTIVITY)){ int x; do x=(int)Math.ceil(Group.random()*Group.size);
                    while (x==myself || x> Group.size);  Sort= Names[x]; }
  else Sort= Group.size> 2 && Test(AskAnybody)?"Any?":
                          Test(VERBOSITY*0.5)?"More": "pTRP";
```

```
    length= 1+(int) Math.round(Group.random()*MaxUtterances/cycleTime); //random#
   midTRPs= around(2.0);    //possible mid-utterance pre-TRPs roughly at each 2s
   speaker= myself;       wannaTalk= false;       huhed=simultalk=hesitate= 0;
   if (IHaveTheFloor && Test(HESITATION))
      { hesitate= around(0.5);  ++length;  return "ta- "; }//hesitate ~~0.2-0.7s
   return "talk";                    //have a hesitating pre-starter ("well...")?
 }
 String GiveFeedback() { feedback= around(0.4);    //feedback length ~~ 0.2-0.6s
                       return Test(NotUnderstand)? "huh?":"uhuh"; }
}




[hmodel.java]      (Talker & Group classes from model.java must be pasted here)

import java.io.*;
public class hmodel //with horizontal output (one agent per line)
{
 static Group group;
 static final int maxSize= 7; // maximum size of 'small' groups
 static public void  main(String[] args)
 {
  System.out.println(
    "Turn-taking model in a simulation of small group discussion v.22/07/04\n");
  int size; //first argument is the group size

  if (args.length < 1 || (size= Integer.parseInt(args[0]))< 2 || size> maxSize)
  {
    group= new Group(1); //getting the default attribute values
    System.out.println(
    "\nProvide the group size as the first argument (max. "+ maxSize +").\n"+
    "Then optionally in any order, the cycle-time defining the simulation's\n"+
    "resolution or granularity (-X), the maximum length of any utterances \n"+
    "in the discussion (+X), the frequency of negative feedback (?X), and the"+
    "\nattributes (cXYZ..), where X (Y,Z,..) in all these parameters is a dig"+
    "it\ncorresponding to their one-tenth probabilities (0.X, 0.Y, 0.Z) or se"+
    "conds\nfor the maximum utterance length (+X), and c is the attribute:\n"+
    "\nt - Talkativeness: want to talk at TRPs or when no one is talking?\n"+
    "e - Eagerness: start immediately or wait for the speaker to stop?\n"+
    "f - Feedback: give feedback (uhuh or huh?) when NOT wanting to talk?\n"+
    "c - Confidence: persist in talking simultaneously with others?\n"+
    "a - Attention: time in secs. of perceiving others talking (unused)\n"+
    "v - Verbosity: continue to talk (by starting new utterances)?\n"+
    "p - Pausing: mean length of pauses before starting new utterances\n"+
    "i - Interactivity: pre-TRPs in mid-utterance and trailing-off talk\n"+
    "h - Hesitation: pausing or disfluencies in mid-utterance\n"+
    "s - Selectivity: select the next speaker to talk?\n"+
    "n - Nonverbal: frequency of non-verbal behaviour.\n\n"+
    "If there is more than one digit (cXYZ..), one is for each agent in the\n"+
    "group; if there is only one instead (cX), the same value is for all.\n"+
    "\nFor example:\t\tjava model 4 -5 ?0 t3456 v5 e6\n"+
    "starts a 4-group discussion with a cycle-time of 0.5s, no negative\n"+
    "feedback, agents with talkativeness 0.3, 0.4, 0.5, 0.6 respectively,\n"+
    "and all with verbosity 0.5 and eagerness 0.6.\n\n"+
```

```
      "Defaults (when no parameter is provided):\n"+
      "- cycle time (-X): "+Talker.cycleTime+"s\n"+
      "- maximum length of utterances (+X): "+Talker.MaxUtterances+"s\n"+
      "- frequency of negative feedback (?X): "+Talker.NotUnderstand+"\n\n"+
      "Attribute defaults (same value for all participants):\n"+
      "- Talkativeness (tXXXX): "+group.participant[1].TALKATIVENESS+"\n"+
      "- Eagerness (eXXXX): "+group.participant[1].EAGERNESS+"\n"+
      "- Feedback (fXXXX): "+group.participant[1].FEEDBACK+"\n"+
      "- Confidence (cXXXX): "+group.participant[1].CONFIDENCE+"\n"+
      "- Attention (aXXXX): "+group.participant[1].ATTENTION+"\n"+
      "- Verbosity (vXXXX): "+group.participant[1].VERBOSITY+"\n"+
      "- Pausing (pXXXX): "+group.participant[1].TRPAUSING+"s\n"+
      "- Interactivity (iXXXX): "+group.participant[1].INTERACTIVITY+"\n"+
      "- Hesitation (hXXXX): "+group.participant[1].HESITATION+"\n"+
      "- Selectivity (sXXXX): "+group.participant[1].SELECTIVITY+"\n"+
      "- Nonverbal (nXXXX): "+group.participant[1].NONVERBAL+"\n");
      return;
  }
  group= new Group( size ); //create the group

  for (int i=1; i< args.length ;i++)
    if (args[i].charAt(0)=='-') Talker.cycleTime= (args[i].charAt(1)-'0')/10.0;
    else if (args[i].charAt(0)=='+')
                              Talker.MaxUtterances= (args[i].charAt(1)-'0');
    else if (args[i].charAt(0)=='?')
                          Talker.NotUnderstand= (args[i].charAt(1)-'0')/10.0;
    else if (args[i].length()==2) for (int k=1; k<= Group.size ;k++)
          group.participant[k].attribute( args[i].charAt(0),
                                          (args[i].charAt(1)-'0')/10.0 );
    else for (int k=1; k<= Group.size && k< args[i].length() ;k++)
          group.participant[k].attribute( args[i].charAt(0),
                                          (args[i].charAt(k)-'0')/10.0 );
  String output[]= new String[size+1];
  int width= 80;   //character width of the output
  int time= 0;     //time of the discussion in multiples of 10 (12 => 1.2s)
  Thread main= Thread.currentThread();
  while (true)
  {
    output[0]= "____|"
    for (int i=1; i<= size ;i++) output[i]= Talker.Names[i]+"|";
    while (output[1].length()< width- 5) //output width minus 1 behaviour width
    {
      String[] blackboard= group.cycle(); //run each cycle of the simulation

      for (int i=1; i<= size ;i++) output[i]+= blackboard[i]+ " ";
      output[0]+= time%10 ==0 ? (time/10+"s___").substring(0,4)+"|":"____|";
      time+= (int)10*Talker.cycleTime; //clock up the time
    }
    for (int i=0; i< output.length ;i++) System.out.println(output[i]);
    try { InputStreamReader x= new InputStreamReader(System.in);
  int key;  if ((key= x.read())!='\n') return;
      }catch (IOException _) {return;}
  }
 }
}
```

Appendix B

Evaluation program

There follows below the evaluation program with the `Talker` class omitted for redundancy, since it is the same as in the first model. Because Java requires it for compilation with the main class (here, `mtest` instead of `model`), the program in this listing must have the `Talker` class from the file `model.java` pasted at the top.

The `Group` class now contains the routines `Measures()` and `showMeasures()`: the first does the evaluation job, the second displays the results at the end of the evaluation suite: 150 simulations of 300 seconds each. The only addition to the original `Group.cycle()` is then a call to `Measures()` at the end of each cycle of simulation so that the *measures* can be identified and counted. To do this, that routine receives the blackboard with the recent agent actions for inspection and returns the text information that in the routine is appended to a variable `test` (originally only for test purposes as the name says). This information is returned in the first (unused) slot of the 'blackboard' array, which is output on the screen by the `main` routine (`mtest` class). This routine runs the evaluation suite whereof all measures are accumulated, averaged and presented at the end. Cycle-by-cycle agent actions together with the information returned from the `Measure()` routine can also be shown, which may be useful in understanding how the various measures are counted.

The listings here and in the previous appendix 1 are just for easy browsing. If you would like to have these files, instead of typing them all, you will probably be able to download them from my site about this work, wherever it may be in the future (as yet undefined at the time of writing).

```
[mtest.java]    (the Talker class from model.java also goes here)

class Group
{
 static int size;        // group size
 Talker participant[];  // the participants of the group discussion
 String environment[],  // the behaviours of the current cycle (next blackboard)
        blackboard[];   // and behaviours of the previous one

 Group(int groupsize)
 {
   size = groupsize;
   participant= new Talker[size+1];    //indexing is [1..size], not [0..size-1]
   environment= new String[size+1];
   blackboard = new String[size+1];     //initialize the blackboards
   for (int i=1;  i<= size  ;i++)  participant[i]= new Talker(i);
   for (int i=0;  i<= size  ;i++)  environment[i]= blackboard[i]= " -  ";
 }
 String[] cycle() //run one cycle of the simulated discussion
 {
   for(int i=1; i<= size ;i++) environment[i]= participant[i].cycle(blackboard);
   String[] t=environment;     environment= blackboard;
   Measures(t); t[0]= test;  //count the measures and return actions
   return blackboard=t;       //returned behaviours are next cycle's blackboard
 }
 static Random seed;
 static double rand;    //uses 8 digits of a generated random number
 static double random() { if (rand<0.00000001)  return rand=seed.nextDouble();
                          rand= rand*10.0- Math.floor(rand*10.0); return rand; }

/*Measures counted in the aggregate of simulations:
  total_singletalk: total time of single talk,
  total_feedback: amount of feedback on its own and backchannels (with talk),
   total_backchannel: only the time of feedback in the background of talk,
  total_notalk: time of all silence including pauses and feedback (uhuh/huh?),
   total_silence: time of all silence without any uhuh/huh?,
  silentrs: occurrences of silent speaker transitions ('gaps' and 'lapses'),
   silentr_sum: total time of silent transitions (=total_notalk - pauses),
   silentr_shortest/longest: their shortest and longest occurrences,
  overlaps/latches: occurrences of overlapping and latched speaker transitions,
  total_simtalk: total time of two or more talking (talk,ta- ,pTRP,tal_),
  simtalks: occurrences of simultaneous talks/starts, except overlaps (tal_),
   simtalk_sum: total time of simultaneous talks/starts but not overlaps (tal_),
   simtalk_shortest/longest: their shortest and longest occurrences,
  longsimtalks: occurrences of long simultaneous talks (of 1 or more seconds),
   longsimtalk_many: total sum of number of speakers in long simult.talks,
  singlestarts: occurrences of single starts of talk (not quasi-simstarts),
  simstarts: occurrences of simultaneous or quasi- starts (off by a cycle),
   simstarts_many: total sum of number of speakers in simstarts (for the mean),
  middlestarts: occurrences of starts in the middle of someone's utterance,
  falsestarts: occurrences of incomplete short talks (of less than 1 second),
   'botcheds': occurrences of "collective" false-starts, all stopping,
   incompletes: occurrences of "longer" false-starts: incomplete utterances,
  utterances/ag_utterances[]: number of complete utterances (total/per agent),
  continueds/ag_continueds[]: number of continuing utterances (total/per agent).
*/
```

```
 int utterances, continueds, total_silence, total_singletalk, total_simtalk,
     silentrs, silentr_shortest,silentr_longest, silentr_sum, silentr_count,
     simtalks, simtalk_shortest,simtalk_longest, simtalk_sum, simtalk_count,
     singlestarts, simstarts, simstart_many, longsimtalks, longsimtalk_many,
     falsestarts, botcheds, incompletes, middlestarts, overlaps, latches,  anys,
     total_notalk, total_feedback, total_backchannel, prior,false_count,  mores,
     ag_utterances[]={0,0,0,0,0,0,0,0}, ag_continueds[]={0,0,0,0,0,0,0,0}, sels;
 String talks="",TRPs="",last="",test;   int speaker,length[]={0,0,0,0,0,0,0};

 void Measures(String was[])
 {
    //threshold of distinguishing false-starts from incomplete utterances (1s)
   int Limit= (int)Math.round(1.0/Talker.cycleTime);
   String lastTalks=talks, lastTRPs=TRPs;   talks=TRPs=test="";
   int talked=0, stops=0;  boolean silence=true, allsilence=true, fback=false;
   for (int i=1; i< was.length ;i++)  if (isTalk(was[i]))
      {                                              //i must be 1 digit only!
        length[i]++; if (talked==0 || talked==speaker) talked= i;
        talks+=i;    if (preTRP(was[i])) { TRPs+= i; countPreTRPs(was[i]); }
      }else{
        if (! was[i].startsWith(" ")) { allsilence= false;  fback=true; }
        if (was[i]==" .. "){ stops++;
           if (length[i]>= Limit) { incompletes++; test+= "incomplete "; }
              else { false_count++; falsestarts++; test+= "false-start "; } }
        if (i!=speaker || was[i]!=" - " || last!="talk" && last!="ta- ")
           length[i]= 0;  else allsilence= silence= false; //exclude hesitations
      }
   if (fback){ total_feedback++;  if (talked> 0) total_backchannel++; }
   int ntalks=talks.length(), lastNtalks=lastTalks.length();
   test= "   "+talked+"("+ntalks+","+TRPs+") "+test;
   if (lastNtalks> 1 && ntalks<= 1 && simtalk_count> 0) //exclude overlaps
   {
     if (ntalks==0 && stops==lastNtalks && length[talked]< Limit)
       { botcheds++;  falsestarts-=false_count;  //"collective" false-start
         test+= "botched(falses-"+false_count+") "; }
     if (simtalk_shortest==0 || simtalk_count< simtalk_shortest)
       { test+= "simtalk_shortest ";  simtalk_shortest= simtalk_count; }
     if (simtalk_count> simtalk_longest)
       { test+= "simtalk_longest ";   simtalk_longest= simtalk_count; }
     test+= "simtalk_sum+="+simtalk_count+" ";
     simtalks++;  simtalk_sum+= simtalk_count;  simtalk_count= 0;
   }
   if (talked==0) //no one is talking (ntalks==0)
   {                                              //ignore silent hesitations
     if (was[speaker]!=" .. " && (last=="start"|| last=="talk"|| last=="ta- "))
       return;
     int l=lastTRPs.length(); if (l> 0) test+="utterance+="+l+" ";
     utterances+= l;               silentr_count++;  if (silence) total_notalk++;
     ag_utterances[speaker]+= l;                     if (allsilence) total_silence++;
     if (speaker> 0) { silentr_count= 1;  silentrs++;  prior= speaker;
                                      test+="silence("+speaker+") "; }
     last= was[speaker];  speaker=0;  false_count= 0;
     return;
   }
   //so from here onwards talked<>0, ntalks>0: one or more ARE talking!
```

```
    if (silentr_count> 0)
       if (ntalks==1 && talked==prior) //only SINGLE continuing speakers!
       {
         if (silentr_count> 1 || last!=" - ") { if (last==" . ") silentrs--;
             test+="continued ";  continueds++;  ag_continueds[talked]++; }
       }else
       {    //distinguish pauses (above) from silent gaps (below)
         if (silentr_shortest==0 || silentr_count< silentr_shortest)
             { test+= "gap_shortest ";  silentr_shortest= silentr_count; }
         if (silentr_count> silentr_longest)
             { test+= "gap_longest ";   silentr_longest= silentr_count; }
         silentr_sum+= silentr_count;  test+= "gap_sum+="+silentr_count+" ";
       }
    silentr_count= 0;
    for (int l, i=0; i< lastTRPs.length() ;i++) //check previous pTRPs
    {
      String t= was[l= lastTRPs.charAt(i)-'0']; //what happens after the pTRP
      if (t=="tal_" || t.startsWith(" "))
         { utterances++;  ag_utterances[l]++;  test+="utterance "; }
      if (i> 0) continue; //prevent more than 1 pTRP counting multiples below
      for (l=0; l< ntalks ;l++)  if (lastTalks.indexOf(talks.charAt(l))==-1)
          if      (t=="tal_") { overlaps++;        test+= "overlap ";
                                 if (simtalk_count==0) simtalk_count=-9999; }
          else if (isTalk(t)) { middlestarts++;  test+= "middlestart "; }
          else                { latches++;        test+= "latch "; }
    }
    if (ntalks> 1) //more than one is talking now
    {
      total_simtalk++;  simtalk_count++;  int n= 0;  boolean longsim= false;
      for (int i=0; i< talks.length() ;i++){ int l=length[talks.charAt(i)-'0'];
                            if (l==Limit) longsim= true;  if (l>=Limit) n++; }
      if (longsim && n>1){ longsimtalks++;  longsimtalk_many+= n;
                       test+="longsimtalk_many+="+n+" ";       }
      if (last=="start" && lastNtalks> 1) //MORE starters in the second cycle
      { if (ntalks>lastNtalks){ simstart_many+= ntalks-lastNtalks;
                              test+="simstart_many+="+(ntalks-lastNtalks)+" ";}
      }else
      if (last=="start" || speaker==0 || preTRP(last) && ntalks>lastNtalks+1
                             && (was[speaker]=="talk" || was[speaker]=="ta- ")
            ) //quasi-simstarts, simstarts or more-than-one middle-starts
      {
        if (last=="start") { total_singletalk--;  singlestarts--;
            /*discount 1st singlestart cycle*/     test+="nosingle "; }
        simstarts++;  simstart_many+=ntalks;  test+="simstart_many+="+ntalks+" ";
    } }
    else{ total_singletalk++;  simtalk_count= 0;  false_count= 0;
          if (speaker==0 && lastTalks.indexOf(talked+'0')< 0)
              { singlestarts++; test+="singlestarts "; }
        }
    if (speaker==0 || !isTalk(was[speaker])) //change 'main' speaker
    {
      test+= (speaker==0?"":"NEW ")+"speaker="+talked+" ";
      if (speaker==0) last="start";  prior= speaker= talked;
    }
    else last= was[speaker];
}
```

```
  boolean isTalk(String did) { return did.startsWith("ta") || preTRP(did)
                                                    || did=="TALK"; }
  boolean preTRP(String did) { for (int i=1;  i<= size  ;i++)
                                    if (did==Talker.Names[i]) return true;
                                return did=="pTRP"|| did=="More"|| did=="Any?"; }
  void countPreTRPs(String did){ if (did=="More") mores++; else
                                 if (did=="Any?") anys++; else
                                 if (did!="pTRP") sels++; }


  void showMeasures(int divisor) //results are all means by the number of runs
  {
    double cycle= Talker.cycleTime;
    System.out.println( "\nTotals: amount of single talk ("+
       d(cycle*total_singletalk/divisor)+ "s), of no-talk/total silence ("+
       d(cycle*total_notalk/divisor)+      "s/"+
       d(cycle*total_silence/divisor)+    "s)\n\tsimult.talk with overlaps ("+
       d(cycle*total_simtalk/divisor)+    "s), feedback/backchannels ("+
       d(cycle*total_feedback/divisor)+   "s/"+
       d(cycle*total_backchannel/divisor)+"s)");
    System.out.println( "silent gaps  (" +silentrs/divisor+ "): total ("+
       d(cycle*silentr_sum/divisor)+      "s), mean ("+
       d(cycle*silentr_sum/silentrs)+     "s), shortest ("+
       d(cycle*silentr_shortest)+         "s), longest ("+
       d(cycle*silentr_longest)+          "s)");
    System.out.println( "simult.talks (" +simtalks/divisor+ "): total ("+
       d(cycle*simtalk_sum/divisor)+      "s), mean ("+
       d(cycle*simtalk_sum/simtalks)+     "s), shortest ("+
       d(cycle*simtalk_shortest)+         "s), longest ("+
       d(cycle*simtalk_longest)+          "s)");
    System.out.println("simult.starts("+simstarts/divisor+"), mean #speakers("+
       d(simstart_many/(double)simstarts)+                "); long sim.talks("+
       longsimtalks/divisor+                              "), mean #speakers("+
       d(longsimtalk_many/(double)longsimtalks)+ ")");
    System.out.println( "middle starts("+
       middlestarts/divisor+ "), false-starts/incompletes("+
       falsestarts/divisor+ "+" +botcheds/divisor+ "/" +incompletes/divisor+
       "), overlaps/latches("  +overlaps/divisor+ "/" +latches/divisor +")" );
    System.out.print( "single starts("+ singlestarts/divisor+ "), utterances ("+
                    utterances/divisor+ ":"+ ag_utterances[1]/divisor);
    for (int i=2; i<=size ;i++)  System.out.print("+" +ag_utterances[i]/divisor);
    System.out.print( "), continueds ("+
                    continueds/divisor+ ":"+ ag_continueds[1]/divisor);
    for (int i=2; i<=size ;i++)  System.out.print("+" +ag_continueds[i]/divisor);
    if (mores+anys+sels==0)  System.out.println(")\n");
    else System.out.println( "),\n\t\t   (numbers of More: " +mores/divisor
          +", Any?: " +anys/divisor+ ", select-next: " +sels/divisor+ ").\n");
  }
 double d(double x){ return Math.round(x*10.0)/10.0; }//round to 1 decimal point
}


public class mtest   //it's the same 'model' with a slightly modified main()
{
 static Group group;
 static final int maxSize= 7; //maximum size of 'small' groups
```

```
static public void  main(String[] args)
{
 int size; //first argument is the group size

 if (args.length < 1 || (size= Integer.parseInt(args[0]))< 2 || size> maxSize)
 {
   group= new Group(1); //getting the default attribute values
   System.out.println(
    "\nProvide the group size as the first argument (max. "+ maxSize +").\n"+
    "Then optionally in any order, the cycle-time defining the simulation's\n"+
    "resolution or granularity (-X), the maximum length of any utterances \n"+
    "in the discussion (+X), the frequency of negative feedback (?X), and the"+
    "\nattributes (cXYZ..), where X (Y,Z,..) in all these parameters is a dig"+
    "it\ncorresponding to their one-tenth probabilities (0.X, 0.Y, 0.Z) or se"+
    "conds\nfor the maximum utterance length (+X), and c is the attribute:\n"+
    "\nt - Talkativeness: want to talk at TRPs or when no one is talking?\n"+
    "e - Eagerness: start immediately or wait for the speaker to stop?\n"+
    "f - Feedback: give feedback (uhuh or huh?) when NOT wanting to talk?\n"+
    "c - Confidence: persist in talking simultaneously with others?\n"+
    "a - Attention: time in secs. of perceiving others talking (unused)\n"+
    "v - Verbosity: continue to talk (by starting new utterances)?\n"+
    "p - Pausing: mean length of pauses before starting new utterances\n"+
    "i - Interactivity: pre-TRPs in mid-utterance and trailing-off talk\n"+
    "h - Hesitation: pausing or disfluencies in mid-utterance\n"+
    "s - Selectivity: select the next speaker to talk?\n"+
    "n - Nonverbal: frequency of non-verbal behaviour.\n\n"+
    "If there is more than one digit (cXYZ..), one is for each agent in the\n"+
    "group; if there is only one instead (cX), the same value is for all.\n"+
    "\nFor example:\t\tjava model 4 -5 ?0 t3456 v5 e6\n"+
    "starts a 4-group discussion with a cycle-time of 0.5s, no negative\n"+
    "feedback, agents with talkativeness 0.3, 0.4, 0.5, 0.6 respectively,\n"+
    "and all with verbosity 0.5 and eagerness 0.6.\n\n"+
    "Defaults (when no parameter is provided):\n"+
    "- cycle time (-X): "+Talker.cycleTime+"s\n"+
    "- maximum length of utterances (+X): "+Talker.MaxUtterances+"s\n"+
    "- frequency of negative feedback (?X): "+Talker.NotUnderstand+"\n\n"+
    "Attribute defaults (same value for all participants):\n"+
    "- Talkativeness (tXXXX): "+group.participant[1].TALKATIVENESS+"\n"+
    "- Eagerness (eXXXX): "+group.participant[1].EAGERNESS+"\n"+
    "- Feedback (fXXXX): "+group.participant[1].FEEDBACK+"\n"+
    "- Confidence (cXXXX): "+group.participant[1].CONFIDENCE+"\n"+
    "- Attention (aXXXX): "+group.participant[1].ATTENTION+"\n"+
    "- Verbosity (vXXXX): "+group.participant[1].VERBOSITY+"\n"+
    "- Pausing (pXXXX): "+group.participant[1].TRPAUSING+"s\n"+
    "- Interactivity (iXXXX): "+group.participant[1].INTERACTIVITY+"\n"+
    "- Hesitation (hXXXX): "+group.participant[1].HESITATION+"\n"+
    "- Selectivity (sXXXX): "+group.participant[1].SELECTIVITY+"\n"+
    "- Nonverbal (nXXXX): "+group.participant[1].NONVERBAL+"\n");
   return;
 }
 group= new Group( size ); //create the group
```

```
   for (int i=1;  i< args.length  ;i++)
     if (args[i].charAt(0)=='-')  Talker.cycleTime= (args[i].charAt(1)-'0')/10.0;
     else if (args[i].charAt(0)=='+')
                     Talker.MaxUtterances= Integer.parseInt(args[i].substring(1));
     else if (args[i].charAt(0)=='?')
                                Talker.NotUnderstand= (args[i].charAt(1)-'0')/10.0;
     else if (args[i].length()==2)  for (int k=1;  k<= Group.size  ;k++)
             group.participant[k].attribute( args[i].charAt(0),
                                           (args[i].charAt(1)-'0')/10.0 );
     else for (int k=1;  k<= Group.size && k< args[i].length()  ;k++)
             group.participant[k].attribute( args[i].charAt(0),
                                           (args[i].charAt(k)-'0')/10.0 );

  System.out.print("Press ENTER, or SPACE+ENTER for nonstop evaluation.");
  boolean nonstop=false;  InputStreamReader x= new InputStreamReader(System.in);
  try{ if (x.read()==' ') nonstop=true; } catch (IOException _) {return;}
  if (!nonstop)
  {                                    System.out.print("\nTime    ");
    for (int i=1;  i<= size  ;i++)  System.out.print(Talker.Names[i]+"|");
    System.out.print("\n         ");
    for (int i=1;  i<= size  ;i++)  System.out.print("-----");
    System.out.println("   speaker(#speakers,TRPs) --identified measures");
  }
  int time=0,  Duration= (int)Math.round(5*60/Talker.cycleTime);   //5 min.=300s
  for (int k=0; k< 3 ;k++)
  {
    for (int l=0; l< 50 ;l++) //runs 50 5-minute discussions thrice
    {
      Group.seed=new Random();  for (int i=0; i< Duration ;i++)
      {
        String[] blackboard=group.cycle();  if (nonstop) continue;
        System.out.print(time/10 +"."+ time%10);
        System.out.print(time<100? "    ":time<1000? "   ":"  ");
        for (int j=1;  j<= size  ;j++)  System.out.print(blackboard[j]+ " ");
                                        System.out.print(blackboard[0]);
        time+= (int)10*Talker.cycleTime; //clock up the time
        try { int key;  if ((key=x.read())==' ')  nonstop=true;  else
                        if (key!='\n') return; } catch (IOException _) {return;}
      }
    }   if (k==2) break;
    System.out.print("(press ENTER for the randomic seeding) ");
    try { if (k==0) x.read(); x.read(); } catch (IOException _) {return;}
  }
  group.showMeasures(150); //show all the averages: sums divided by 150
 }
}
```

Appendix C

Evaluation results

This appendix arrays the charts for twenty main measures (8 total *times* and 12 *numbers*) from all the twelve evaluations of the simulation described in chapter 6, eight in the basic model and four in the three extended models. The evaluations of the basic model were:

- in four group sizes with EAGERNESS variation,

- in two group sizes with a different maximum utterance length,

- and in 5-agent groups for CONFIDENCE and INTERACTIVITY variation.

Here I separated the total *times* from *numbers*, putting together the total averaged times of all the first eight evaluations of the basic model so that they can be more easily compared together. These are the charts that follow on the next eight pages, one per evaluation, in order:

- times of 3-agent groups (triads, referred to as "3's"), 4-agent groups (tetrads: "4's"), 5-agent groups (pentads: "5's"), and 6-agent groups (hexads: "6's"), ranging in TALKATIVENESS at the horizontal axis and in five EAGERNESS settings for all agents according to the side legend ('e0', 'e0.2', etc), with all other relevant attributes (FEEDBACK, CONFIDENCE, INTERACTIVITY, VERBOSITY) at a middle likelihood of 0.5, and with maximum utterance length of 4 seconds;

- times in 3- and 5-agent groups with the maximum utterance length increased to 12 seconds, noted by the '+12' in the chart titles (all other attributes the same);

- times for 5's now varying in CONFIDENCE (note the side legends: 'c0', 'c0.2',

etc), with all other parameters the same, including EAGERNESS at 0.5; and

- times for 5's now varying in INTERACTIVITY (see the legend: 'i0', 'i0.2', etc), with EAGERNESS and CONFIDENCE at the 0.5.

The charts are positioned in each page so that associated measures can be compared side-by-side. Thus, comparing the amount of silent gaps to that of 'no talk' to the right in each page, one can see the additional amount of time that corresponds to the total of pauses between same-speaker utterances, which is significant for low values of TALKATIVENESS but disappears completely in higher ones.

Likewise, comparing the time of simultaneous talk *excluding* overlaps (non-conflicting simultaneous talk) with the chart *including* them to the side, one can see the extra time they take at speaker transitions, as TALKATIVENESS and specially EAGERNESS increase. And comparing the amount of *backchannels* (feedback in the background of talk) with the total of all feedback vocalizations, one sees the extra amount of time of feedback that goes in the silences between utterances: in pauses or gaps. This difference is the same as that between the time of 'total silence' to that of 'no talk'.

Finally, note that the charts of single talk, the three charts for 'silences', and the four other ones in each page are in different scales of the vertical axis (seconds). Times of single talk are in the scale of the total time of each simulated discussion that constitutes these averages: 300 seconds. The three 'silence' times are in a 60-second scale (save one case), which is one-fifth of the total time of the discussions. And the rest, the simultaneous talk and feedback charts, are in a 100-second scale: one-third of the total time of discussions.

Time of single talk in 3's

Time of total silence in 3's

Time of silent gaps in 3's

Time of gaps+pauses (no talk) in 3's

Time of simult. talk in 3's

Time of sim.talk+overlaps in 3's

Time of backchannels in 3's

Time of total feedback in 3's

Time of single talk in 4's

Time of total silence in 4's

Time of silent gaps in 4's

Time of gaps+pauses (no talk) in 4's

Time of simult. talk in 4's

Time of sim.talk+overlaps in 4's

Time of backchannels in 4's

Time of total feedback in 4's

Time of single talk in 5's

Time of total silence in 5's

Time of silent gaps in 5's

Time of gaps+pauses (no talk) in 5's

Time of simult. talk in 5's

Time of sim.talk+overlaps in 5's

Time of backchannels in 5's

Time of total feedback in 5's

Time of single talk in 6's

Time of total silence in 6's

Time of silent gaps in 6's

Time of gaps+pauses (no talk) in 6's

Time of simult. talk in 6's

Time of sim.talk+overlaps in 6's

Time of backchannels in 6's

Time of total feedback in 6's

Time of single talk in 3's (+12)

Time of total silence in 3's (+12)

Time of silent gaps in 3's (+12)

Time of gaps+pauses (no talk) in 3's (+12)

Time of simult. talk in 3's (+12)

Time of sim.talk+overlaps in 3's (+12)

Time of backchannels in 3's (+12)

Time of total feedback in 3's (+12)

Time of single talk in 5's (+12)

Time of total silence in 5's (+12)

Time of silent gaps in 5's (+12)

Time of gaps+pauses (no talk) in 5's (+12)

Time of simult. talk in 5's (+12)

Time of sim.talk+overlaps in 5's (+12)

Time of backchannels in 5's (+12)

Time of total feedback in 5's (+12)

Time of single talk in 5's (e0.5)

Time of total silence in 5's (e0.5)

Time of silent gaps in 5's (e0.5)

Time of gaps+pauses (no talk) in 5's (e0.5)

Time of simult. talk in 5's (e0.5)

Time of sim.talk+overlaps in 5's (e0.5)

Time of backchannels in 5's (e0.5)

Time of total feedback in 5's (e0.5)

Time of single talk in 5's (e,c0.5)

Time of total silence in 5's (e,c0.5)
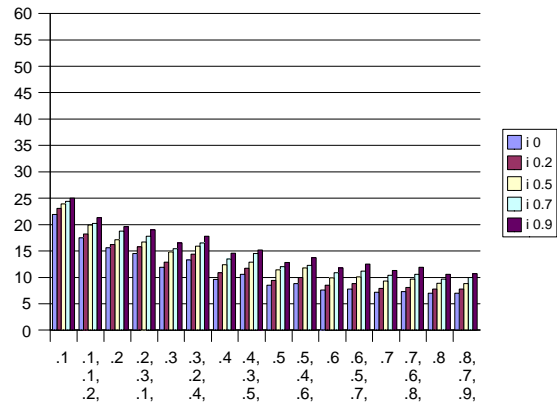
Time of silent gaps in 5's (e,c0.5)
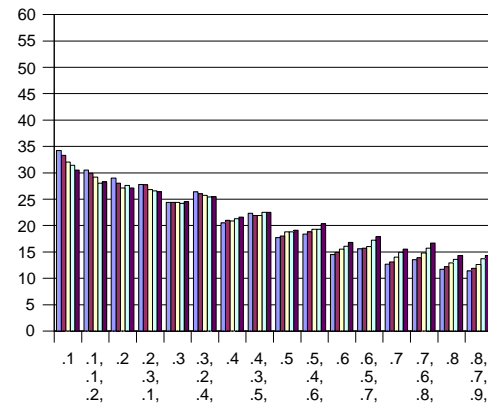
Time of gaps+pauses (no talk) in 5's (e,c0.5)
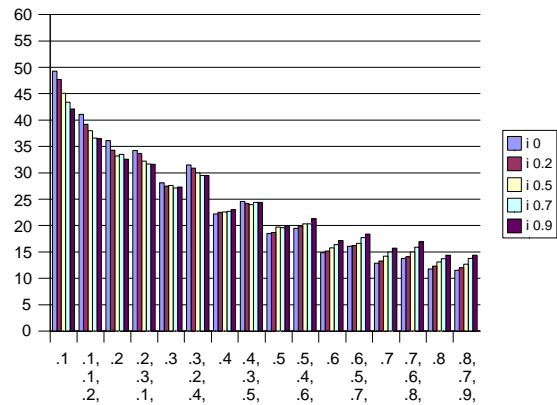
Time of simult. talk in 5's (e,c0.5)
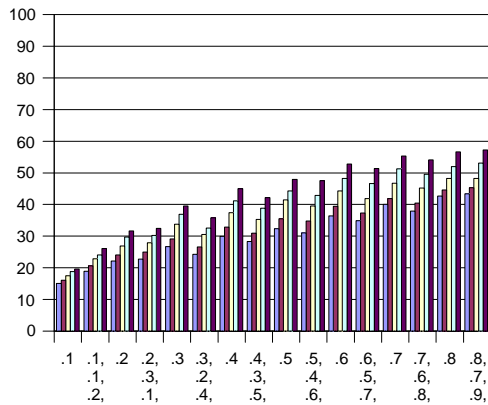
Time of sim.talk+overlaps in 5's (e,c0.5)
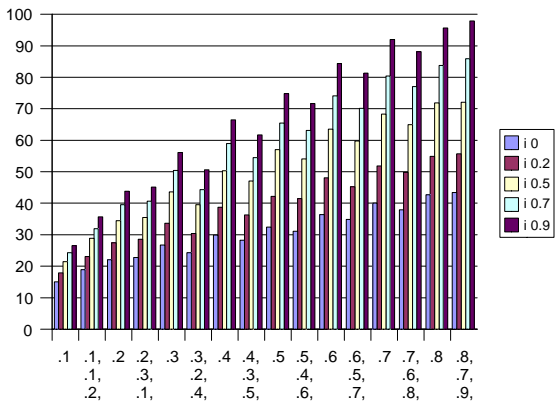
Time of backchannels in 5's (e,c0.5)
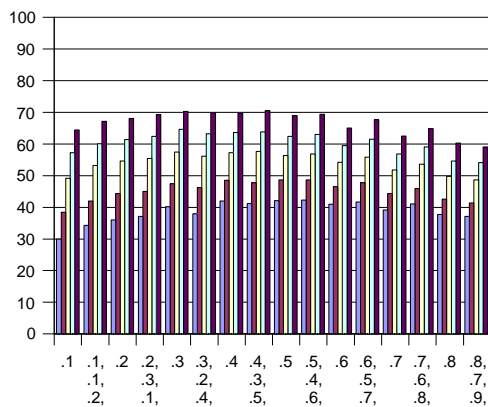
Time of total feedback in 5's (e,c0.5)
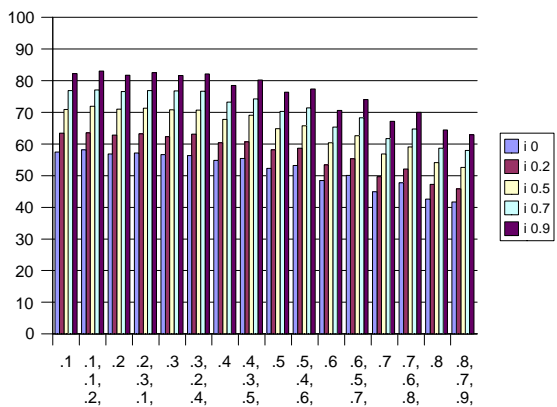
The charts on the next pages present the 'number' measures of the same eight evaluations of the basic model described previously:

- for 3's, 4's, 5's and 6's groups with EAGERNESS variation,

- for 3's and 5's with an increased maximum utterance length of 12 seconds,

- for 5's varying in CONFIDENCE instead of EAGERNESS, and

- for 5's varying in INTERACTIVITY, with all other parameters at 0.5.

There are twelve measures now, so each evaluation occupies one and a half pages. This means that the charts for each two evaluations are arrayed in three consecutive pages, with the middle one grouping for the two the *failure* measures of false-starts (individual and collective), incomplete utterances and 'long' simultaneous talks. Notice that these are all related to one another in their positions on the page, horizontally and vertically: individual false-starts and incomplete utterances are, respectively, self-interrupted talk of less than one second and of one second or more; 'collective' false-starts and long simultaneous talks are simultaneous talks that are either all-interrupted or not, and either less than one second or of one second or more.

Of the other measures, it is useful to re-stress the following. First, the number of continuing utterances is also part of the total number of utterances. Second, transitions between any two utterances can either be via a *pause* (of same-speaker continuing utterances), a silent *gap* (between non-continuing utterances), an *overlap* of different-speaker utterances, or a pefect *latch* when the transition leaves no gap or overlap. In the evaluations which have INTERACTIVITY fixed at 0.5, the number of overlaps was generally a couple more than the number of latches.

The scales in the vertical axis now indicate number of occurrences instead of seconds. Whenever possible, charts were rendered in the same scale, or at least compared to equivalent ones (such as e.g. those for false-starts).
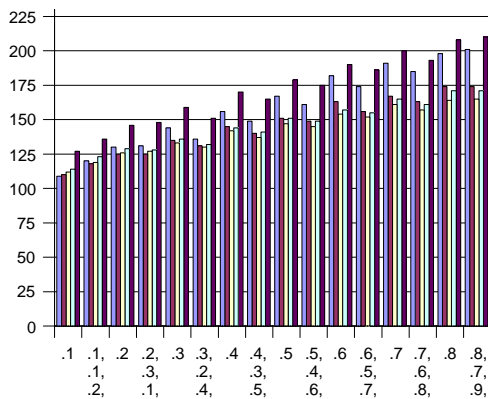
Number of utterances in 3's
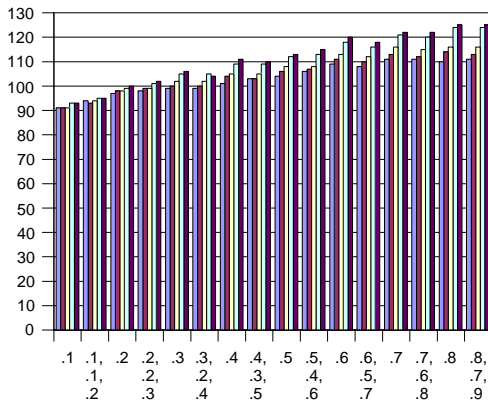
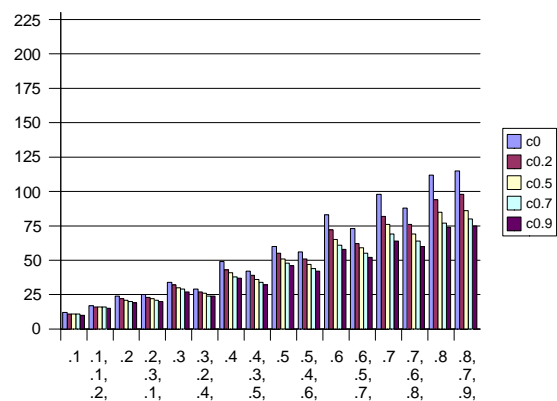## Continuing utterances in 3's
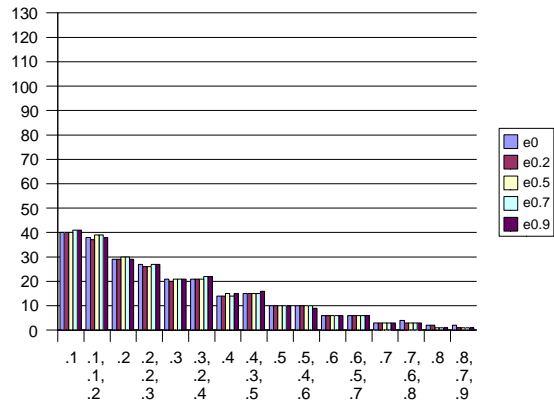
Number of silent gaps in 3's
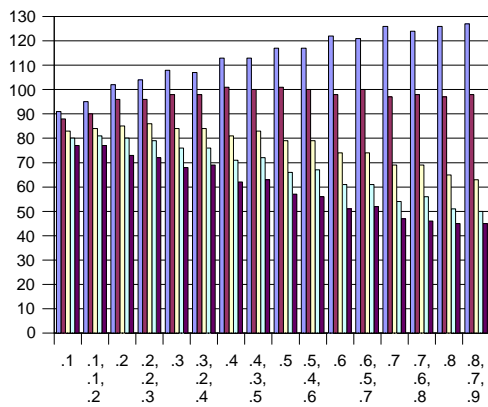
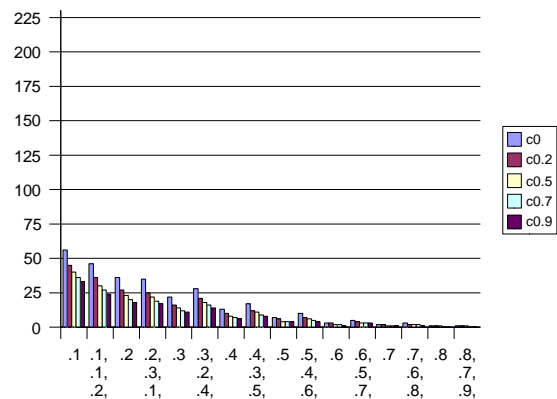## Number of overlaps in 3's
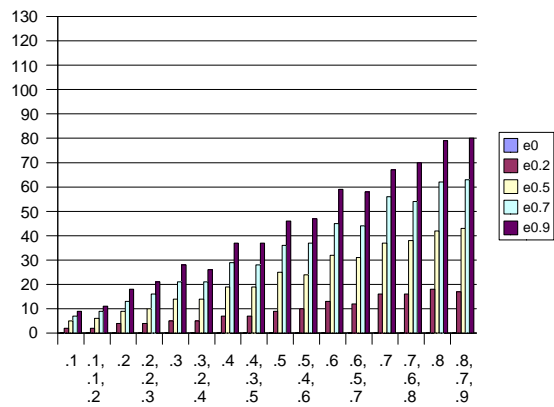
Number of sim. talks in 3's
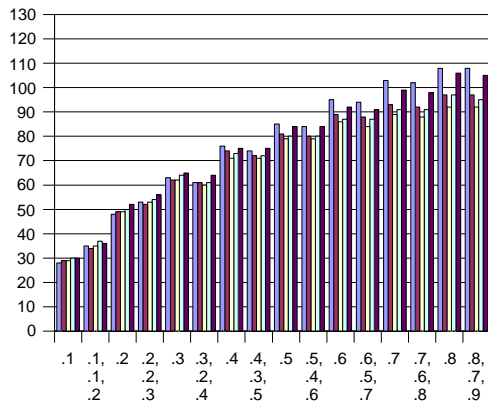
Number of single starts in 3's
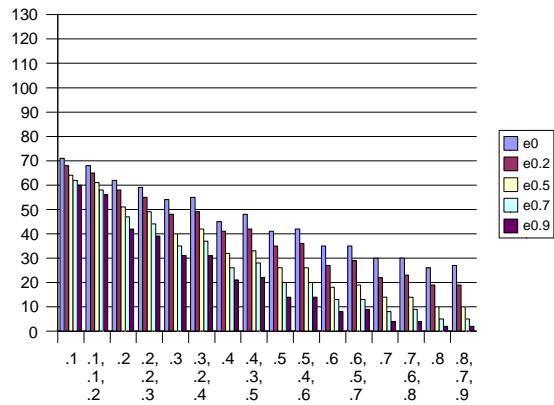
Number of sim. starts in 3's
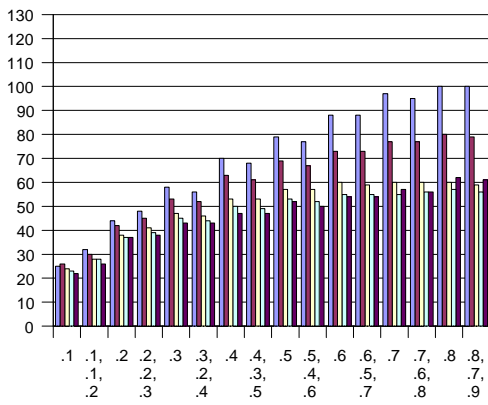
Number of middle starts in 3's
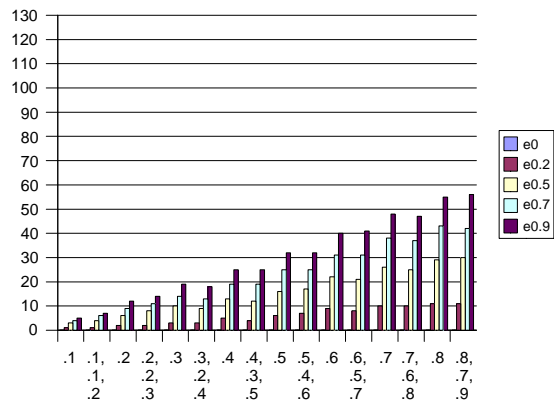
Number of false-starts in 3's
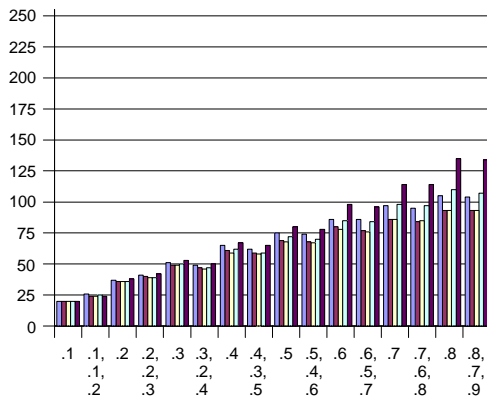
'Collective' false-starts in 3's
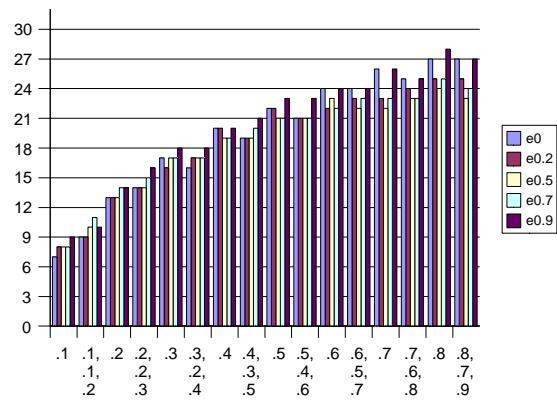
Incomplete utterances in 3's
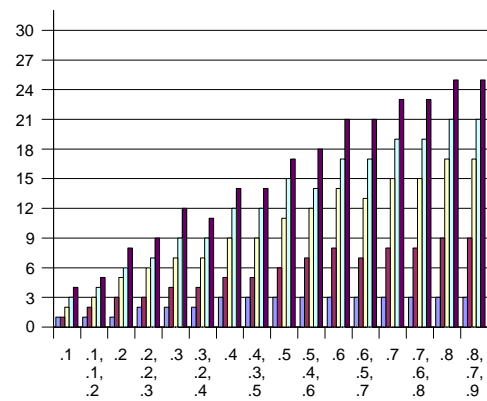
Number of long sim. talks in 3's
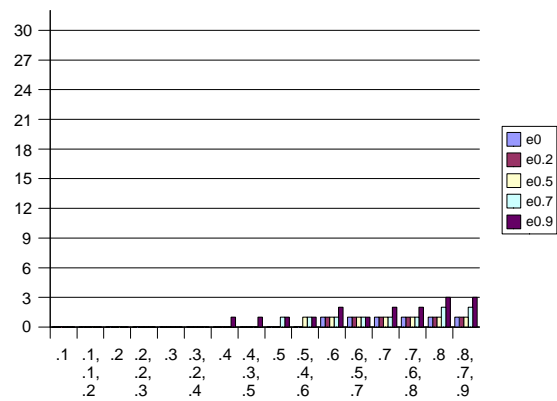
Number of false-starts in 4's
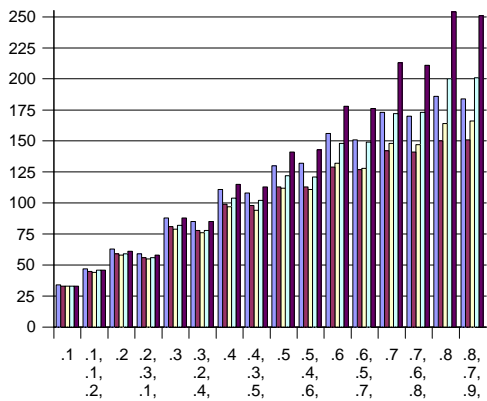
'Collective' false-starts in 4's
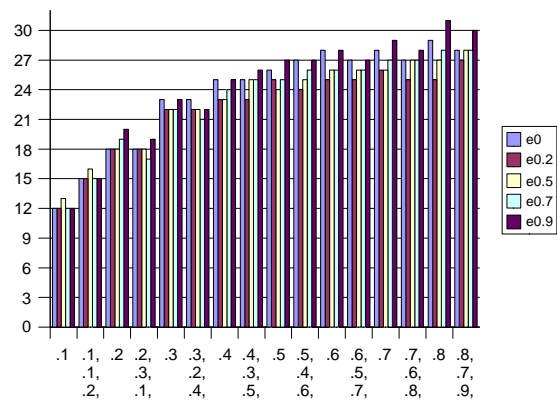
Incomplete utterances in 4's
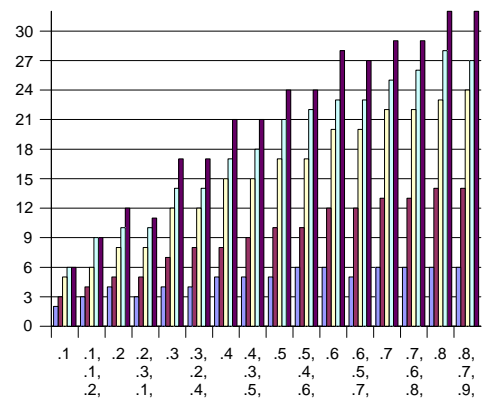
Number of long sim. talks in 4's
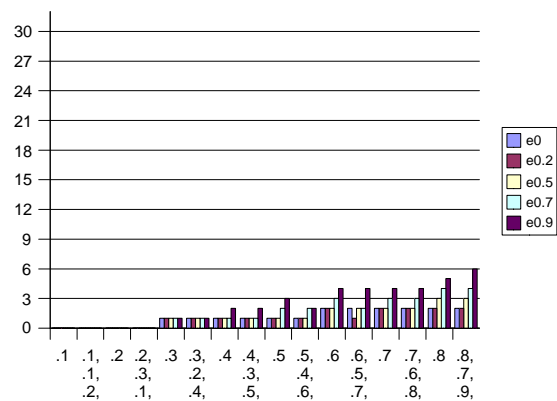
Number of utterances in 4's
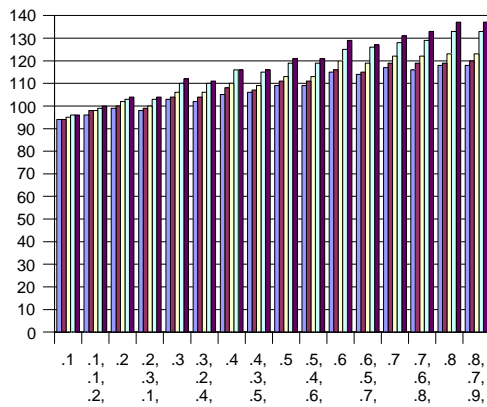
Continuing utterances in 4's
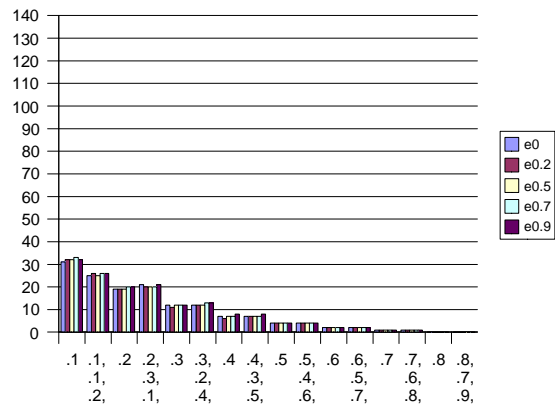
Number of silent gaps in 4's
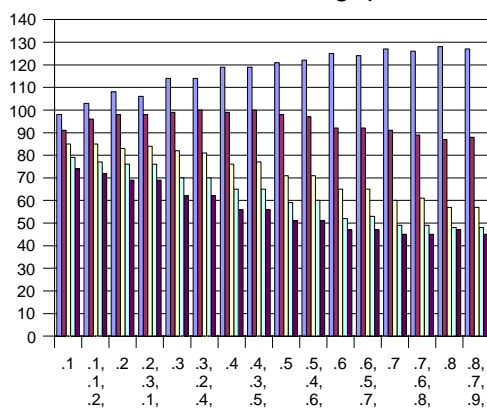
Number of overlaps in 4's
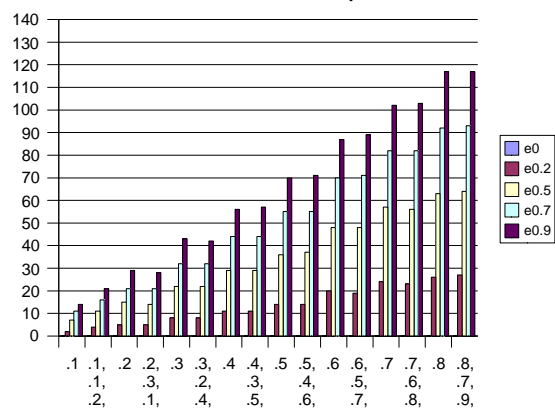
Number of sim. talks in 4's
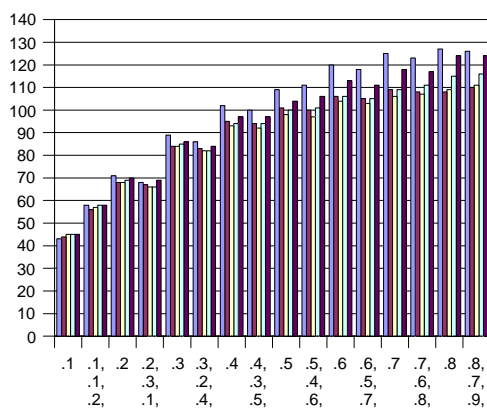
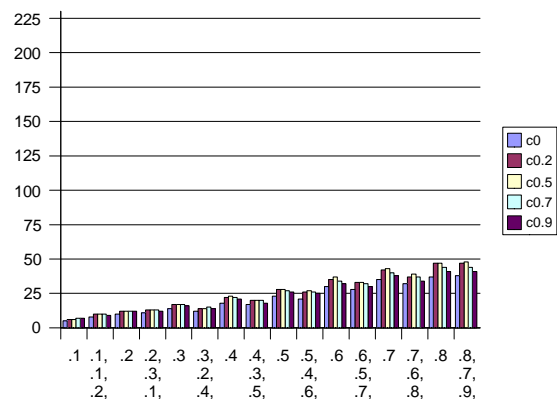### Number of single starts in 4's
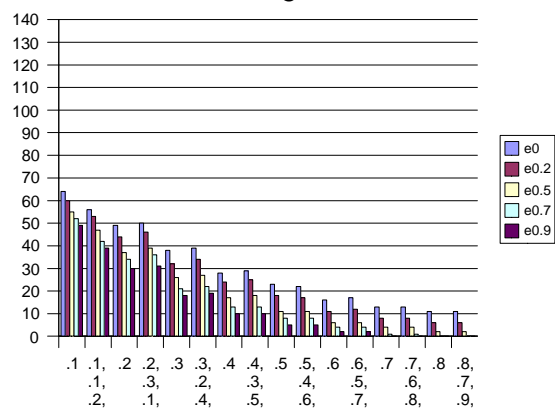
Number of sim. starts in 4's

Number of middle starts in 4's

Number of utterances in 5's

Continuing utterances in 5's

Number of silent gaps in 5's

Number of overlaps in 5's

Number of sim. talks in 5's

Number of single starts in 5's

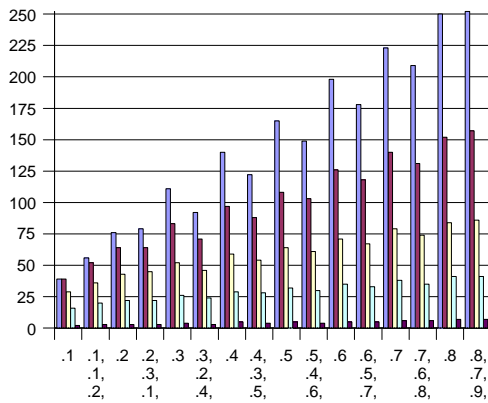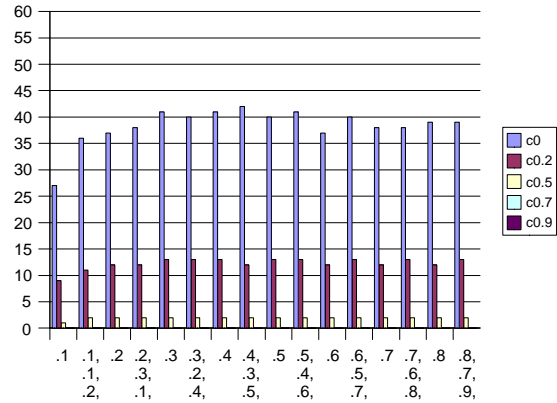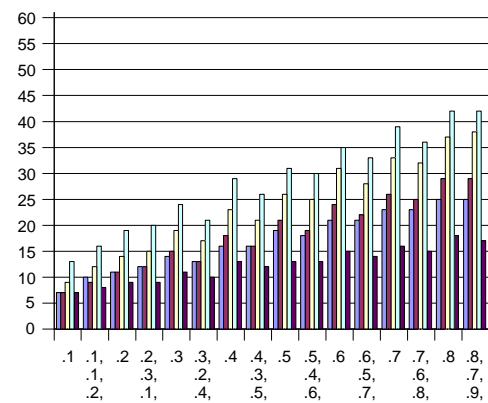### Number of sim. starts in 5's
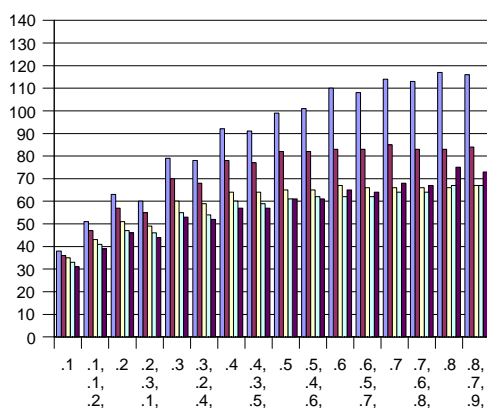
Number of middle starts in 5's
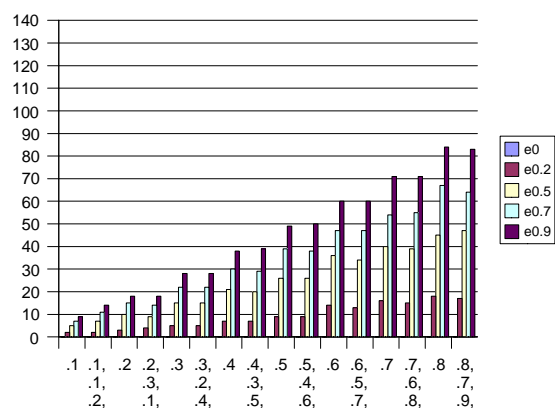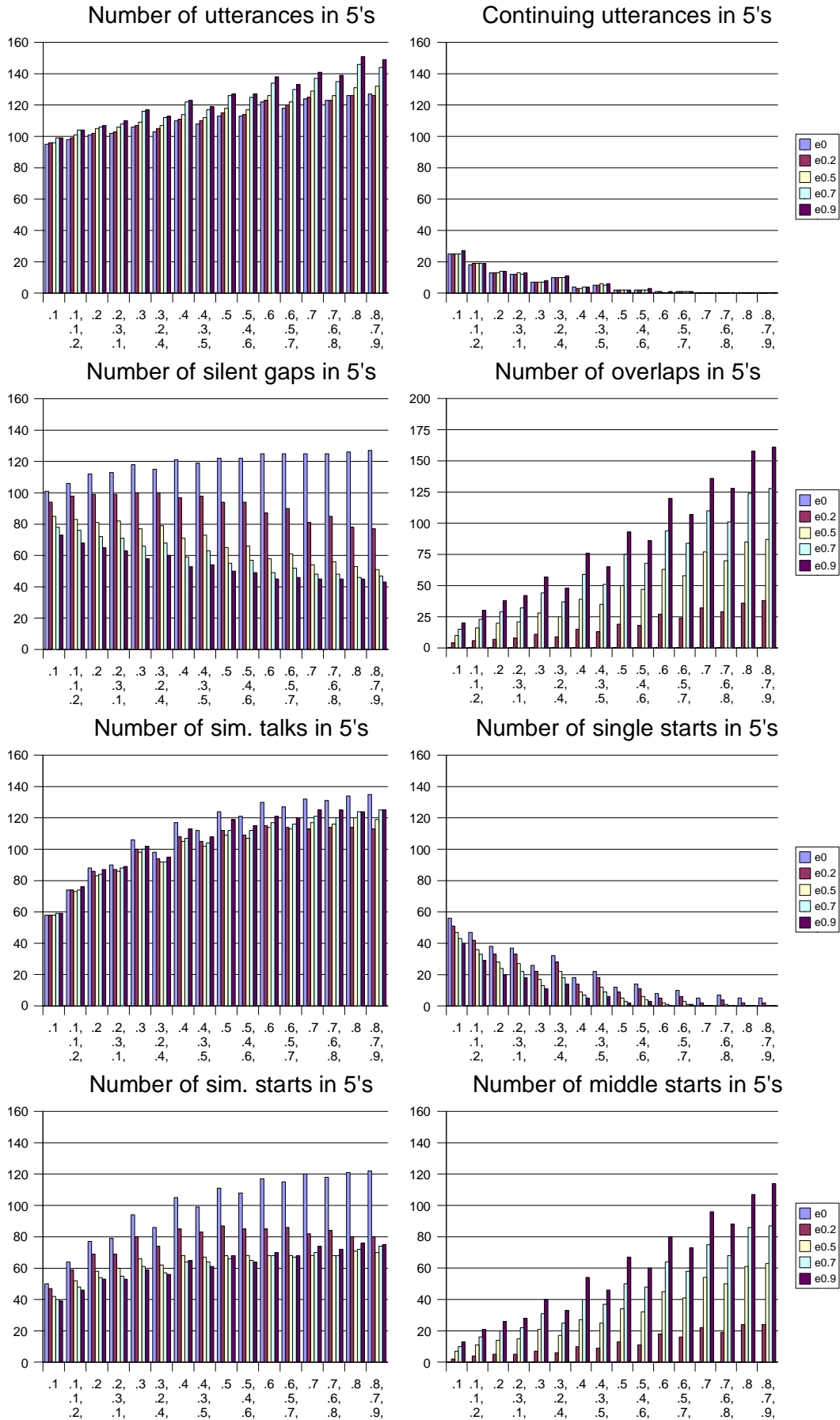
Number of false-starts in 5's
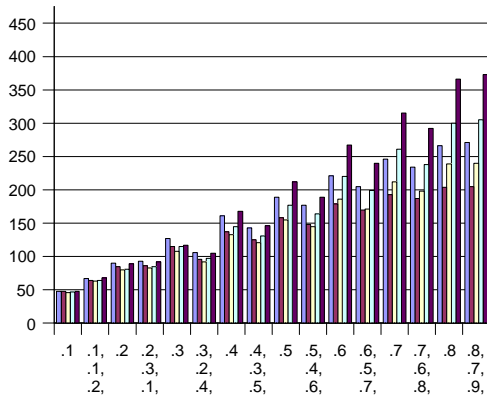
'Collective' false-starts in 5's
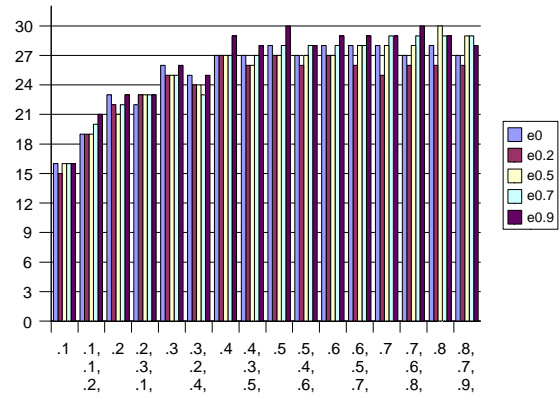
Incomplete utterances in 5's
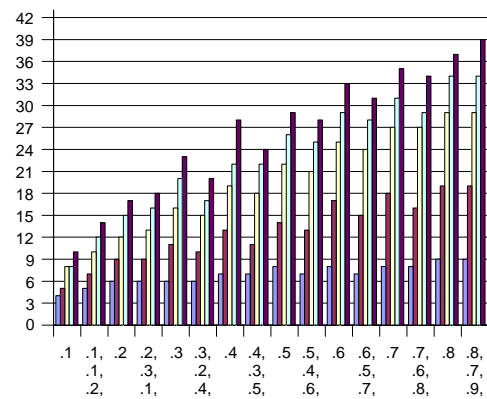
Number of long sim. talks in 5's
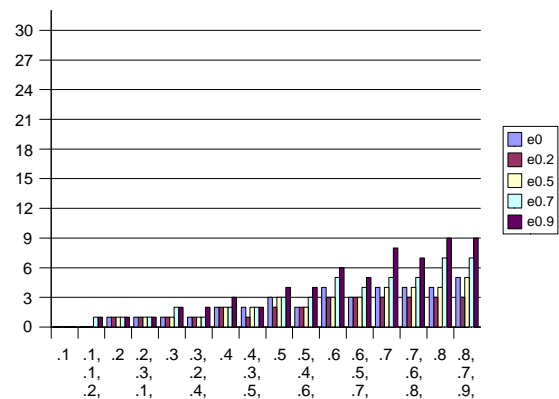
Number of false-starts in 6's
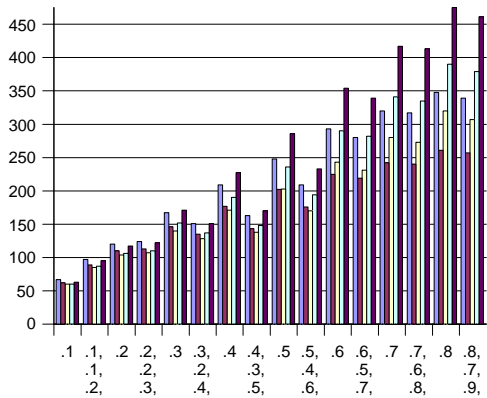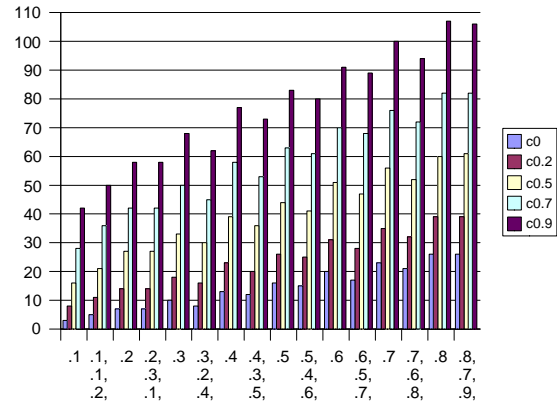
'Collective' false-starts in 6's
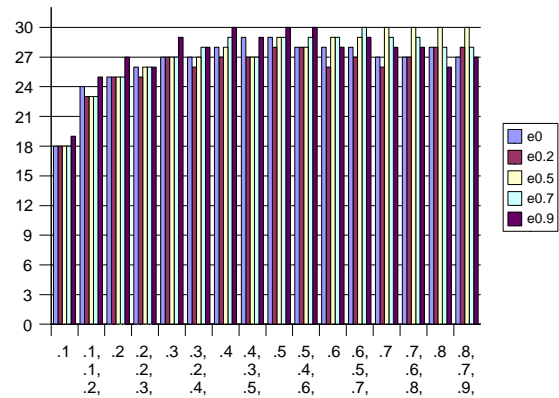
Incomplete utterances in 6's
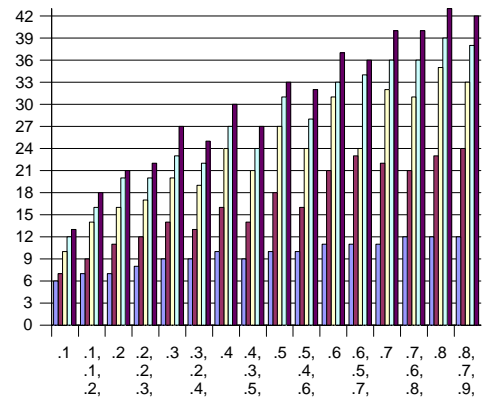
Number of long sim. talks in 6's
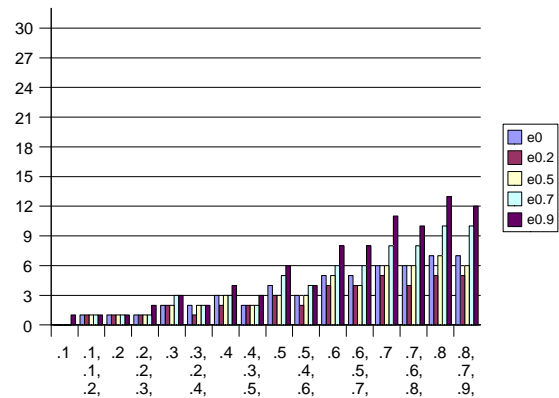
Number of utterances in 6's
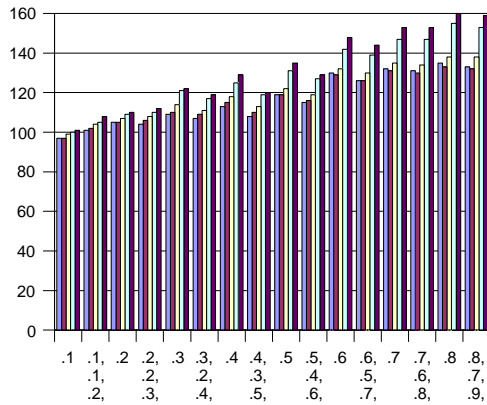
Continuing utterances in 6's
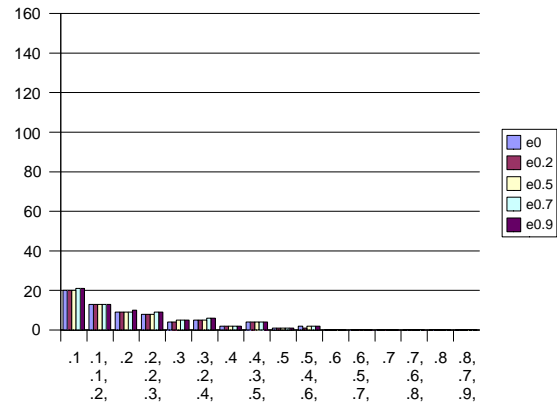
Number of silent gaps in 6's
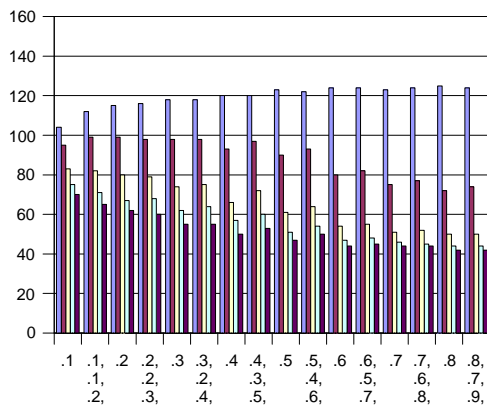
Number of overlaps in 6's
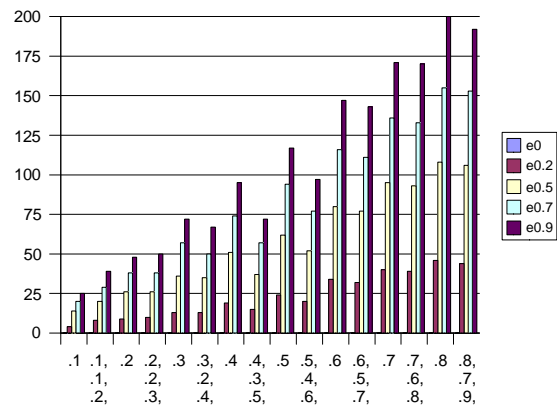
Number of sim. talks in 6's
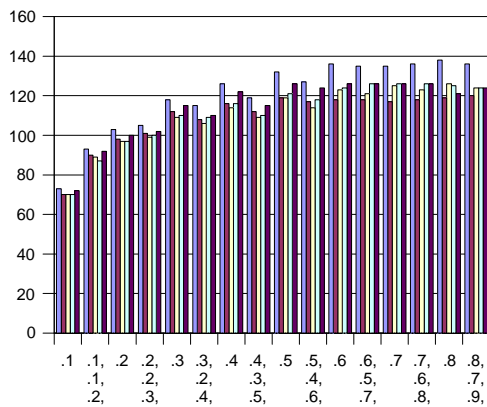
Number of single starts in 6's
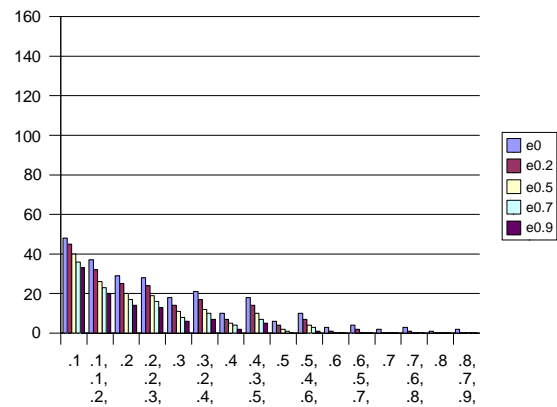
Number of sim. starts in 6's
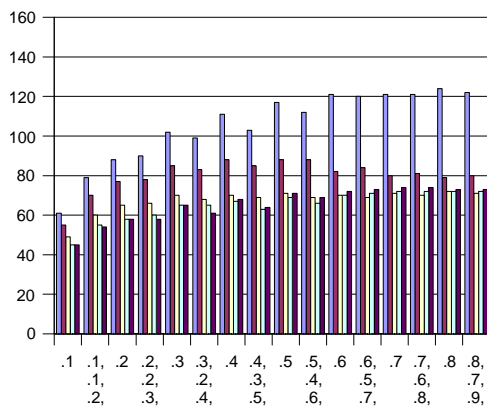
Number of middle starts in 6's
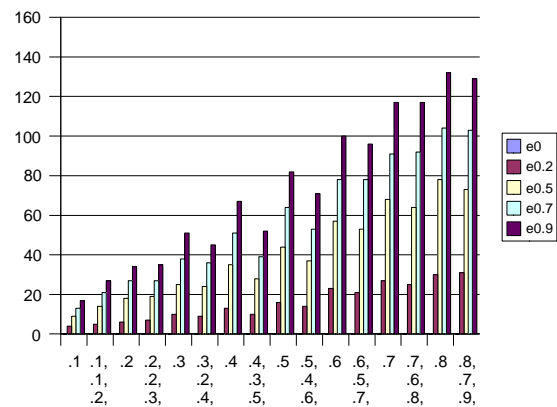
Number of utterances in 3's (+12)
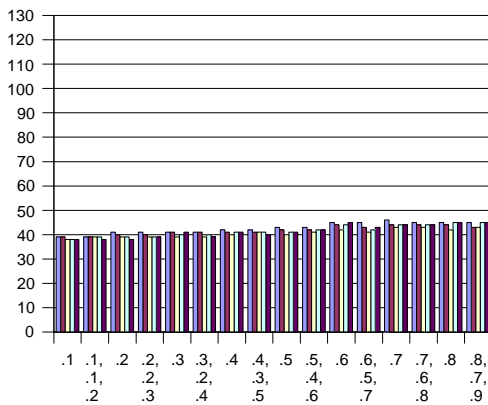
Continuing utterances in 3's (+12)
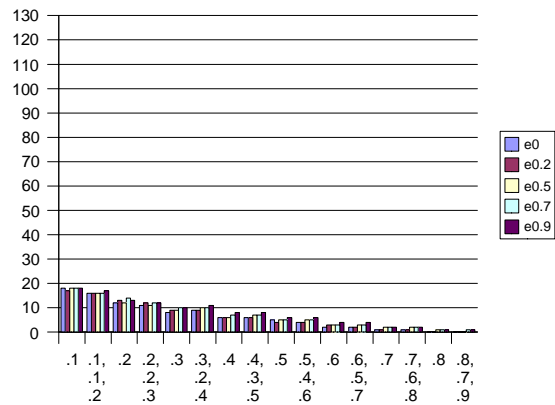
Number of silent gaps in 3's (+12)
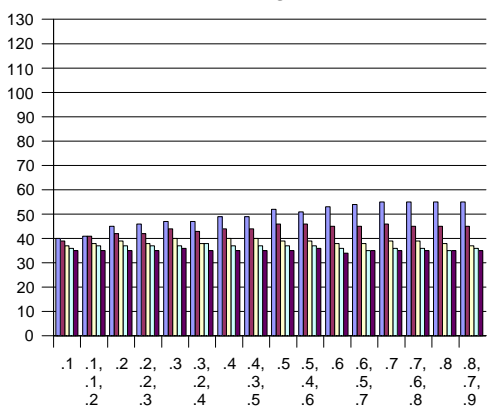
Number of overlaps in 3's (+12)
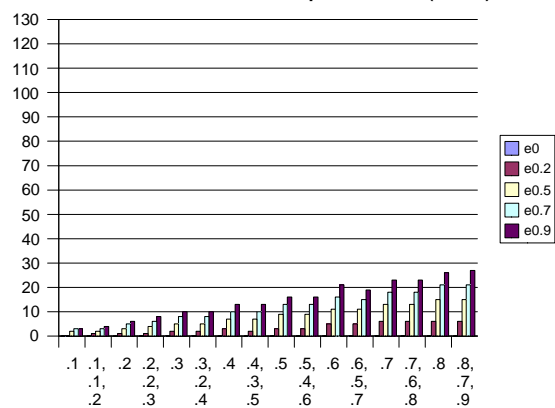
Number of sim. talks in 3's (+12)
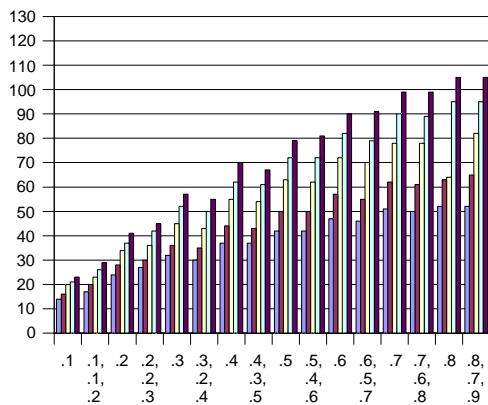
Number of single starts in 3's (+12)
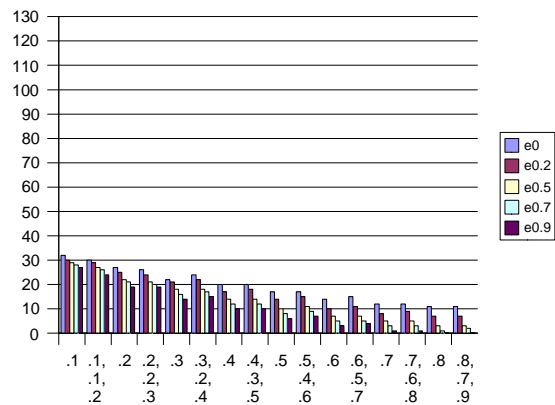
Number of sim.starts in 3's (+12)
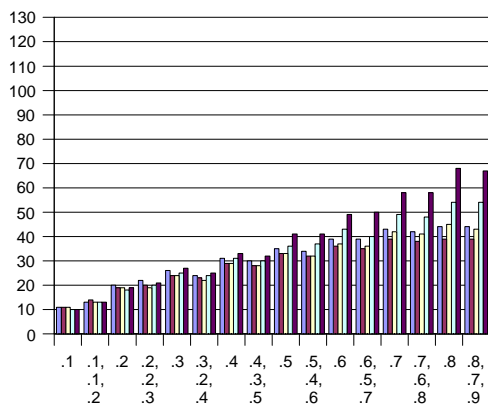
Number of middle starts in 3's (+12)
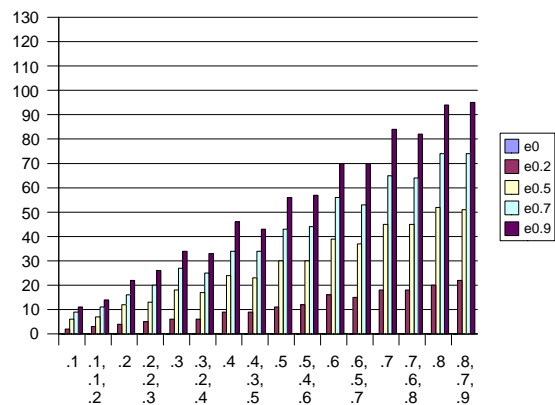
Number of false-starts in 3's (+12)
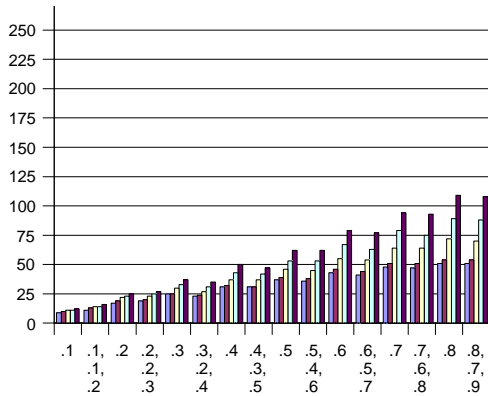
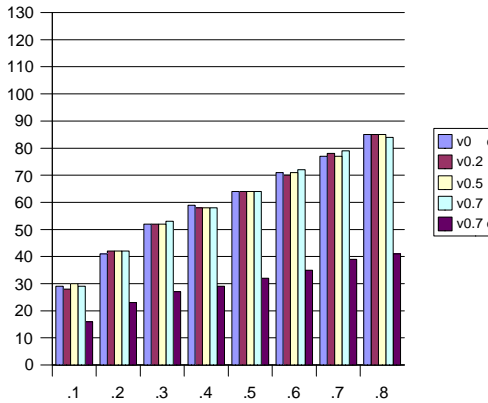## 'Collective' false-starts in 3's (+12)
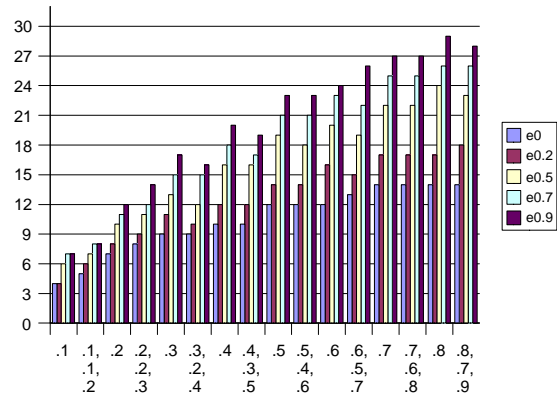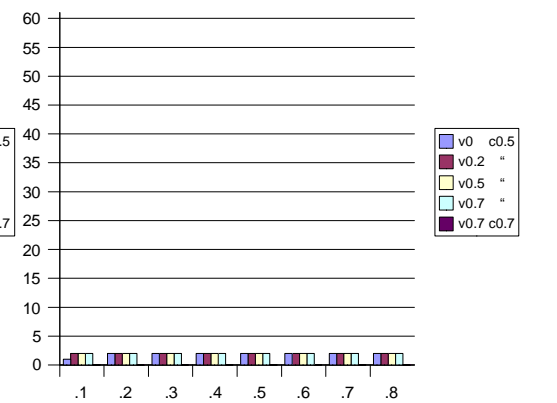
Incomplete utterances in 3's (+12)
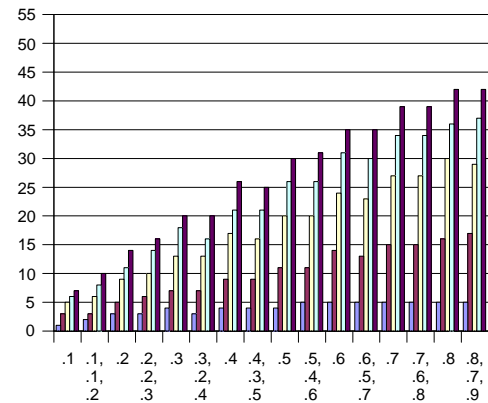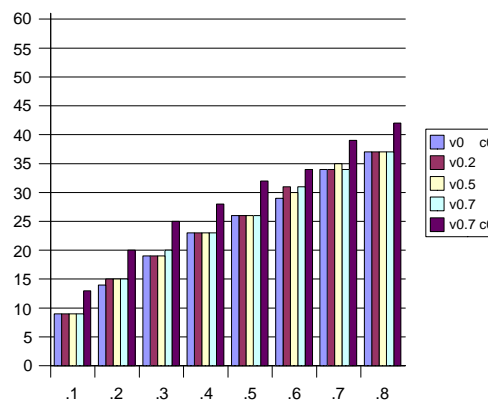
Number of long sim. talks in 3's (+12)
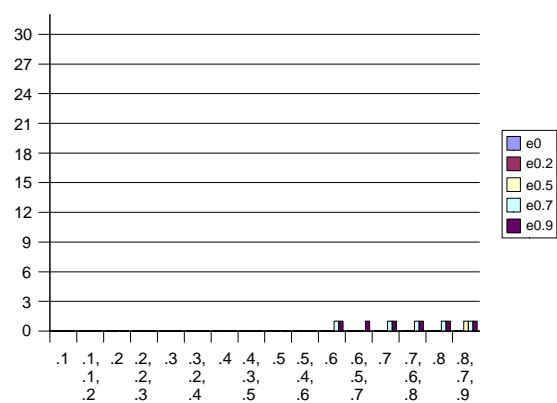
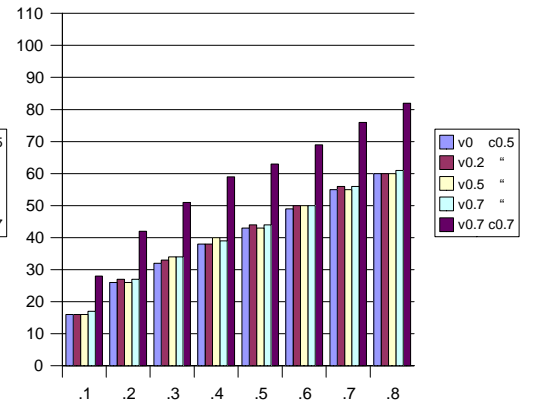## Number of false-starts in 5's (+12)
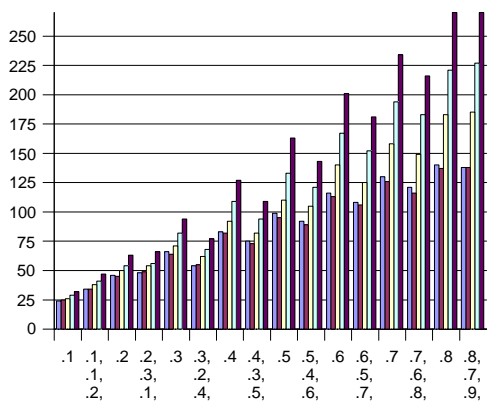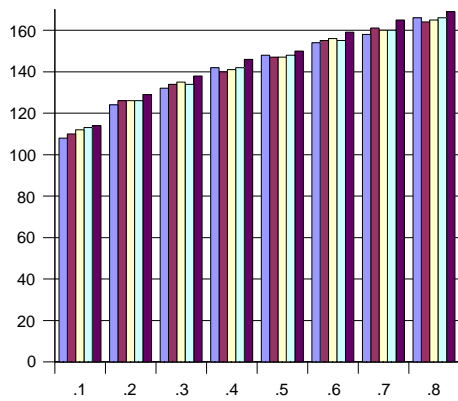
'Collective' false-starts in 5's (+12)
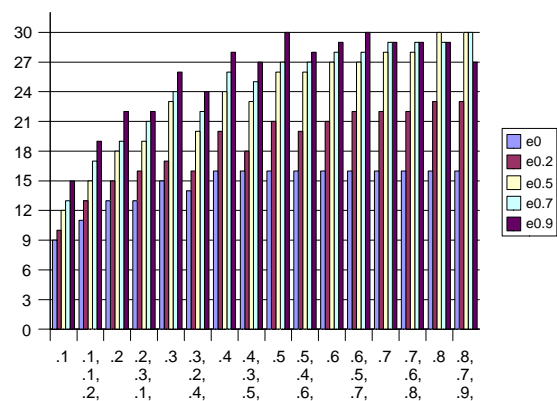
Incomplete utterances in 5's (+12)
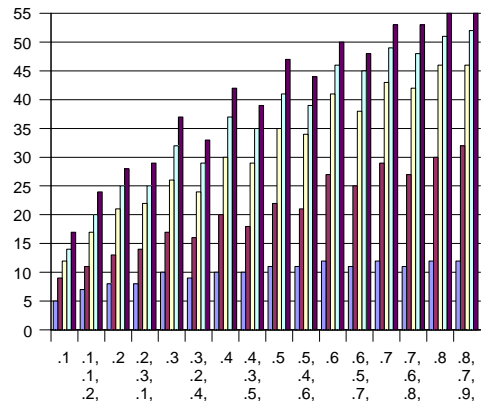
Number of long sim. talks in 5's (+12)
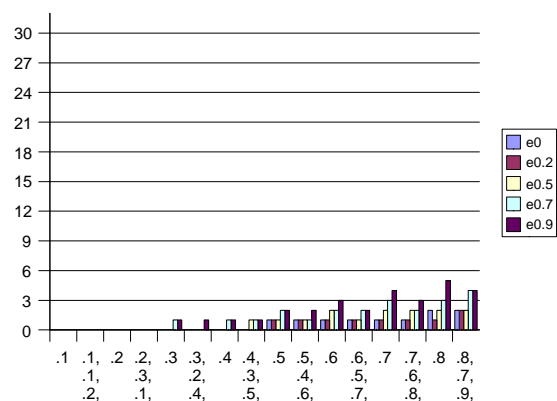
Number of utterances in 5's (+12)
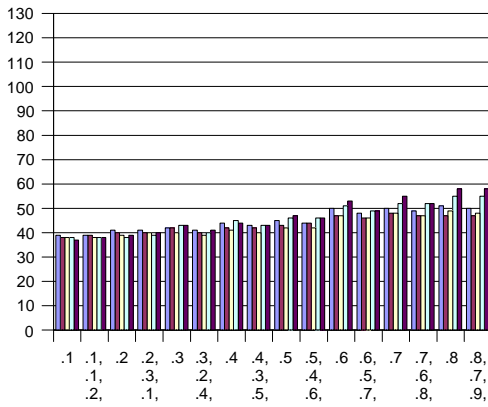
Continuing utterances in 5's (+12)
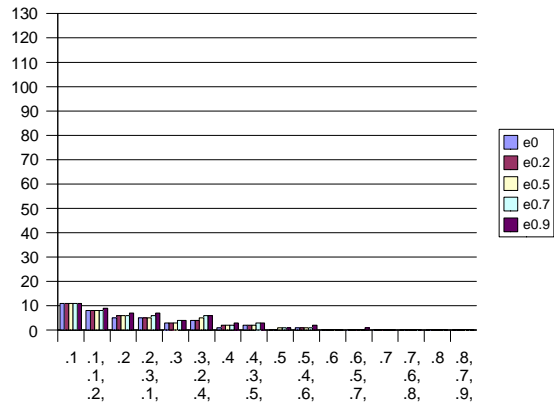
Number of silent gaps in 5's (+12)
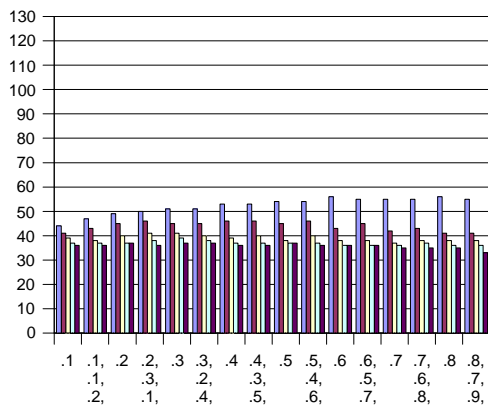
Number of overlaps in 5's (+12)
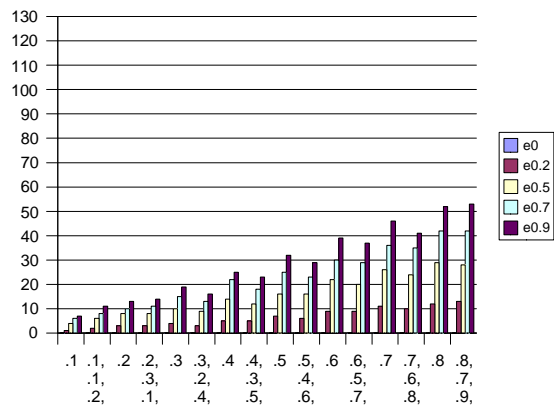
Number of sim. talks in 5's (+12)
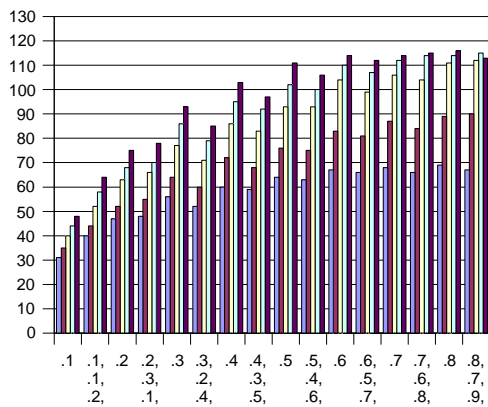
Number of single starts in 5's (+12)
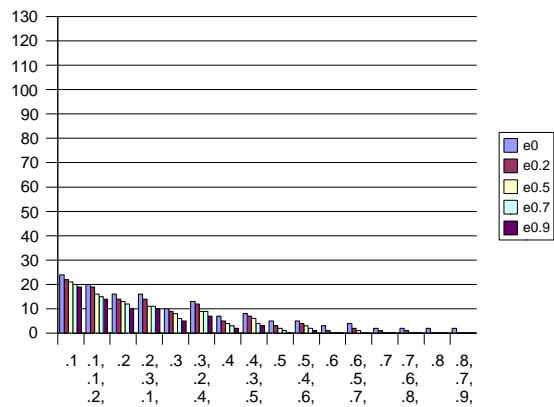
Number of sim.starts in 5's (+12)
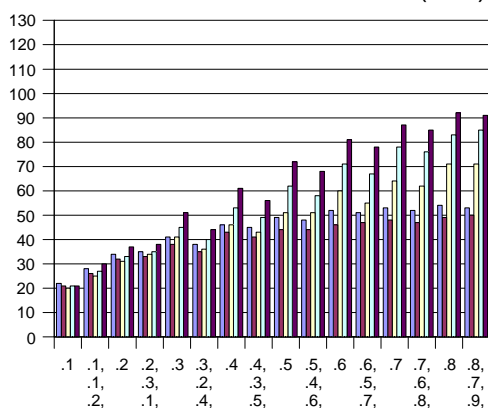
Number of middle starts in 5's (+12)
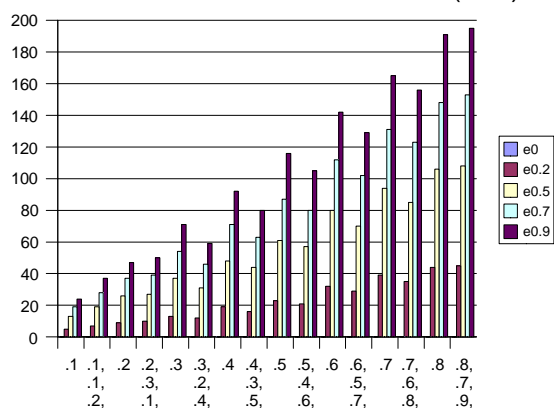
Number of utterances in 5's (e0.5)
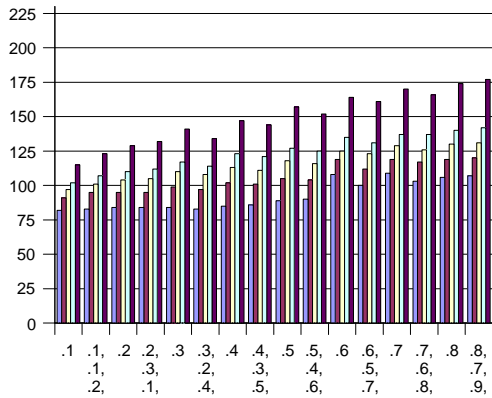
Continuing utterances in 5's (e0.5)
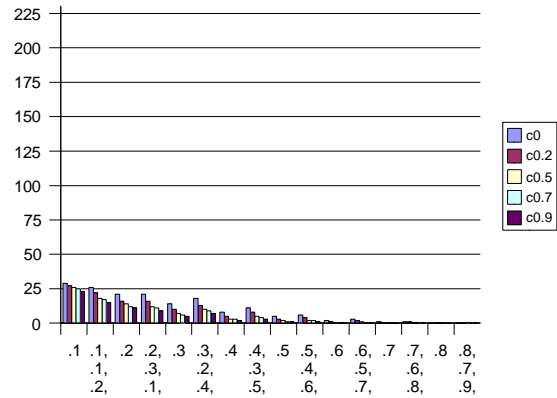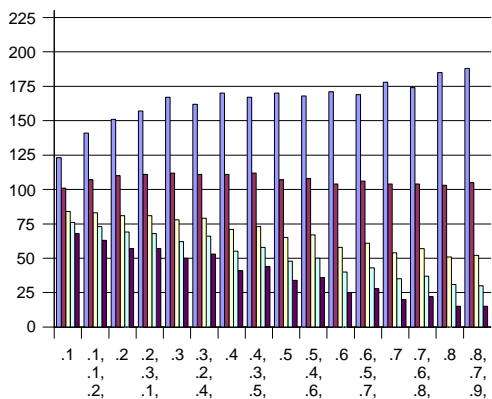
Number of silent gaps in 5's (e0.5)

Number of overlaps in 5's (e0.5)
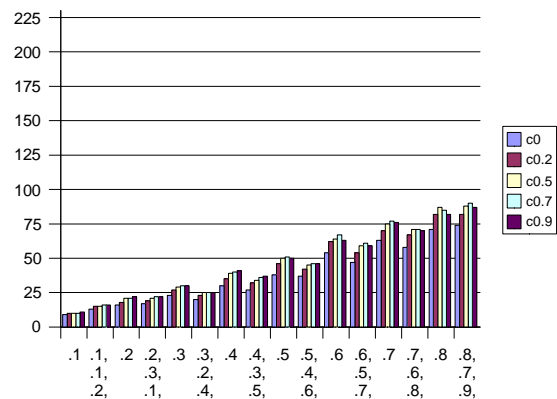
Number of sim. talks in 5's (e0.5)
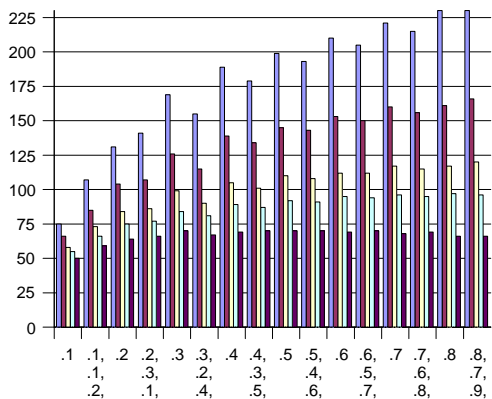
Number of single starts in 5's (e0.5)
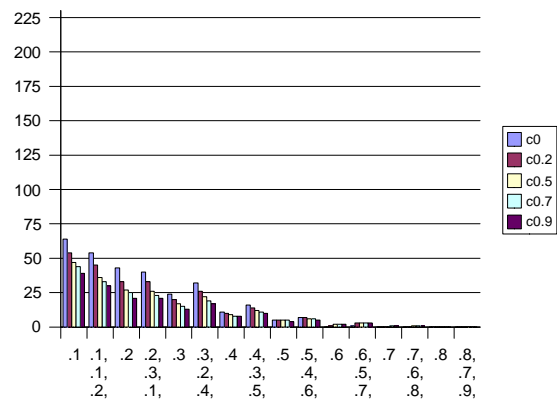
Number of sim.starts in 5's (e0.5)
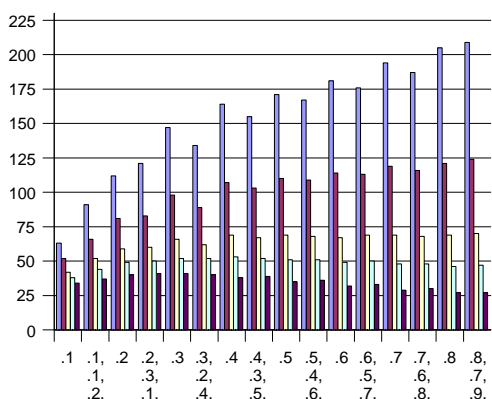
Number of middle starts in 5's (e0.5)
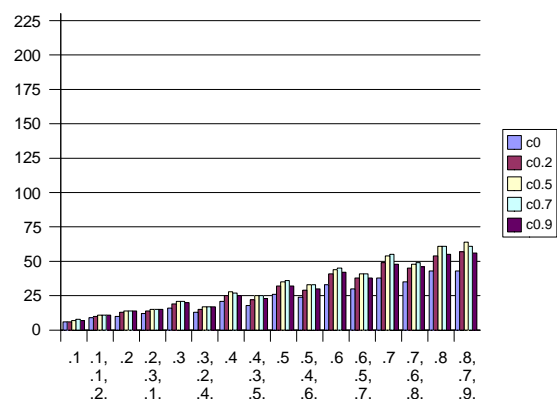
Number of false-starts in 5's (e0.5)
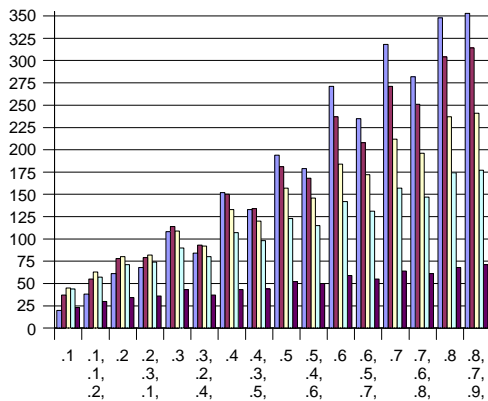
'Collective' false-starts in 5's (e0.5)
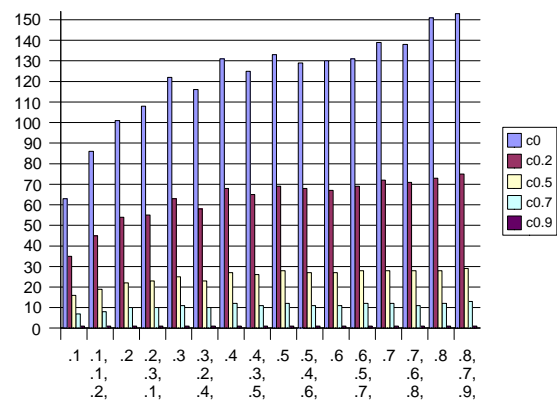
Incomplete utterances in 5's (e0.5)
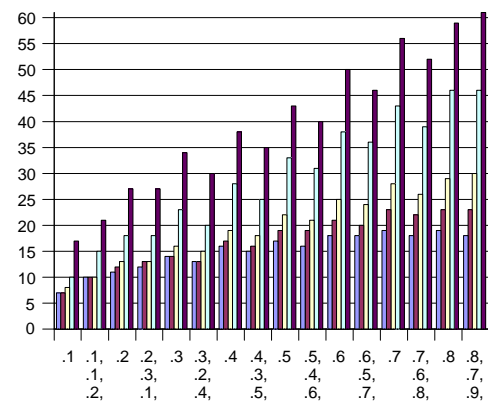
Number of long sim. talks in 5's (e0.5)
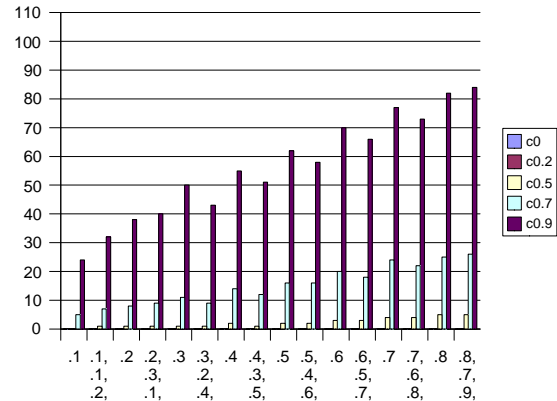
Number of false-starts in 5's (e,c0.5)

'Collective' false-starts in 5's (e,c0.5)

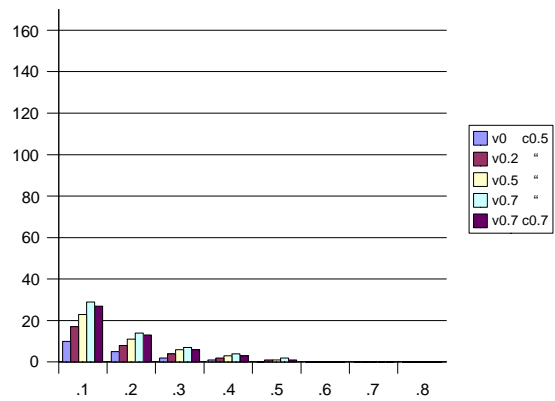### Incomplete utterances in 5's (e,c0.5)

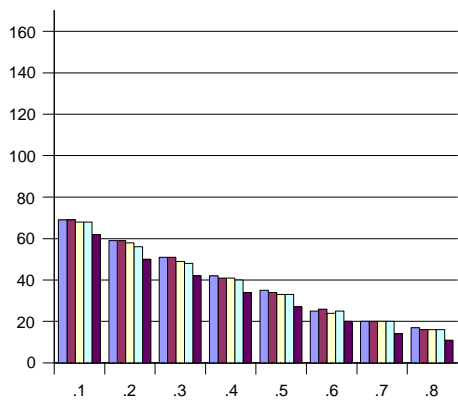### Number of long sim. talks in 5's (e,c0.5)
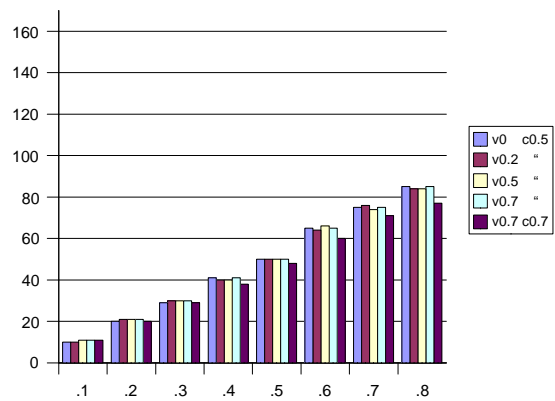
Number of utterances in 5's (e,c0.5)

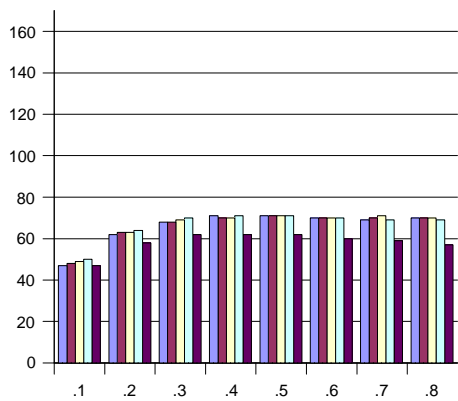## Continuing utterances in 5's (e,c0.5)
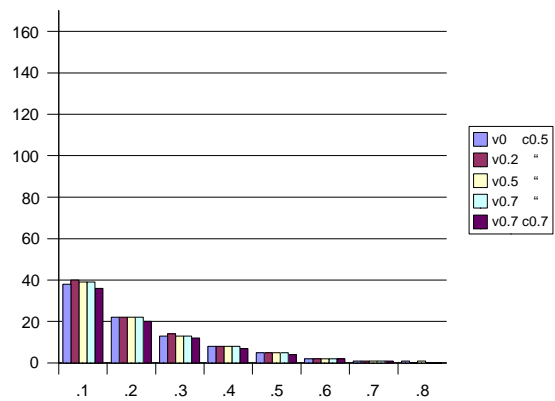

Number of silent gaps in 5's (e,c0.5)

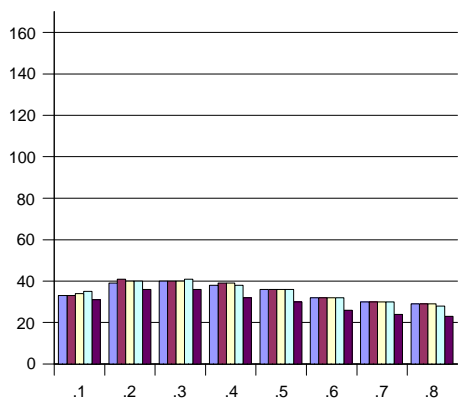

Number of overlaps in 5's (e,c0.5)

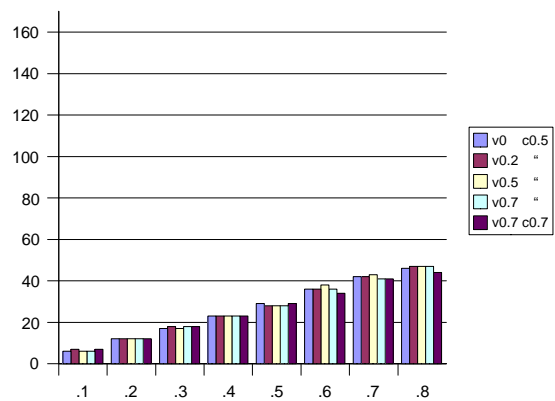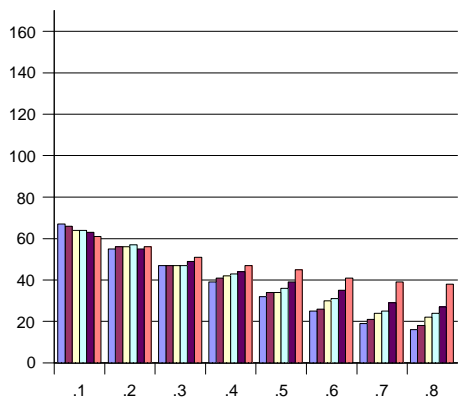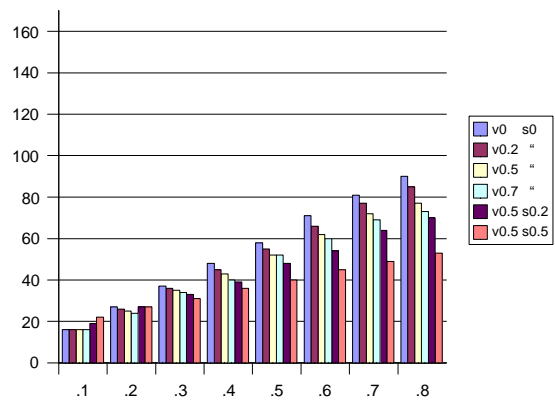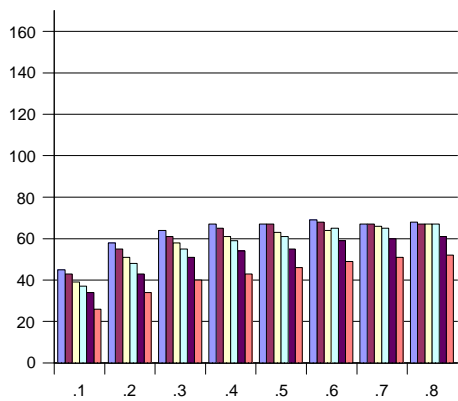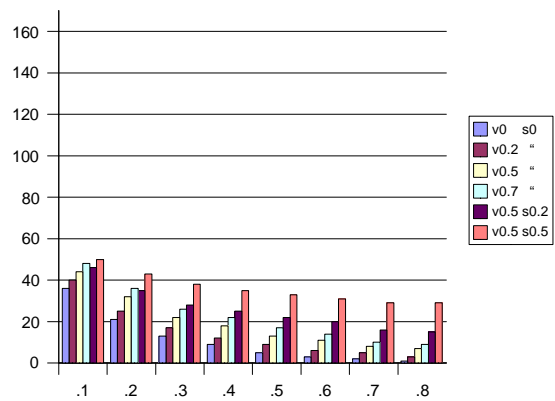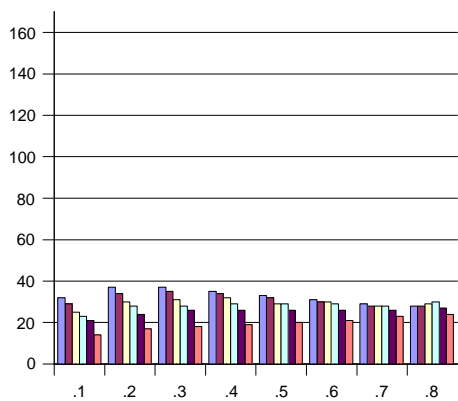## Number of sim. talks in 5's (e,c0.5)
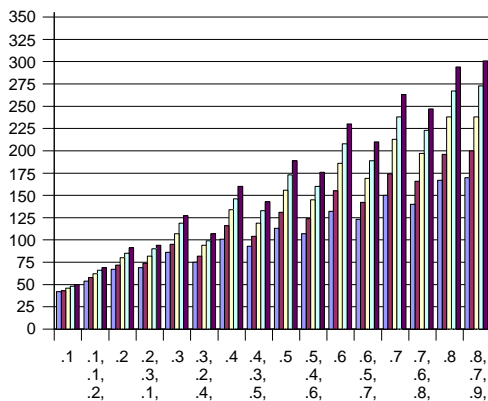
Number of single starts in 5's (e,c0.5)
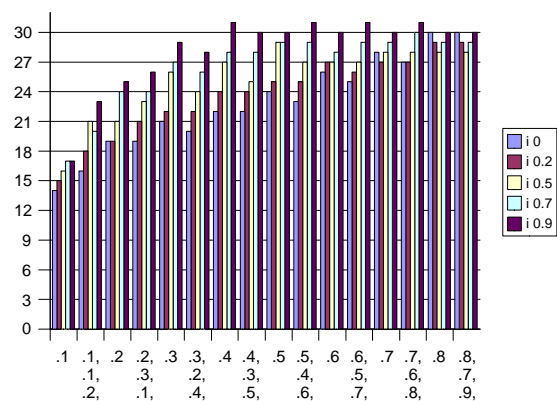
Number of sim.starts in 5's (e,c0.5)
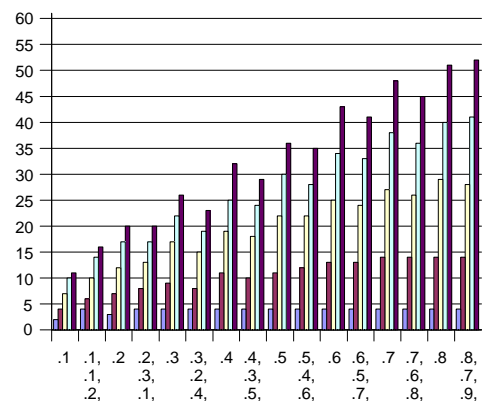
Number of middle starts in 5's (e,c0.5)
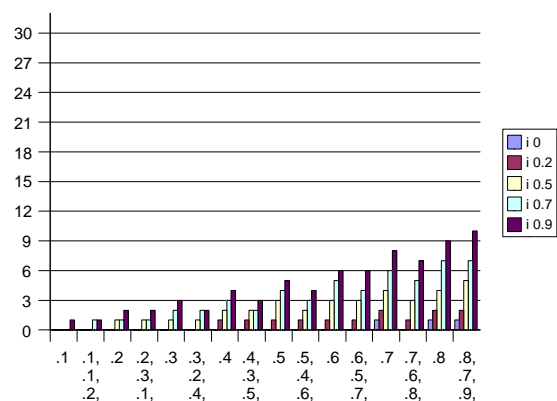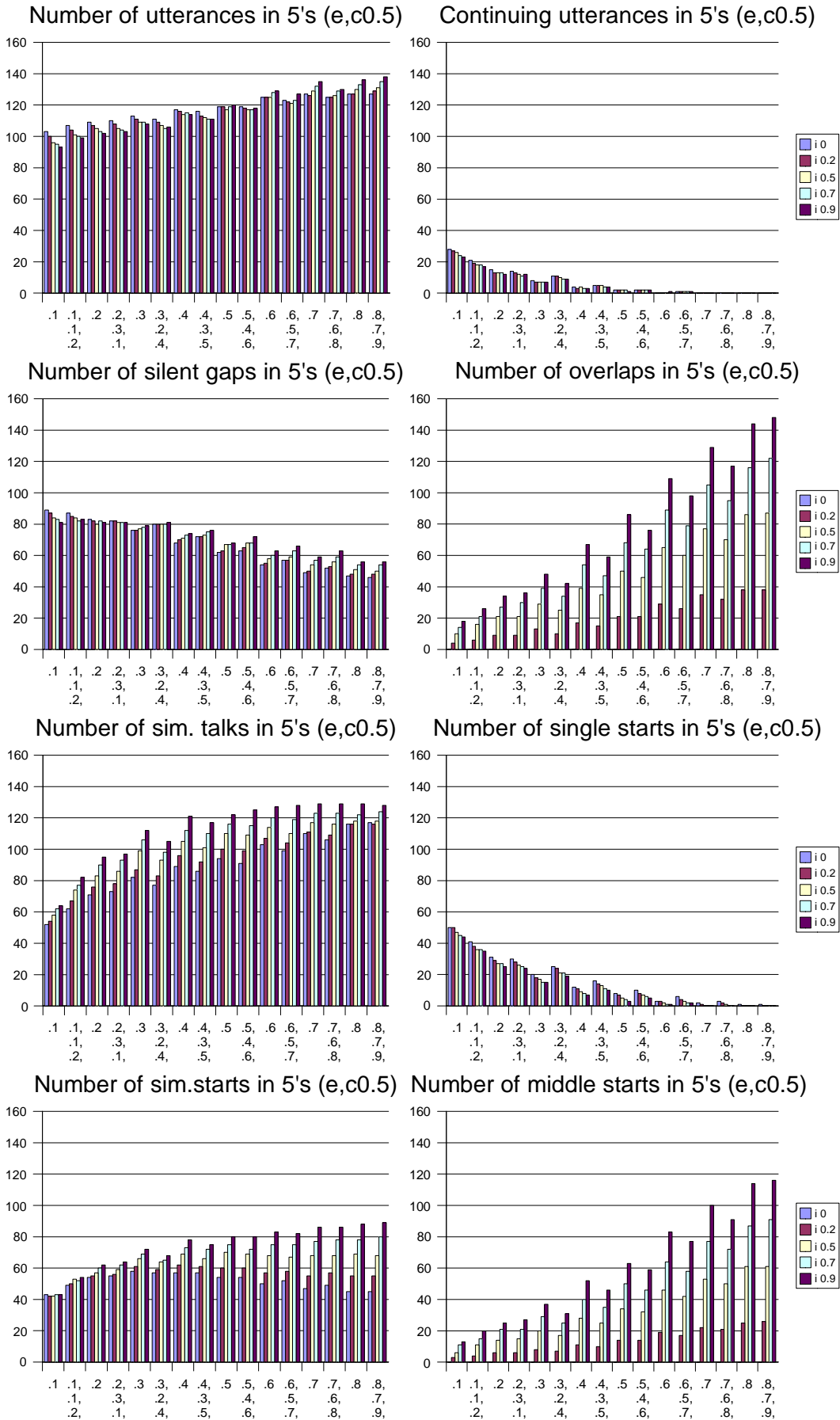
The following set of charts presents the measures counted from the last four evaluations: two for the second model (with more elaborate procedures of simultaneous talk to prevent 'fickle' agents stopping forthwith whenever there is simultaneous talk), and one for each of the other extensions, including 'types' of TRPs giving different restrictions and obligations to turn-taking, and with hesitations. There comes first the 'time' measures for all these evaluations, then the 'number' measures in the same organization of one-and-a-half pages per evaluation described earlier.

The two evaluations of the second model, beyond varying in TALKATIVENESS at the horizontal axis, range in CONFIDENCE and VERBOSITY (see the lateral legends) which are the relevant attributes of this model. The last evaluations of the third and fourth (extended) models range in VERBOSITY as the main parameter (changing the frequency of 'More' utterances that allow a speaker to continue talking in spite of others wanting to), and in two settings of SELECTIVITY for the third model and HESITATION for the fourth; check the lateral legends: 's0', 's0.2', 'h0.2', 'h0.5'. SELECTIVITY means that speaking agents have a certain likelihood (here 0.2 and 0.5) in deciding to 'select' someone to speak next instead of leaving the turn-taking free; this apparently made the discussions much more 'productive', with more single talk, less simultaneous talk, and so forth. HESITATION for the fourth model means that all agents hesitate in a given frequency (here 0.2 and 0.5) in the middle of their talk and when selected to speak, which did not seem to make much of a change in the stats as compared to the third model.

Finally, the last three evaluations were not evaluated with different TALKATIVENESS parameters for the group, only with the same value for all, so they omit the extra set of different parameters in the horizontal axis, ranging only in the eight basic settings of the attribute for all agents. These are indicated in the horizontal axis as '.1', '.2', '.3', etc.

Time of single talk in 5's (model 2)

Time of total silence in 5's (e0.5)

Time of silent gaps in 5's (e0.5)

Time of gaps+pauses (no talk) in 5's

Time of simult. talk in 5's (e0.5)

Time of sim.talk+overlaps in 5's (e0.5)

Time of backchannels in 5's (e0.5)

Time of total feedback in 5's (e0.5)

Time of single talk in 5's (model 2)

Time of total silence in 5's (e0.5)

Time of silent gaps in 5's (e0.5)

Time of gaps+pauses (no talk) in 5's

Time of simult. talk in 5's (e0.5)

Time of sim.talk+overlaps in 5's (e0.5)

Time of backchannels in 5's (e0.5)

Time of total feedback in 5's (e0.5)

Time of single talk in 5's (model 3)

Time of total silence in 5's (e,c0.5)

Time of silent gaps in 5's (e,c0.5)

Time of gaps+pauses (no talk) in 5's

Time of simult. talk in 5's (e,c0.5)

Time of sim.talk+overlaps in 5's (e,c0.5)

Time of backchannels in 5's (e,c0.5)

Time of total feedback in 5's (e,c0.5)

Time of single talk in 5's (model 4)

Time of total silence in 5's (s0,e/c0.5)

Time of silent gaps in 5's (s0,e/c0.5)

Time of gaps+pauses (no talk) in 5's

Time of simult. talk in 5's (s0,e/c0.5)

Time of sim.talk+overlaps in 5's (s0,..0.5)

Time of backchannels in 5's (s0,..0.5)

Time of total feedback in 5's (s0,e/c0.5)

Number of utterances in 5's (model 2)

Continuing utterances in 5's (e0.5)

Number of silent gaps in 5's (e0.5)

Number of overlaps in 5's (e0.5)

Number of sim. talks in 5's (e0.5)

Number of single starts in 5's (e0.5)

Number of sim.starts in 5's (e0.5)

Number of middle starts in 5's (e0.5)

Number of false-starts in 5's(model 2)

## 'Collective' false-starts in 5's (e0.5)

Incomplete utterances in 5's (e0.5)

Number of long sim.talks in 5's (e0.5)

Number of false-starts in 5's (e0.5)

'Collective' false-starts in 5's (e0.5)

## Incomplete utterances in 5's (e0.5)

## Number of long sim.talks in 5's (e0.5)

Number of utterances in 5's (model 2)

Continuing utterances in 5's (e0.5)

Number of silent gaps in 5's (e0.5)

Number of overlaps in 5's (e0.5)

Number of sim.talks in 5's (e0.5)

Number of single starts in 5's (e0.5)

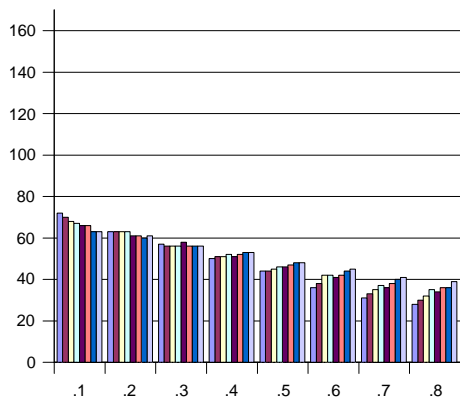Number of sim. starts in 5's (e0.5)

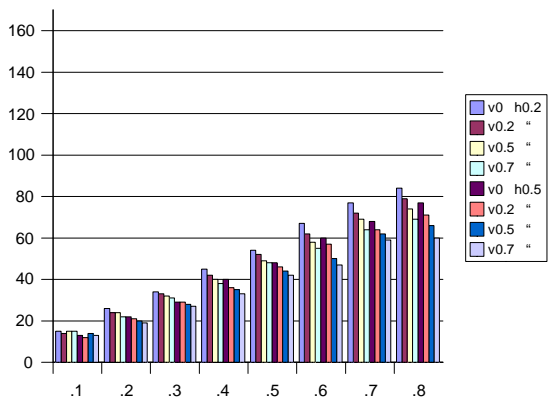Number of middle starts in 5's (e0.5)

Number of utterances in 5's (model 3)

Continuing utterances in 5's (e,c0.5)

Number of silent gaps in 5's (e,c0.5)

Number of overlaps in 5's (e,c0.5)

Number of sim. talks in 5's (e,c0.5)

Number of single starts in 5's (e,c0.5)

Number of sim.starts in 5's (e,c0.5)

Number of middle starts in 5's (e,c0.5)

Number of false-starts in 5's (model 3)

'Collective' false-starts in 5's (e,c0.5)
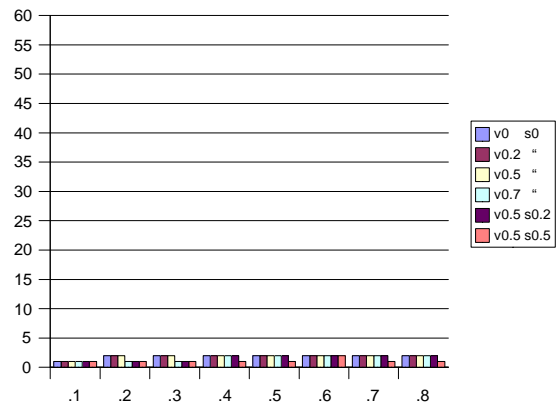
Incomplete utterances in 5's (e,c0.5)
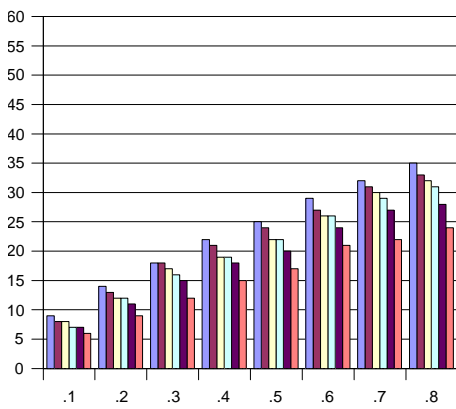
Number of long sim.talks in 5's (e,c0.5)
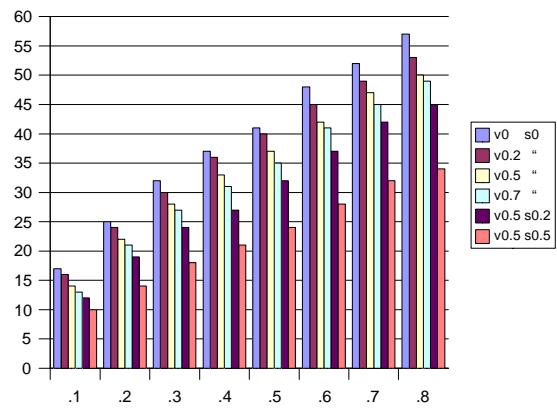
Number of false-starts in 5's (model 4)
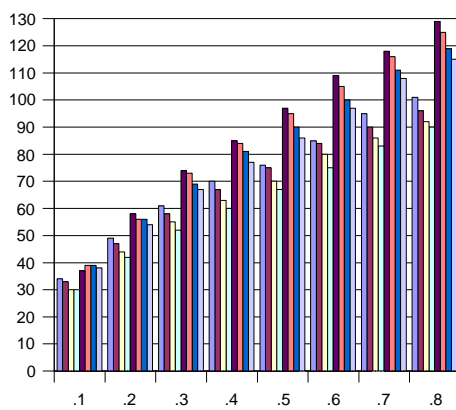
'Collective' false-starts in 5's (s0,e/c0.5)
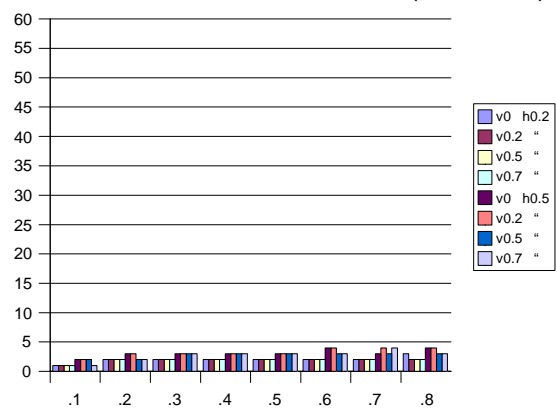
Incomplete utterances in 5's (s0,..0.5)
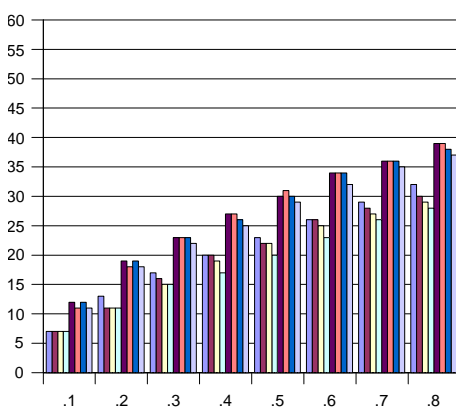
Number of long sim.talks in 5's (s0,e/c0.5)
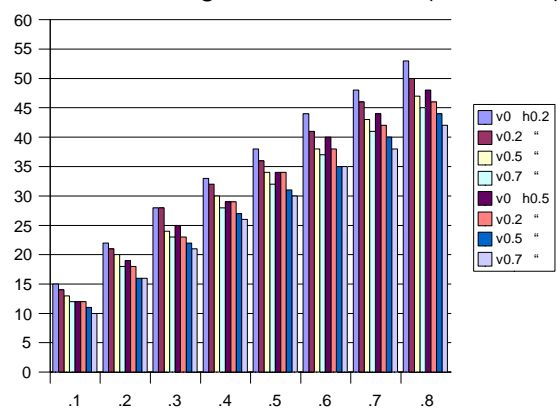
Number of utterances in 5's (model 4)
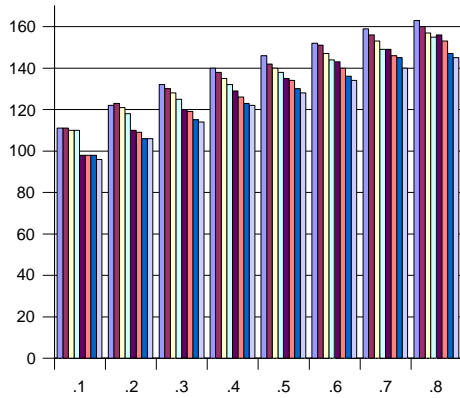
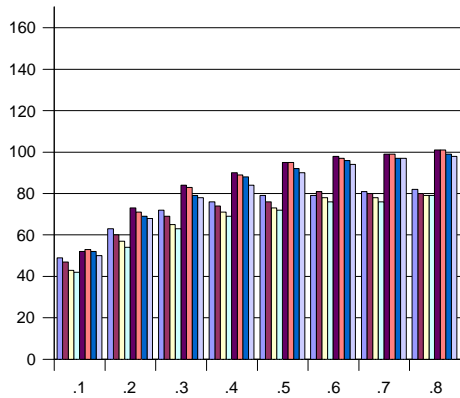## Continuing utterances in 5's (s0,e/c0.5)
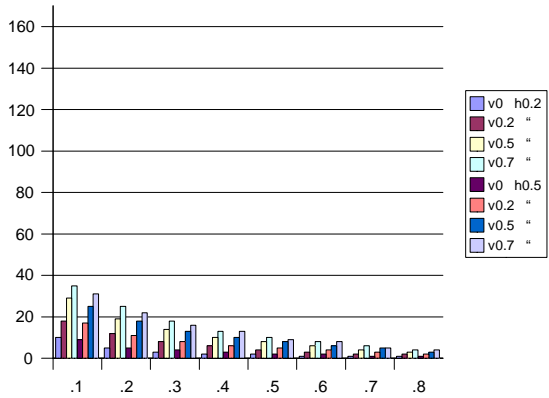
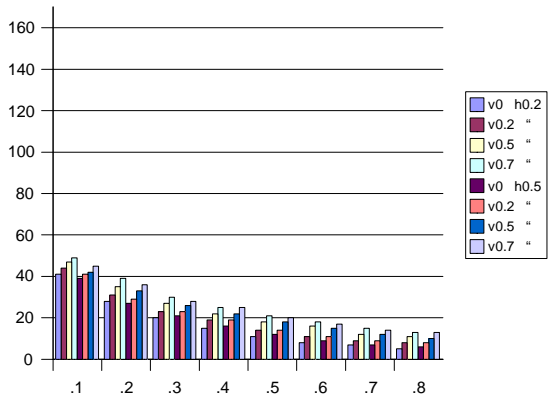## Number of silent gaps in 5's (s0,e/c0.5)

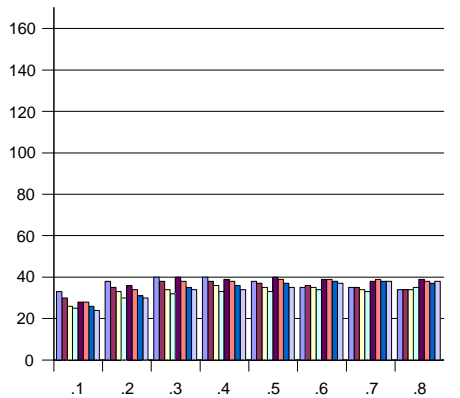## Number of overlaps in 5's (s0,e/c0.5)

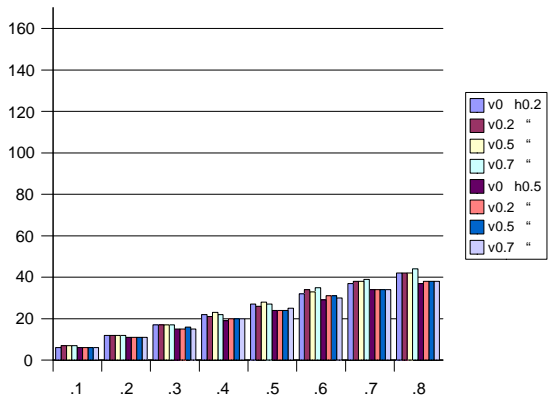
Number of sim. talks in 5's (s0,e/c0.5)

Number of single starts in 5's (s0,e/c0.5)

Number of sim.starts in 5's (s0,e/c0.5)

Number of middle starts in 5's (s0,e/c0.5)

Bibliography

Aleguire, D. (1978). Interruptions as turn-taking. In *Ninth World Congress of Sociology*, Uppsala, Sweden.

Argyle, M. (1972). Non-verbal communication in human social interaction. In R. A. Hinde (Ed.), *Non-verbal Communication*, London: Cambridge University Press, pp. 243–269.

Argyle, M. and M. Cook (1976). *Gaze and mutual gaze*. Cambridge, USA: Cambridge University Press.

Bales, R. F. (1950). *Interaction process analysis*. Cambridge, USA: Addison Wesley.

Beattie, G. W. (1978). Floor apportionment and gaze in conversational dyads. *British Journal of Social and Clinical Psychology 17*, 7–16.

Beattie, G. W. (1979). Contextual constraints on the floor-apportionment function of speaker gaze in dyadic conversation. *British Journal of Social and Clinical Psychology 18*, 391–392.

Beattie, G. W. (1981). The regulation of speaker turns in face-to-face conversation: some implications for conversation in sound-only communication channels. *Semiotica 34*, 55–70.

Bell, L. and J. Gustafson (2000). Positive and negative user feedback in a spoken dialogue corpus. In *Proceedings of the 6th International Conference on Spoken Language Processing – ICSLP'00*.

Birdwhistell, R. L. (1952). *Introduction to kinesics: an annotation system for the analysis of body motion and gestures*. Louisville, USA: University of Louisville Press.

Bull, M. and M. Aylett (1998). An analysis of the timing of turn-taking in a corpus

of goal-oriented dialogue. In *Proceedings of the 5th International Conference on Spoken Language Processing – ICSLP'98*, Volume 4, Sydney, Australia, pp. 1175–1178.

Carletta, J., R. Caley, and S. Isard (1993). A collection of self-repairs from the map task corpus. Technical Report HCRC/TR-47, HCRC, Edinburgh University, Edinburgh, UK. (citeseer.ist.psu.edu/carletta93collection.html).

Carlson, R., J. Hirschberg, and M. Swerts (2005). Cues to upcoming swedish prosodic boundaries: Subjective judgment studies and acoustic correlates. *Speech Communication 46*, 326–333.

Caspers, J. (1998). Who's next? The melodic marking of question vs. continuation in dutch. *Language and Speech 41*(3–4), 375–398.

Cassell, J., O. E. Torres, and S. Prevost (1999). Turn taking vs. discourse structure: How best to model multimodal conversation. In Y. Wilks (Ed.), *Machine conversations*, The Hague, Netherlands: Kluwer, pp. 143–154.

Cassell, J., C. S. Y. Nakano, T. W. Bickmore, and C. Rich (2001). Non-verbal cues for discourse structure. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, Toulouse, France, pp. 106–115.

Clancy, P., S. Thompson, R. Suzuki, and H. Tao (1996). The conversational use of reactive tokens in english, japanese, and mandarin. *Journal of Pragmatics 26*(3), 355–387.

Clark, H. H. (1992). *Arenas of Language Use*. Chicago, USA: University of Chicago Press.

Clark, H. H. and S. Brennan (1991). Grounding in communication. In L. B. Resnick, R. M. Levine, and S. D. Teasley (Eds.), *Perspectives on socially shared cognition*, Washington, USA: APA, pp. 127–149.

Clark, H. H. and E. F. Schaefer (1987). Collaborating on contributions to conversation. *Language and Cognitive Processes 2*(1), 19–41.

Clark, H. H. and E. F. Schaefer (1989). Contributing to discourse. *Cognitive Science 13*, 259–284.

Clark, H. H. and D. Wilkes-Gibbs (1986). Referring as a collaborative process. *Cognition 22*, 1–39.

Coates, J. (1989). Gossip revisited: Language in all female groups. In J. Coates and D. Cameron (Eds.), *Women in their speech communities*, New York, USA: Longman, pp. 94–122.

Cutler, A. and M. Pearson (1986). On the analysis of prosodic turn-taking cues. In C. Johns-Lewis (Ed.), *Intonation in discourse*, London, England: Croom Helm, pp. 139–155.

Dabbs, J. M. and R. B. Ruback (1987). Dimensions of group process: amount and structure of vocal interaction. *Advances in Experimental Social Psychology 20*, 123–169.

Derriks, B. and D. Willems (1998). Negative feedback in information dialogues: identification, classification and problem-solving procedures. *International Journal of Human-Computer Studies 48*, 577–604.

Donaldson, T. and R. Cohen (1996). Turn-taking in discourse and its application to the design of intelligent agents. In M. Tambe and P. Gmytrasiewicz (Eds.), *Working Notes of the AAAI-96 Workshop on Agent Modeling*, Portland, USA, pp. 17–23. (citeseer.ist.psu.edu/33078.html).

Duncan, S. D. (1972). Some signals and rules for taking speaking turns in conversation. *Journal of Personality and Social Psychology 23*(2), 283–293.

Duncan, S. D. (1974). On the structure of speaker-auditor interaction during speaking turns. *Language in Society 3*(2), 161–180.

Duncan, S. D. and G. Niederehe (1974). On signalling that it's your turn to speak. *Journal or experimental social psychology 10*, 234–247.

Eco, U. (1986). Eine palette von grautönen. *Die Zeit ??*, 51–51.

Edelsky, C. (1981). Who's got the floor? *Language and Society 10*(3), 383–421.

Edlund, J., M. Heldner, and J. Gustafson (2005). Utterance segmentation and turn-taking in spoken dialogue systems. In B. Fisseni, H. C. Schmitz, B. Schröder, and P. Wagner (Eds.), *Sprachtechnologie, mobile Kommunikation und linguistische Ressourcen*, pp. 576–587. Peter Lang. (citeseer.ist.psu.edu/edlund05utterance.html).

Edlund, J. and M. Nordstrand (2002). Turn-taking gestures and hour-glasses in a multi-modal dialogue system. In *Proceedings of the ISCA Workshop on Multi-Modal Dialogue in Mobile Environments*, Kloster Iresee, Germany.

Ervin-Tripp, S. (1979). Children's verbal turn-taking. In E. Ochs and B. B. Schieffelin (Eds.), *Developmental Pragmatics*, New York, USA: Academic Press, pp. 391–414.

Fay, N. (2000). *The effect of group size upon influence and process in group communication*. Ph. D. thesis, University of Glasgow, Glasgow, Scotland.

Fay, N., S. Garrod, and J. Carletta (2000). Group discussion as interactive dialogue or serial monologue: the influence of group size. *Psychological Science 11*(6), 487–492.

Ferguson, N. (1977). Simultaneous speech, interruptions and dominance. *British journal of social and clinical psychology 16*, 295–302.

Ferrer, L., E. Shriberg, and A. Stolcke (2002). Is the speaker done yet? Faster and more accurate end-of-utterance detection using prosody. In *Proceedings of the 7th International Conference on Spoken Language Processing – ICSLP'02*, Volume 3, Denver, USA, pp. 2061–2064.

Ferrer, L., E. Shriberg, and A. Stolcke (2003). A prosody-based approach to end-of-utterance detection that does not require speech recognition. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing – ICASSP'03*, Volume 1, Hong Kong, China, pp. 608–611.

Ford, C. and S. Thompson (1996). Interactional units in conversation: Syntactic, intonational, and pragmatical resources for the management of turns. In E. Ochs, E. Schegloff, and S. Thompson (Eds.), *Interaction and Grammar*, Cambridge, USA: Cambridge University Press, pp. 134–184.

Freedman, N. and S. P. Hoffman (1967). Kinetic behaviour in altered clinical states. Approach to objective analysis of motor behaviour during clinical interviews. *Perceptual and Motor Skills 24*, 527–539.

Fries, C. (1952). *The structure of English: an introduction to the construction of English sentences*. New York, USA: Harcourt Brace.

Furo, H. (2001). *Turn-taking in English and Japanese: Projectability in Grammar, Intonation, and Semantics*. New York, USA: Routledge.

Goffman, E. (1964). The neglected situation. *American Anthropologist 66*, 133–136.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. New York, USA: Academic Press.

Grosjean, F. (1983). How long is the sentence? Prediction and prosody in the on-line processing of language. *Linguistics 21*, 501–529.

Grosz, B. J. and C. L. Sidner (1986). Attention, intentions, and the structure of the discourse. *Computational Linguistics 12*, 175–204.

Harrigan, J. A. and J. J. Steffen (1982). Gaze as a turn-exchange signal in group conversations. *Journal of Personality and Social Psychology 22*, 167–168.

Hayashi, R. and S. Iwasaki (1998). Cognition, empathy, and interaction: Floor management of english and japanese conversation. *Language in Society 27*(3), 393–397.

Hulstijn, J. and G. Vreeswijk (2003). Turntaking: A case for agent-based programming. Technical Report UU-CS-2003-045, IICS, Utrecht University, Utrecht, Netherlands. (citeseer.ist.psu.edu/hulstijn03turntaking.html).

Iwasaki, S. and P. Horie (1998). The 'northridge earthquake' conversations: conversational patterns in japanese and thai and their cultural significance. *Discourse and Society 9*(4), 501–529.

Janin, A., D. Baron, J. Edwards, D. Ellis, D. Gelbart, N. Morgan, B. Peskin, T. Pfau, E. Shriberg, A. Stolcke, and C. Wooters (2003). The ICSI meeting corpus. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing – ICASSP'03*, Hong Kong, China, pp. 364–367.

Jefferson, G. (1973). A case of precision timing in ordinary conversation: overlapped tag-positioned address terms in closing sequences. *Semiotica 9*, 47–96.

Kalma, A. (1992). Gazing in triads – a powerful signal in floor apportionment. *British Journal of Social Psychology 31*, 21–39.

Kamp, J. A. W. (1981). A theory of truth and semantic representation. In J. Groenendijk, T. Janssen, and M. Stockhof (Eds.), *Formal methods in the study of language, Part 1*, Amsterdam, Netherlands: Mathematical Centre Tracts, pp. 177–321.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica 26*, 22–63.

Kendon, A. (1970). Movement coordination in social interaction: some examples described. *Acta Psychologica 32*, 101–105.

Kendon, A. (1972). Some relationships between body motion and speech. In A. W.

Siegman and B. Pope (Eds.), *Studies in Dyadic Communication*, Elmsford, USA: Pergammon Press, pp. 177–210.

Kim, K. H. (1999). Phrasal unit boundaries and organization of turns and sequences in korean conversation. *Human Studies 22*(2–4), 425–446.

Kirchhoff, K. and M. Ostendorf (2003). Directions for multi-party human-computer interaction research. In *Proceedings of the HLT-NAACL 2003 Workshop on Research Directions in Dialogue Processing*, Edmonton, Canada, pp. 7–9. (citeseer.ist.psu.edu/kirchhoff03directions.html).

Levelt, W. J. M. (1989). *Speaking: from Intention to Articulation*. Boston, USA: MIT Press.

Lewis, D. K. (1969). *Convention: A Philosophical Study*. Cambridge, USA: Harvard University Press.

Litman, D. J. and J. F. Allen (1987). A plan recognition model for subdialogues in conversation. *Cognitive Science 11*, 163–200.

Lobb, M. O. H. (1982). Seating arrangement as a predictor of small group interaction. *Journal of Advanced Nursing 7*, 163–166.

McGrath, J. E. (1984). *Groups: Interaction and Performance*. Englewood Cliffs, USA: Prentice-Hall.

Meltzer, L., W. Morris, and D. Hayes (1971). Interruption outcomes and vocal amplitude: Explorations in social psycho-physics. *Journal of Personality and Social Psychology 18*, 392–402.

Nakajima, S. and J. F. Allen (1993). A study on prosody and discourse structure in cooperative dialogues. *Phonetica 50*(3), 197–210.

Novick, D. G., B. Hansen, and K. Ward (1996). Coordinating turn-taking with gaze. In *Proceedings of the 4th International Conference on Spoken Language Processing – ICSLP'96*, Volume 3, Philadelphia, USA, pp. 1888–1891.

O'Connell, D. C., S. Kowal, and E. Kaltenbacher (1990). Turn-taking: A critical analysis of the research tradition. *Journal of Psycholinguistic research 19*(6), 345–373.

Oreström, B. (1983). *Turn-taking in English Conversation.* Lund Studies in English. Malmö, Sweden: LiberFörlag Lund.

Padilha, E. G. and J. Carletta (2002). A simulation of small group discussion. In *Proceedings of the Sixth workshop on the semantics and pragmatics of dialogue – EDILOG'02*, pp. 117–124. (http://www.ltg.ed.ac.uk/edilog/papers/117.pdf).

Padilha, E. G. and J. Carletta (2003). Nonverbal behaviours improving a simulation of small group discussion. In *Proceedings of the 1st Nordic Symposium on Multimodal Communication*, Helsingør, Denmark, pp. 93–105.

Parker, K. C. H. (1988). Speaking turns in small group interaction: A context-sensitive event sequence model. *Journal of Personality and Social Psychology 54*(6), 965–971.

Polanyi, L. and R. Scha (1985). A discourse model for natural language. In L. Polanyi (Ed.), *The structure of discourse*, Norwood, USA: Ablex.

Power, R. J. D. and F. D. Martello (1986). Some criticisms of sacks, schegloff and jefferson on turn taking. *Semiotica 58*(1), 29–40.

Psathas, G. (1995). *Conversation Analysis: The Study of Talk-in-Interaction.* London, UK: Sage Publications.

Reisman, K. (1977). Contrapuntal conversations in an antiguan village. In R. Bauman and J. Scherzer (Eds.), *Explorations in the ethnography of speaking*, Cambridge, USA: Cambridge University Press, pp. 110–124.

Rosenfeld, H. M. (1977). Conversational control functions of nonverbal behavior. In A. W. Siegman and S. Feldstein (Eds.), *Nonverbal Behavior and Communication*, Hillsdale, USA: Lawrence Erlbaum Associates, pp. 291–328.

Rutter, D. R., G. M. Stephenson, K. Ayling, and P. A. White (1978). The timing of looks in dyadic conversation. *British Journal of Social and Clinical Psychology 17*, 17–21.

Sacks, H. (1974). An analysis of the course of a joke's telling in conversation. In R. Baumann and J. Sherzer (Eds.), *Explorations in the ethnography of speaking*, Cambridge, USA: Cambridge University Press.

Sacks, H. and E. A. Schegloff (1979). Two preferences in the organization of refer-

ence to persons in conversation and their interaction. In G. Psathas (Ed.), *Everyday Language: Studies in Ethnomethodology*, New York, USA: Irvington Publishers.

Sacks, H., E. A. Schegloff, and G. Jefferson (1974). A simplest systematics for the organization of turn taking for conversation. *Language 50*(4), 696–735.

Schaffer, D. (1983). The role of intonation as a cue to turn-taking in conversation. *Journal of Phonetics 11*, 243–259.

Schegloff, E. A. (1980). Preliminaries to preliminaries: 'can I ask you a question'. *Social Inquiry 50*, 104–152.

Schegloff, E. A. (1982). Discourse as an interactional achievement: some uses of 'uh huh' and other things that come between sentences. In D. Tannen (Ed.), *Analyzing Discourse: Text and Talk*, Washington, USA: Georgetown Univ. Press, pp. 71–93.

Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in Society 29*, 1–63.

Schegloff, E. A., G. Jefferson, and H. Sacks (1977). The preference for self-correction in the organization of repair in conversation. *Linguistcs 53*, 361–382.

Schegloff, E. A. and H. Sacks (1973). Opening up closings. *Semiotica 8*, 283–327.

Schenkein, J. N. (Ed.) (1978). *Studies in the Organization of Conversational Interaction*. New York, USA: Academic Press.

Schober, M. F. and H. H. Clark (1989). Understanding by addressees and overhearers. *Cognitive Psychology 21*, 211–232.

Searle, J. R. (1992). Conversation. In H. Parret and J. Verschueren (Eds.), *(On) Searle on Conversation*, Amsterdam, Netherlands: John Benjameens Publishing.

Selting, M. (1996). On the interplay of syntax and prosody in the constitution of turn-constructional units and turns in conversation. *Pragmatics 6*, 357–388.

Selting, M. (2000). The construction of units in conversational talk. *Language in Society 29*, 477–517.

Sidnell, J. (2001). Conversational turn-taking in a caribbean english creole. *Journal of Pragmatics 33*(8), 1263–1290.

Spelke, E. S., W. Hirst, and U. Neisser (1976). Skills of divided attention. *Cognition 4*, 215–230.

Stasser, G. and L. A. Taylor (1991). Speaking turns in face-to-face discussions. *Journal of Personality and Social Psychology 60*(5), 675–684.

Steinzor, B. (1950). The spatial factor in face to face discussion groups. *Journal of Abnormal and Social Psychology 45*, 552–555.

Stephan, F. F. and E. G. Mishler (1952). The distribution of participation in small groups: An exponential approximation. *American Sociological Review 17*, 598–608.

Svartvik, J. and R. Quirk (1980). *A corpus of English conversation*. Lund, Sweden: Gleerup.

Takagi, K. and S. Itahashi (1996). Segmentation of spoken dialogue by interjection, disfluent utterances and pauses. In *Proceedings of the 4th International Conference on Spoken Language Processing – ICSLP'96*, Volume 1, Philadelphia, USA, pp. 693–697.

Tanaka, H. (1999). *Turn-taking in Japanese conversation : a study in grammar and interaction*. Philadelphia, USA: J. Benjamins Publishing Co.

Tanaka, H. (2001). Adverbials for turn-projection in japanese: toward a demystification of the 'telepathic' mode of communication. *Language in Society 30*(4), 559–588.

Traum, D. R. and P. A. Heeman (1996). Utterance units in spoken dialogue. In *The ECAI Workshop on Dialogue Processing in Spoken Language Systems*, pp. 125–140. (citeseer.ist.psu.edu/traum97utterance.html).

Ward, N. and W. Tsukahara (2000). Prosodic features which cue back-channel responses in english and japanese. *Journal of Pragmatics 32*(8), 1177–1207.

Wennerstrom, A. and A. F. Siegel (2003). Keeping the floor in multiparty conversations: intonation, syntax, and pause. *Discourse Processes 36*(2), 77–107.

Wiemann, J. and M. Knapp (1975). Turn-taking in conversation. *Journal of Communication 25*, 75–92.

Yngve, V. (1970). On getting a word in edgewise. In *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, pp. 567–578.