DISCOURSE PROCESSES 9, 375–390 (1986)

#29

The Structure of Silence between Turns in Two-party Conversation

THOMAS P. WILSON
DON H. ZIMMERMAN
University of California, Santa Barbara

Turn taking is a fundamental structural feature of social interaction. Three major approaches to describing turn taking have emerged: stochastic, signaling, and sequential-production models. The first two treat silences between speakers as simple response latencies, whereas the third views silence as generated collaboratively by the parties to the conversation. The simple response-latency interpretation predicts a distribution of between-turn silences that declines monotonically with duration, whereas the sequential production model predicts a periodic pattern of peaks and valleys, with an overall decline in the heights of the peaks as duration increases. Analysis of the frequency distributions of durations of silences between speakers in two-party conversations finds the periodic structure predicted by the sequential-production model. The finding is interpreted as supporting a view of social interaction as a fundamentally collaborative activity.

Turn taking is one of the most salient features of social interaction. In the usual and expected course of events just one person talks at a time, silences are infrequent and brief, and the role of speaker changes hands frequently and easily. Of course, interaction does not always run smoothly, but simultaneous talking, long silences, and other disruptions are notable precisely because they stand out against the background of normally unproblematic speech exchange. For this reason, the process of turn taking can be exploited as a resource for pursuing individual ends, making social identities visible and effective in concrete form, and assembling social occasions and pursuing socially organized sequences of activity within them. Consequently, elucidation of the mechanisms of turn taking should further our understanding both of routine and strategic aspects of social interaction per se and also the bases of social organization.[1]

This research was supported by a grant from the Academic Senate of the University of California, Santa Barbara. We are indebted to R. A. Berk for invaluable comments throughout the course of this research, and to D. E. Rauma for help with the statistical analysis.

Correspondence and requests for reprints should be addressed to Thomas P. Wilson, Department of Sociology, University of California, Santa Barbara, CA 93106.

[1]Social interaction has long been a subject of research in its own right, and the vast literature is readily accessible and requires no citation here. For a review of research specifically concerned with turn taking, see Wilson, Wiemann, & Zimmerman (1984). A different tradition of research concerned with pauses but not turn taking per se is represented by the papers and references in Dechert

Turn taking is potentially problematic for those engaged in interaction when there is no prior arrangement providing for the order, lengths, or contents of turns, that is, in what Sacks, Schegloff, and Jefferson (1974) refer to as "conversation."[2] In the absence of such arrangements, the participants must manage turn taking themselves in the course of the interaction and consequently require some mechanism to accomplish the exchange of turns in an orderly manner.

Systematic concern with turn taking in conversation can be traced back at least to Goffman (1955), and in recent years a number of investigators have addressed the problem of how people manage the orderly change of turns in conversation. Three principle strategies have been proposed to describe the process of turn taking: stochastic models, signaling models, and sequential-production models. In current stochastic-modeling approaches, the data consist of each participant's sequence of silences and stretches of speech, and a conversation is assumed to be generated by a Markov or similar probabilistic process with several distinct "states," such as a particular person speaking, everyone silent, and the like (e.g., Brady, 1969; Cappella, 1979; Jaffe & Feldstein, 1970). A conversation is then described as a sequence of transitions of the system from one state to another with probabilities specified by the model or estimated from data, and change of speaker is treated as a probabilistic event associated with transitions between states of the system. In contrast, in a signaling model, change of speakers is assumed to be mediated by the exchange of discrete cues that can be identified independently of context but whose functions may vary depending on which other cues are displayed (e.g., Duncan, 1972; Duncan & Fiske, 1977; Yngve, 1970). Thus, in a signaling model, it is assumed that the current speaker offers the turn to another by means of a signal activated by one or more specific cues, analogous to the "over" convention used in some radio communication, and the recipient is free to accept or decline the offer. And, in sequential-production models, orderly change of speaker is seen as managed by the participants actively collaborating to construct the units of talk at the end of which speaker change is appropriate, as well as to effect the actual exchange of turns (e.g., Sacks et al, 1974).

We have reviewed the conceptual and theoretical issues raised by these approaches to research on turn taking in some detail elsewhere (Wilson, Wiemann, and Zimmerman, 1984). In this paper we report an empirical study designed to

and Raupach (1980). More recently, the interpenetration of social structure and social interaction has begun to receive systematic attention; see Giddens (1984), Goffman (1984), Maynard and Wilson (1980), and Wilson (1982) for one current line of thinking.

[2]As defined here, conversation is a particular form of speech exchange. In contrast, debates, formally chaired meetings, ceremonies, rituals, and the like, though forms of speech exchange, are not conversation for present purposes. For further discussion, see Sacks et al. (1974) and Wilson et al. (1984).

locate phenomena predicted by the sequential-production model but difficult t explain within the framework of stochastic or signaling models.[3]

HYPOTHESES

The concern here is with silences occurring at transition-relevance places, that is places in the flow of conversation at which change of speaker is appropriate, an in particular with the frequency distribution of the durations of silences betwee turns in a conversation. Although silence clearly is important in relation to tur taking, none of the current models is designed specifically to explain the dura- tions of silences between turns and, indeed, apart from the observation that shor silences tend to be more common than long ones, no detailed data on silence between turns have been available that could guide theory construction.[4] Nev- ertheless, these models have implications for the distribution of durations o silences between turns that are open to empirical investigation.

The status of silence between turns is quite different in stochastic and signal- ing models, on the one hand, and the sequential-production model, on the other In the former two, between-turn silence is treated essentially as a respons latency and thus is seen as arising from processes within the next speaker whereas in the latter, between-turn silence is seen as interactionally generated involving both the current and the next speaker.

Silence as Simple Response Latency

From a standard psychological point of view, it is natural to think of a perso starting to speak as a response to some preceding stimulus, and of silenc intervening between that stimulus and the talk terminating the silence as a re- sponse latency. Such a response latency can be construed as arising in variou ways, depending on the details of the particular psychological theory on

[3]Although current models of turn taking are each grounded in particular empirical observations they have not been subjected to empirical tests that go significantly beyond the phenomena motivat ing their original formulations. Thus one might well be uneasy that, though they perhaps summariz the facts that inspired them, these models may fail to capture the fundamental structure of th processes of turn taking. Consequently, we need to derive empirically testable hypotheses that ar more than minor variations on the original facts on which a particular model was based, and whic involve the central concepts of the model rather than details that can be discarded or altered withou threatening the model itself.

[4]Brady (1968) presents data suggesting that the frequency of silences decreases exponentiall with duration, but unfortunately he aggregated his data across conversations, which obscures th detailed structure of the distributions for individual conversations. Work following Jaffe and Felds tein (1970) uses a very large sampling interval (300 ms), which again obscures important detail However, these problems with the available data are immaterial because current models were no developed with the question of the duration of silence in mind.

chooses to employ. However, what is important for present purposes is not the variety of possible processes that could be involved in generating response latencies, but rather the implication that because short latencies are more common than long ones, the frequency of silences should decline monotonically with duration (see Figure 1).

Both stochastic and signaling models lead fairly directly to a response-latency conception of silences. In a stochastic model, when the system gets into a state in which no one is speaking, its movement to some other kind of state is probabilistic. Hence a long silence would be generated by the system moving between silent states for some number of cycles of the system, ending with a transition to a nonsilent state, the result being silence for the duration of the process. But the probability of a sequence of successive transitions from silent state declines with the length of the sequence, and so models of this sort provide a stochastic basis for response latencies. In the case of a signaling model, Duncan (1972) and Duncan and Fiske (1977) found no evidence that silence functions as a cue in relation to turn taking. Consequently, within the general framework of a signaling approach, the only natural alternative is to interpret silence as a response latency generated by some set of psychological processes within the next speaker. For, once the current speaker has offered the turn to another through an appropriate signal, the next move is entirely up to the other person. Both the stochastic and signaling models, therefore, predict that the frequency of between silences will decline monotonically with duration, as in Figure 1.[5]

Silence as Interactionally Generated

The sequential-production model of Sacks et al. (1974) leads to a distinctly different prediction. The central mechanism of the sequential-production model is an ordered set of options that becomes available when the current speaker reaches a transition-relevance place, which may be at the end of a word, a phrase, a clause, a sentence, or longer utterance, depending on the context.[6] The options in order are as follows:

[5]Smoothly declining distributions of a roughly exponential shape are, of course, quite familiar in social phenomena and can be generated by a variety of different underlying processes (e.g., Zipf, 1949), and Brady (1968) finds just such a size-frequency relationship for silences in his aggregated data (see Note 4, above). However, the point here is not that a stochastic or a signaling model is needed to predict a distribution of this particular kind, for obviously it is not. Rather, what is important is that these models do in fact predict a smooth monotonically declining distribution, and that this is quite different from the periodic distribution superimposed on an overall decline implied by the sequential-production model of turn taking (see below). Consequently, an opportunity arises for a fairly sharp empirical test between models predicting a smoothly declining distribution of silences and those predicting a periodic distribution.

[6]Turns longer than a sentence can be constructed by the speaker using particular devices to indicate that a story, complex question, or other lengthy turn is impending (Schegloff, 1980, 1982) and Wilson et al. (1984).

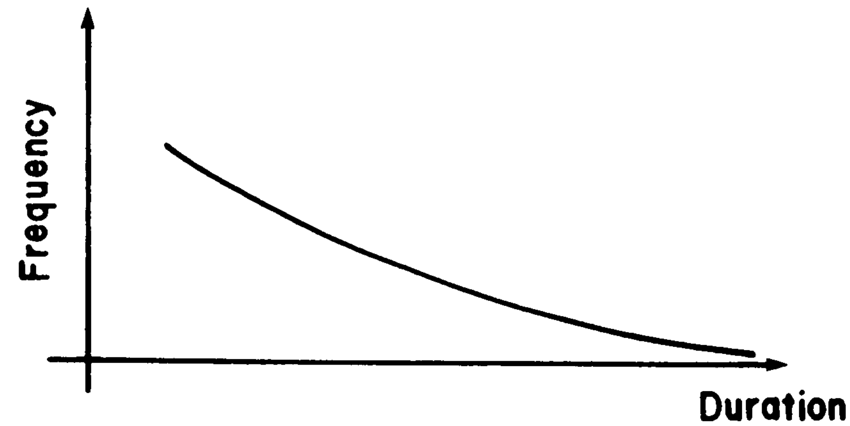

FIG. 1. Theoretical Distribution of Between-turn Silences according to Stochastic and Signaling Models of Turn Taking

(1) During the current turn, the person speaking may select the next speaker, for example, by asking a question addressed specifically to another;
(2) if the current speaker does not select the next, then the next speaker may select him- or herself, where the first person to start talking gains the turn;
(3) if no other person selects him- or herself, the current speaker may continue;
(4) if the current speaker does not continue, the options recycle back to (2) until either the current speaker continues or another person takes the turn. (See Figure 2.)

An important feature of this model is the assumption that, by monitoring the current turn as it develops, speakers and hearers can anticipate the end of the current unit type in advance of its occurrence and so anticipate an upcoming transition-relevance place. This *projectability* of transition-relevance places plays a fundamental role in the model. For example, if the current speaker does not select the next in the course of the current turn, option (2) will become available at the end of the present unit type, and thus a hearer can time his or her next utterance to begin precisely when the current speaker reaches the end of the unit type (Jefferson, 1973).

According to the sequential-production model, the cycling of the options is a process in real time (Sacks et al. 1974, p. 715). Hence when the conversationalists fail to exercise options (2) and (3) for some number of cycles, a silence is generated that terminates when option (2) or (3) is finally exercised.

The period during which a conversationalist can exercise option (2) or (3) can be thought of as a slot in which he or she can initiate a turn. If at any given transition-relevance place in a conversation the cycling of the options occurs at a regular pace, then a fixed length of time, say S ms, is required for each option,

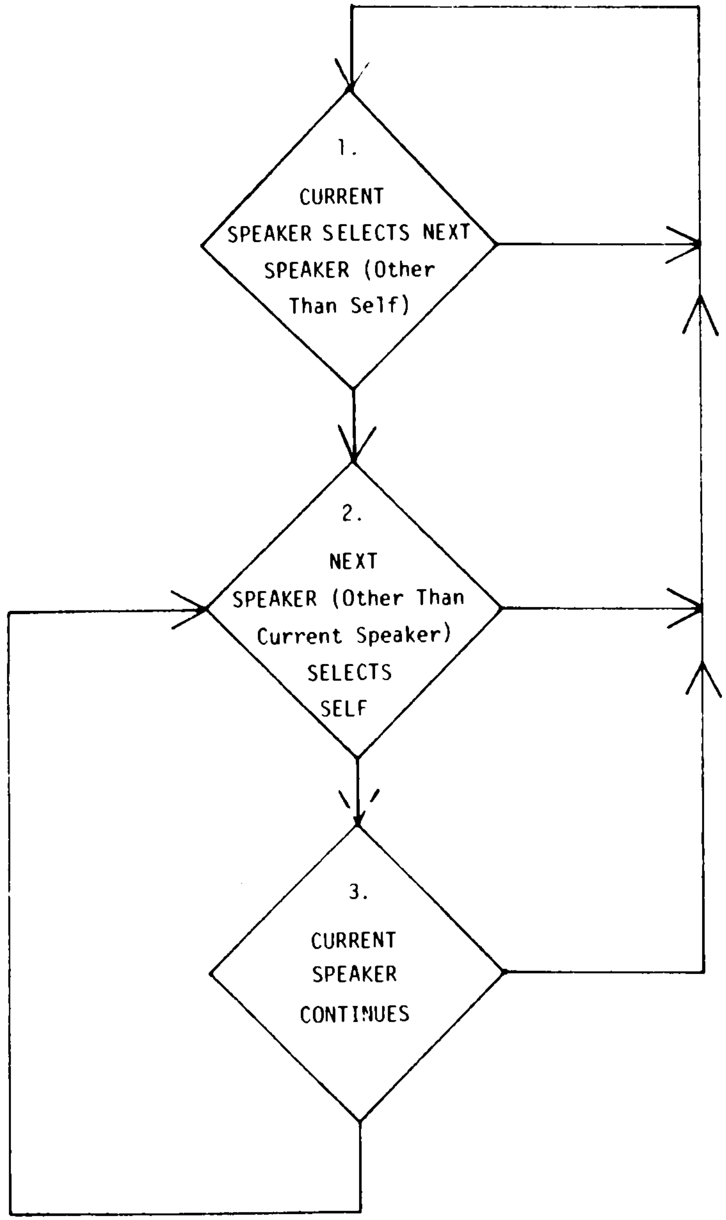

FIG. 2. Schematic Representation of Sacks et al. Turn-allocation Model

which can be called "slot length." It follows that a silence between two different speakers will last a length of time equal to an even multiple of the slot length S. For, if a silence begins with one speaker falling silent and ends with another speaker taking the turn, then both options (2) and (3) must have been passed up some number of times before option (2) is finally exercised. Hence the silence will last $2S \times k = 2kS$ ms, where 2S ms is the length of time needed to pass up options (2) and (3) in one cycle of the options, and $k = 0, 1, 2, \ldots$, is the number of times options (2) and (3) are passed up before the silence ends. Note that if the next speaker exercises option (2) at the first opportunity and so latches his or her turn immediately to the preceding turn without gap, the "multiple" k can be taken to be zero.

Now consider the frequency distribution of durations of between-turn silences in a conversation. If the slot length S remains constant throughout the course of the conversation, then, according to the model, this distribution will show peaks at durations of 2kS ms and valleys at durations of $(2k + 1)S$ ms, $k = 0, 1, 2, \ldots$. Moreover, because silences of long duration are less common than shorter silences, the heights of the peaks should decline regularly with the value of k (see Figure 3).

By a parallel argument, the model also predicts an alternating distribution for the durations of silences at transition relevance places within turns, but with the peaks occurring at intervals of $(2k + 1)S$ ms and the valleys at 2kS ms. However, for reasons noted below, attention here is confined to between-turn silences.

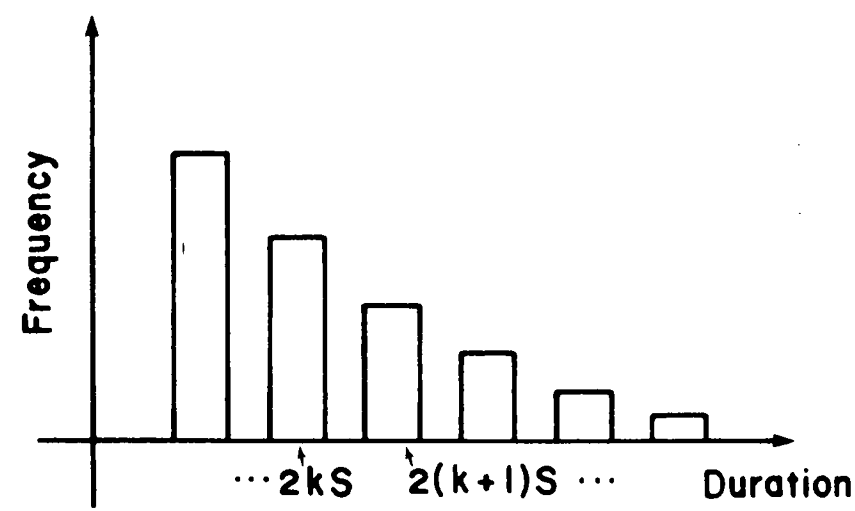

FIG. 3. Theoretical Distribution of Between-turn Silences according to the Sequential Production Model of Turn Taking

METHODS

The data were gathered as part of a program of research on conversational interaction, and for the present investigation seven trials were chosen unsystematically from a substantially larger corpus of conversations in a laboratory setting between randomly paired unacquainted university students.[7] Four of the conversations were male-female, two were female-female, and one was male-male, with no subject participating in more than one conversation. Consequently, each conversation was treated as a separate experiment for statistical purposes, and the entire study consists of seven independent replications of the tests of the two competing hypotheses.

A 9-minute segment, starting with the 2nd minute of talk, was used from each conversation. Each speaker was originally recorded on a separate track of a two-channel audio recorder to facilitate transcribing the conversations, but for the purpose of analog-to-digital conversion in the present study, the seven trials were rerecorded onto four-channel tapes with a 5 kHz timing signal placed on a third channel to provide a reference for pulsing on ADC 600-11 analog-to-digital converter at intervals of 0.2 ms. The data from the converter were fed directly to a PDP 11/45 computer for initial processing and storage.

Data Reduction

Each data channel was converted to digital form separately. Because the speech signal recorded on the tape generates a rapidly alternating positive and negative electrical signal as input to the analog-to-digital converter, the digitized output consisted of positive and negative values. However, only the magnitudes were of interest in this study, and so the negative values were converted to positive, and then all the values were summed in blocks of 50 observations to yield an effective sampling interval of 10 ms.[8]

Because of the laboratory recording conditions, the speech signals were substantially stronger than the background noise, and so it was possible to separate signal from noise by using a threshold. A first approximate threshold was established by inspecting plots of intensity as a function of time and noting the level of the background noise from which the speech signals stood out clearly. Each conversation was then analyzed at several thresholds bracketing the initial approximation, and the final threshold for each conversation was chosen to yield the sharpest results. The final threshold differed from the first approximation in three cases and in each instance was a level immediately above or below the first approximation.

Initially each speaker's channel was processed separately. Silence was defined as a signal below threshold, but the processing algorithm was designed to ignore below-threshold signals lasting 100 ms or less embedded in a single speaker's talk, because energy drops of this duration are common in normal speech production (e.g., Denes & Pinson, 1973, pp. 160–161). Following this, above-threshold signals lasting 100 ms or less surrounded by silence were treated as noise and converted to silence. (Cf. Brady, 1965; Barik, 1972.)

In the next step, the two channels were processed simultaneously. A *between-speaker* silence was defined as one or more contiguous 10-ms intervals in which both channels were below threshold, bounded by segments in which different channels were above threshold before and after. Silences in which the identity of the preceding or following speaker could not be established unequivocally, that is, situations of possible simultaneous speech, were discarded. A check was carried out to determine if leakage of sound from one speaker to the other's

[7]See West (1979) for a description of the data-gathering procedure. The conversations selected for this investigation happened to all be on a single tape, and so represent a contiguous sequence of experimental sessions; however, subjects were selected for the study and assigned to pairs randomly; pairs were then scheduled at times convenient for both the subjects and the experimenters. Transcripts or other data about the conversations were not consulted in selecting this particular tape for analysis.

The fact that these conversations were between unacquainted students in a laboratory setting could raise the objection that they do not provide data appropriate for a test of models of turn taking in "naturally occurring" conversations. However, this concern rests on several misconceptions.

First, Sacks et al. (1974) provide a formal definition of conversation as a type of speech-exchange system in which turn size, turn order and turn content are free to vary (see Note 2 above, and Wilson et al., 1984, for further discussion). Although laboratory conversations typically are subject to certain constraints (e.g., in their manner of starting or ending) there were no prior specifications affecting the order, length, or substance of turns at talk. However unusual their auspices, the dyadic encounters studied here clearly fall within Sacks et al.'s definition.

Second, although it is obvious that strangers interact differently in some respects than do the previously acquainted persons, there is no evidence that these differences lie at the level of such basic mechanisms of interaction as turn taking. Instead, the differences reflect the different circumstances of speakers or the requirements of the setting (e.g., the resources available to speakers to introduce topics (Maynard and Zimmerman, 1984)).

Third, it should be emphasized that Sacks, et al. (1974) posit a turn-taking mechanism for *any* conversation, in the technical sense of the term, that furnishes a context-sensitive means of organizing a wide diversity of conversational activities which hinge on the management of turn taking. There is no a priori reason to suppose that it, or stochastic and signaling models, would not pertain to conversations (again in the technical sense) between unacquainted persons in unfamiliar circumstances.

Finally, it is apparent that the term "natural" invokes a conception of casual conversation in relatively informal settings—an ill-defined notion at best. It seems inappropriate to implicitly define one vaguely specified assemblage of conversational encounters as "natural" and then to employ this unexplicated classification to question the relevance of findings based on allegedly different kinds of conversational occasions. The fact of the matter is that *any* two occasions can be shown to be different from each other in some respects, and the real issue is whether the differences are ones that matter for the purpose at hand—in the present case, the management of turn taking.

In sum, talk between strangers in a laboratory may not be a common or even comfortable event, but there is no evident reason to assume that it is not conversation at the fundamental level of the management of turn taking.

[8]The measurements were rectified for convenience in subsequent processing, because there was no interest in attempting spectral analyses of the speech stream.

microphone might affect the measurements; however, in no case was there evidence that an above-threshold signal on one channel accompanied by silence in the other channel could be attributed to immediately preceding talk by the other speaker.

By this technique, between-speaker silences as short as 10 ms could be measured. Since silences longer than 1 s were quite rare in these data, attention was confined to silences ranging from 10 to 1000 ms in duration. The number of silences in this range varied from 91 to 182 in the seven conversations. The resulting data were assembled in the form of a frequency distribution for each conversation of the durations of between-speaker silences. These distributions formed the basis of the statistical analyses described below.

It should be noted that the data concern between-*speaker* rather than between-*turn* silences. The difference is that between-speaker silences include events that are not silences between turns: For example, if the current speaker elects option (1) and selects the next speaker by some device such as a question, the turn immediately goes to the next speaker, and a silence at that point must be counted as within the turn of the next speaker. With the resources available for this study, it was not possible to remove those between-speaker silences that are not between-turn from the digitized data. Examination of one transcript showed that approximately 25% of the between-speaker silences are not between-turn silences. Consequently, the present data contain substantial measurement error, which needs to be taken into account in interpreting the results. In particular, this form of measurement error should bias the results against the sequential-production hypothesis, because it would tend to obscure a periodic pattern in the data.

Only data on between-speaker silences were analyzed, because approximately 50% of the silences within turns in the transcript did not occur at transition-relevance places and so would not be generated by the mechanism posited by the sequential-production model. As noted above, it was not feasible to cull misclassified silences out of the computerized data, so that meaningful analysis of the within-speaker distributions was not possible within the constraints of available resources, and attention was confined to between-speaker silences.

Analysis

The concern here is with two hypotheses concerning the distribution of silences between turns in a single conversation: The first, derived from the stochastic and signaling models, predicts a monotonic decline of frequency with duration (Figure 1); the second, derived from the sequential-production model, predicts a distribution having periodic peaks and valleys, with an overall decline in the heights of the peaks (Figure 3). For statistical purposes, it is useful to view the matter in the following way: If the stochastic-modeling or signaling approach is correct, then once the monotonic decline of frequency with duration has been removed, no systematic pattern should remain in the data; in contrast, if the sequential-production model is correct, a systematic periodic pattern should re-

main in the data after the monotonic decline has been removed.[9] Because an overall decline is predicted in both cases, these hypotheses can be tested against one another by first removing the monotonic decline and then by testing for the presence of periodicity. In sum, the prediction from the stochastic and signaling approaches serves as the null hypothesis in the statistical analyses and the prediction from the sequential-production model is the relevant alternative.

Two steps were taken as a preliminary to the statistical analysis. First, the decline in frequency with duration of silences was removed by differencing the data: That is, each value was subtracted from the following one.[10] Second, there was a marked tendency for frequencies of silences of immediately adjacent durations to differ substantially from one another, resulting in extremely ragged frequency distribution; because this tendency is irrelevant to both of the hypotheses in question, it was removed from the data by fitting a first-order moving-average (MA1) component after the data had been differenced. The result was a new data set for each conversation for which the sotchastic and signaling models predict no systematic pattern, and for which the sequential-production model predicts a periodic structure.

The question of periodicity was addressed using the ARIMA methods frequently employed in the analysis of time series (Box and Jenkins, 1976).[11]

---

[9]It is sometimes argued that it is inappropriate to try to employ statistical analysis in connection with work in the conversation analysis tradition. Apparently the idea is that while conversation analysis is based on a contextual conception of interaction processes, statistical techniques require "decontextualizing" the data. The basic issue here has to do with the manner in which the objects subjected to statistical analysis are identified. The point is well illustrated by the distinction made in the text between between-*turn* and between-*speaker* silences. Identification of between-turn silences requires analysis of the sequential organization of the particular stretch of talk in order to identify transition relevance places and the manner in which the turn has been exchanged, whereas between-speaker silences are those preceded by talk of one speaker and followed by talk of another, and can be identified mechanically. Thus, between-speaker silences are are in some sense "decontextualized" events, but between-turn silences clearly are not. It is important to re-emphasize, then, that use of between-speaker rather than between-turn silence data in this study results, not from any intrinsic inability to identify the latter, but rather reflects inadequate resources, with resulting measurement error. Thus, the use of "decontextualized" data in the present study has nothing to do with the use of statistical methods. It should also be noted here that Sacks et al. propose their model as a context-free, context-sensitive mechanism for the management of turn taking at any point in the course of a conversation. Thus, while the mechanism requires orientation to context for its proper use, the mechanism itself is available at any transition relevance place, and hence it can be employed to generate silences in exactly the same way at any transition relevance place at which the current speaker does not select the next.

[10]Differencing is an appropriate technique to remove a linear trend, but more complicated methods are required to deal with nonlinear trends such as an exponential. However, visual inspection of the frequency distributions suggested that a linear trend was a good approximation and that techniques more sophisticated than differencing were not required.

[11]ARIMA is an acronym for "AutoRegressive Integrated Moving Average," where "integration" refers to the inverse operation of differencing. An alternative approach would be to employ spectral analysis using the Fourier transform. However, in the present situation, a significant periodicity detected by one method should also be found by the other

Although ARIMA techniques were designed originally to deal with time series, they can be applied to any sequence of data in which the intervals between observations are equal, for example, values separated by equal distances in physical space. In the present case, an "observation" is the frequency of silences of a given duration, and these occur at durations of 10 ms, 20 ms, . . . , 1000 ms, yielding a series 100 observations long.[12] Basically, an ARIMA model attempts to analyze a sequence of observations for periodic structure by asking whether an observation at one point is correlated with observations at previous and later points. Thus, for example, if there is periodicity in the data, as called for by the sequential-production model, then observations one period apart should be positively correlated, and observations half a period apart should be negatively correlated. These correlations are measured by coefficients called "autoregressive" (AR) parameters, and a cyclic pattern shows up as two parameters: a positive AR parameter representing the positive correlation of values a full period apart; and a negative AR parameter representing the negative correlation of values half a period apart.[13] The stochastic and signaling models predict no such periodic components, whereas the sequential production model predicts that they will be present.

FINDINGS

After the data had been differenced and the MAI component had been removed, it was possible in each conversation to fit a model consisting of two components:

—a positive periodic AR component with a period of 80 to 180 ms;
—a negative periodic AR component at approximately half the period of the positive component (40 to 90 ms).

The results are reported in Table 1 for each trial. The "lags" given in the table are in 10-ms units and refer to differences between the durations of silences. Thus, for example, in Trial A the half-period AR parameter −.1337 at a lag of 4 means that frequencies of silences differing in duration by 40 ms (= 4 × 10 ms) are negatively correlated, and the full-period AR parameter .2776 with a lag of 8 means that frequences of silences 80 ms apart in duration are positively correlated. This is the pattern presented in ideal form in Figure 3 above.

With regard to the statistical analysis, three facts stand out. First, in each of the seven conversations, the estimated full-period autoregressive parameter is significant at the 1% level by a one-tailed criterion, which is appropriate since the sequential-production hypothesis predicts a negative sign for the half-period parameters and a positive sign for the full-period parameters. Second, the t statistics for the half-period parameters are all greater than 1 in magnitude, two being significant at the 5% level or better, and five significant at least at the 10% level. Moreover, when the full-period component is omitted from the model, the resulting new half-period parameter for each conversation (not shown in Table 1) is significant at least at the 10% level. This suggests that the borderline statistical significance for these half-period parameters may result from the larger full-period effect swamping the half-period effect. Third, the fitted model is successful in accounting for the systematic variation in the data: For each conversation, the chi-square tests for departure from white noise indicate that once the systematic components formulated in the model are removed, the residuals do not differ discernibly from a random pattern. In sum, the hypothesis of a periodic pattern in the distribution of between-speaker silences is strongly supported in each of seven independent replications.[14]

TABLE 1
Arima Models for Periodicity in Silence Distributions for Conversations[a] (Values of t given in parentheses)

	MAI	Half Period AR		Full Period AR		White Noise Chi-Square	
Conversation	Value	Value	Lag[b]	Value	Lag	1–12 Lags	1–24 Lags
A	.8334 (14.47)***	−.1337 (−1.29)	4	.2776 (2.88)***	8	9.6	17.9
B	.8125 (14.70)***	−.1357 (−1.48)*	9	.2050 (2.51)***	14	9.9	14.4
C	.6577 (8.48)***	−.1346 (−1.35)*	7	.2472 (2.53)***	13	4.3	11.3
D	.8384 (17.18)***	−.2617 (−3.44)***	7	.3497 (4.62)***	13	4.8	10.8
E	.7391 (3.29)***	−.0987 (−1.25)	9	.2012 (2.57)***	18	8.1	15.6
F	.7650 (11.23)***	−.1374 (−1.30)*	5	.2490 (2.46)***	10	4.2	14.8
G	.7698 (11.63)***	−.1823 (−1.92)**	7	.3248 (3.48)***	9	7.9	21.6

[a]Parameters fitted after first differencing. See text for terminology.
[b]Multiply lag by 10 to get half and full periods in ms (see text).
*Significant at .10 level. AR t-tests are one-tailed.
**Significant at .05 level. AR t-tests are one-tailed.
***Significant at .01 level. AR t-tests are one-tailed.

[12]Note that although time is involved here, the data do not constitute a time series: The observations are not separated by equal intervals of time, but rather represent frequencies of silences of successively longer durations, where the differences between successive durations are equal.

[13]In standard time-series analyses, these are called "seasonal" autoregressive parameters.

[14]Readers used to pooling data across subjects or trials would generally regard a mean based on 7 subjects or trials as very unstable. Consequently, it is important to emphasize that aggregation across

DISCUSSION

'he periodic structure in the distribution of between-speaker silences was pre-licted from the sequential-production model prior to the analysis (Wilson & 'immerman, 1979). However, this finding is not easy to explain on the basis of a imple response-latency interpretation of between-speaker silences. One might f course seek to accommodate the finding to a stochastic or signaling model by dding additional assumptions. But unless there is some motivation for such ssumptions arising from substantive considerations more fundamental than sim-ly a desire to account for the finding, an attempt of this sort must be regarded as d hoc and unconvincing. Thus, the sequential-production model receives signif-cant support vis-à-vis models of turn taking leading to a simple response-latency nterpretation of between-speaker silences. This, of course, does not deny the elevance of strategic and expressive use of silence, nor of other psychological rocesses, but it does emphasize the importance of seeing these in an interac-ional context rather than as purely individual matters.

Sacks et al. (1974) are not explicit about one detail of the turn-taking mecha-ism. On the one hand, it is possible that once a next speaker has declined to elect him- or herself on the first round, then the ensuing silence operates as a timulus or cue. In this case, in order for the current speaker to continue under ption (3), he or she would have to wait until it was clear that no other potential peaker had begun talking in the preceding slot before setting the vocal apparatus n motion. And a similar consideration would apply at each cycle of the mecha-ism until someone spoke. On the other hand, it is possible that the mechanism is projective one: A person electing to speak in an upcoming slot belonging to him r her does not wait for manifest silence in the slot immediately preceding the argeted one before setting the vocal apparatus in motion, and instead initiates the peech process in advance but timed so as to actually produce speech at the argeted upcoming slot when the option to speak passes to him or her. It seems more in the spirit of the sequential-production approach to think in terms of a projective mechanism than stimuli or cues, because such a mechanism already plays an essential role in the model,[15] and indeed there is a suggestion in the data that a projective mechanism may be more plausible.

The estimated slot lengths, derived from the periods for the AR parameters shown in Table 1, range from 40 to 90 ms with a mean of 60.00 ms. These are implausibly short as response times if the mechanism involves one party waiting until it is clear that the other has not taken up the option to speak in his or her slot

---

conversations is inappropriate here, and that the statistical analyses in this research must be conducted within each conversation separately. In each conversation, the tests for the parameters are based on 96 degrees of freedom, whereas the white noise tests are based on 9 and 21 degrees of freedom. For an aggregate assessment of the results, it can be noted that the probability of exceeding the .01 critical value in 7 independent trials if the null hypothesis in fact is true is 10^{-14}, which is infinitesimal.

[15]Notably when the next speaker latches his or her talk to the previous speaker's without gap (see Sacks et al., 1974).

before initiating the speech process, including, possibly, intake of breath. However, if the mechanism instead is projective, then all that is required is that a prospective speaker inhibit the incipient speech process if another person begins speaking first. No relevant data were located in the literature concerning the time required to shut down the vocal apparatus, once in motion, on receipt of a stimulus such as speech by another person, but it seems plausible that this may be much shorter than the time required to initiate the process and carry it through to the actual production of speech. An empirical test of this conjecture might be possible by examining the placement of breath intakes prior to speaker changes involving between-turn silences: If next speakers often start breath intakes well in advance of their slots, the idea of a projective mechanism would gain support. These reflections are, of course, quite speculative, and this is clearly an area in which further research is required.

Another question concerns the assumption underlying the analysis that the slot length S, though perhaps varying from one conversation to another, is constant over the course of a single conversation. Clearly, talk differs in pace both between conversations and, more importantly, over the course of a single conversation; moreover, there is some impressionistic evidence that proper assessment of the duration of a silence depends on the pace of the surrounding talk (E. Schlegloff & G. Jefferson, personal communications, 1978, 1978). This suggests that in fact the slot length is not constant within a single conversation, introducing another source of measurement error, though again this should tend to obscure any systematic pattern in the data. Consequently, another area for further research is to find a way to normalize measurements of durations of silences to the pace of the surrounding talk.

Finally, this investigation has been confined to between-speaker silences. Subsequent research should be directed to within-turn silences as well as replicating the present study.

CONCLUSION

The research reported here was designed to test the adequacy of the sequential-production model of turn taking proposed by Sacks et al. (1974). The approach taken was to derive a nontrivial novel hypothesis from the model and then examine the appropriate data. The prediction in question, a periodic structure in the distribution of between-turn silences, is supported by the evidence and runs counter to what would be anticipated if turn taking were adequately described by current stochastic or signaling models. This finding suggests that social interaction is a collaborative activity in a more fundamental sense than is recognized in approaches that view it as an exchange of stimuli and responses.

REFERENCES

Barik, H.C. (1972). Some innovations in a computer approach to the analysis of speech patterns. *Language and Speech, 15*, 196–207.

Box, G.E.P., & Jenkins, G.M. (1976). *Time series analysis, forecasting and control*. San Francisco: Holden-Day.

Brady, P.T. (1965). A technique for investigating on-off patterns of speech. *Bell System Technical Journal, 44*, 1–22.

Brady, P.T. (1968). A statistical analysis of on-off patterns of speech. *Bell System Technical Journal, 47*, 73–91.

Brady, P.T. (1969). A model for generating on-off speech patterns in two-way conversation. *Bell System Technical Journal, 48*, 2445–2472.

Cappella, J.N. (1979). Talk-silence sequences in informal conversations I. *Human Communication Research, 6*, 130–145.

Dechert, H.W., & Raupach, M. (Eds.). (1980). *Temporal variables in speech*. The Hague: Mouton.

Denes, P.B., & Pinson, E.N. (1973). *The speech chain*. Garden City, NY: Anchor Books.

Duncan, S. (1972). Some signals and rules for taking speaking turns in conversation. *Journal of Personality and Social Psychology, 23*, 283–292.

Duncan, S., & Fiske, D.W. (1977). *Face-to-face interaction*. Hillsdale, NJ: Erlbaum.

Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. Berkeley: University of California Press.

Goffman, E. (1955). On face work: An analysis of ritual elements in social interaction. *Psychiatry: Journal for the Study of Interpersonal Processes, 18*, 213–231.

Goffman, E. (1984). The interaction order. *American Sociological Review, 48*, 1–17.

Jaffe, J., & Feldstein, S. (1970). *Rhythms of dialog*. New York: Academic.

Jefferson, G. (1973). A Case of precision timing in ordinary conversation: Overlapped tag-positioned address terms in closing sequences. *Semiotica, 9*, 47–96.

Maynard, D.W., & Wilson, T.P. (1980). On the reification of social structure. In S. McNall & G. Howe, (Eds.), *Current perspectives in social theory* (Vol. 1, pp. 287–322) Greenwich, CT: JAI.

Maynard, D.W., & Zimmerman, D.H. (1984). Topical talk, ritual and the social organization of relationships. *Social Psychology Quarterly, 47*, 301–316.

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*, 696–753.

Schegloff, E. (1980). Preliminaries to preliminaries: Can I ask you a question? *Sociological Inquiry, 50*, 104–152.

Schegloff, E. (1982). Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. In D. Tennen (Ed.), *Analyzing discourse: text and talk* (pp. 71–93). Washington, DC: Georgetown University Press.

West, C.C. (1979). *Communicating gender, a study of dominance and control in conversation*. Unpublished doctoral dissertation, University of California, Santa Barbara.

Wilson, T.P. (1982). Qualitative "oder" quantitative methoden in der sozialforschung [Qualitative "versus" quantitative methods in social research] *Kölner Zeitschrift für Sociologie und Sozialpsychologie, 34*, 487–508.

Wilson, T.P., & Zimmerman, D.H. (1979, August). *Turn-taking in two party conversation*. Paper presented at the annual meeting of the American Sociological Association, Boston, MA.

Wilson, T.P., Wiemann, J.M., & Zimmerman, D.H. (1984). Models of turn taking in conversational interaction. *Journal of Language and Social Psychology, 15*, 159–183.

Yngve, H.V. (1970). On getting a word in edgewise. In M. A. Campbell, (Ed.), *Papers from the sixth regional meeting* (pp. 567–578). Chicago, IL: University of Chicago, Chicago Linguistics Society, Department of Linguistics.

Zipf, G.K. (1949). *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley.