

Technical Note 496 • October 1990

SPOKEN LANGUAGE IN INTERPRETED TELEPHONE DIALOGUES*

Prepared by:

SHARON L. OVIATT

PHILIP R. COHEN

ANN PODLOZNY

Artificial Intelligence Center
Computing and Engineering Sciences Division

SPOKEN LANGUAGE IN INTERPRETED TELEPHONE DIALOGUES *

Sharon L. Oviatt, Philip R. Cohen, and Ann M. Podlozny

Artificial Intelligence Center
SRI International

1 Abstract

This research outlines the predominant dialogue and performance characteristics of three-person interpreted telephone speech during service-oriented dialogues, in comparison with those of two-person noninterpreted dialogues. An empirical study was conducted in which 12 native English speakers each made one telephone call through an experienced telephone interpreter to a Japanese confederate who did not speak English, and a second call to a Japanese confederate fluent in English. In total, 24 dialogues were collected, each one containing two successfully completed service tasks, or 48 tasks total. This paper reports on comparisons performed between three-person interpreted and two-person noninterpreted speech, based on the same pool of tasks and English subjects. The unique characteristics of interpreted telephone dialogues are outlined, including structural and referential features, miscommunications and other performance characteristics, confirmatory language, and linguistic indirection. In addition, an analysis is presented of interpreters' strategic management of turn shifts, and of the content, sequencing, and chunking of information passed among speakers. The long-term goal of this exploratory research is the modelling of human speech, and the specification of preliminary target requirements for future automatic systems.

2 Introduction

The concept of a telephone system capable of automatically interpreting Japanese-English conversations, no matter how limited the domains selected, is a challenging proposition. For scientists and technologists to arrive at an optimal system design, it will be important to

*This research was supported in part by a contract from ATR International to SRI International, a grant from the System Development Foundation to the Center for the Study of Language and Information at SRI and Stanford, and a donation of services from AT&T's Language Line. Parts of this research have been presented at the International Conference on Spoken Language Processing, November, 1990, in Kobe, Japan. Authors' address: Artificial Intelligence Center, SRI International, 333 Ravenswood Avenue, Menlo Park, California, 94025

clarify how spoken human communication takes place during interpretation, during use of the telephone, and with speakers as culturally and linguistically different as Japanese and Americans. To date, very little research has considered the topic of human dialogues during telephone interpretation, with the exception of recent research by Iida and colleagues [3, 4], and theoretical work by Oviatt [5]. Likewise, little is known of the compensatory behaviors that speakers naturally engage in to support this demanding form of communication. However, in a broader context, interpretation is an example of mediated communication, and a form of multiparty communication. Both of these characteristics have received attention from researchers interested in facilitating computer-mediated communication and, more specifically, cooperative group work [7, 8]. To some extent, the properties of being mediated and multiparty suggest how interpreted telephone conversation may differ from two-person conversation, although it is not clear how these features interact with channel limitations and cultural/linguistic differences in defining interpreted telephone speech. Given the lack of available literature and the complexities of this modality, the present research was designed to begin exploring and providing an empirical foundation on its basic properties.

This study aimed to examine the unique discourse and performance characteristics of interpreted telephone speech during service-oriented exchanges. Three-person interpreted telephone dialogues that included a professional telephone interpreter were examined and compared with two-person noninterpreted calls, using the same English subjects and pool of tasks. Oviatt [5] previously has hypothesized that pressure on rapid and successful task accomplishment and the limitations of the telephone channel would encourage telephone interpreters to assume the role of an independent agent who actively manages information during the interpreted call. If true, then unlike the relatively literal, passive, and transparent interpretive style of face-to-face interpreters, telephone interpreters may be fulfilling regulatory functions that include managing the content, sequencing, and chunk size of information exchanged among parties. In addition to these possible functions, it was hypothesized that telephone interpreters would need to be pivotal in regulating turn shifts among speakers during interpreted calls, since multiparty interpreted conversations could potentially be more difficult to coordinate than two-person conversations. In particular, interpreted calls could be prone to higher rates of simultaneous speech and third-party interruptions than noninterpreted ones. As one consequence of interpreters' assumption of an independent role, it was anticipated that they would use first-person pronouns in self-reference, and third-person pronouns to refer to the person not currently speaking. That is, unlike face-to-face interpretation, telephone interpreters were not expected to use first person pronouns to literally represent the waiting third party, while remaining transparent themselves. In addition, it was expected that the lack of visual access among multiple telephone speakers would encourage interpreters to explicitly refer to all three participants in order to promote referential clarity.

With respect to dialogue structure, it was hypothesized that the conversation would be structured as a series of extended two-person subdialogues, English interleaved with Japanese. Frequent turn shifts were considered unlikely because of their high cost in loss of efficiency [2]. It also was hypothesized that interpreted calls would take speakers considerably longer to complete, and would entail a higher rate of miscommunications and speech dysfluencies

than noninterpreted calls in which the same subjects completed the same pool of tasks. As a result, these strains on performance were expected to prompt speakers to both request and provide confirmations more frequently. Of course, the structure of interpreted calls results in delayed reception of propositional confirmations by each of the main speakers, as well as the complete loss of meaningful backchannel confirmations. These factors could contribute further to speaker uncertainty about whether mutual knowledge was sufficient to support continued conversation, which also could prompt speakers to be conservative about double and triple-checking information. Finally, higher rates of linguistic indirection were anticipated during interpreted calls than noninterpreted ones, due to the pragmatic demands of mediated communication. The present report will outline the empirical results of this exploratory research on the basic characteristics of three-person interpreted telephone speech, and how it differs from the spoken language observed during two-person noninterpreted calls.

3 Method

3.1 Subjects

Twelve paid volunteers from the professional staff at SRI International participated in the study. All volunteers were fluent in English, with neither knowledge of Japanese nor experience with telephone interpretation services. Each subject made one interpreted and one noninterpreted telephone call in which they assumed the role of an American researcher who needed to collect information to register for a Japanese conference. During interpreted calls, each subject used a professional Japanese-English telephone interpreter to converse with a Japanese conference registrar who neither spoke nor understood English. In contrast, during noninterpreted calls, each subject conversed directly in English with an English-speaking Japanese registrar.

Two native Japanese women played the role of the registrar, one for the interpreted and the other for the noninterpreted calls. Both confederates were native Japanese with respect to language and cultural background, but they differed in their knowledge of English. The confederate for the interpreted calls neither spoke nor understood any English, whereas the confederate for the noninterpreted calls was a Japanese-English bilingual with outstanding English language skills. Both confederates were trained on the same set of conference registration and travel information materials, except that they received materials in the language they were instructed to speak during the study. After the confederates were thoroughly familiar with these materials, they were given practice fielding calls from American researchers with different requests. Given this preparation, both confederates were able to act as skilled agents within the specified task domain.

Three professional Japanese-English telephone interpreters from AT&T's Language Line participated during the interpreted calls. All interpreters rated "Advanced/Superior +" on the English/Japanese sections of the U.S. Government ACTFL's Language Proficiency Test, and had the full-time equivalent of between 3 and 12 months' experience working specifically as telephone interpreters. Interpreters were oriented to the type of participants and tasks

they would be handling, but only the Japanese registrars had detailed reference materials on registration and travel. In addition to general orientation, each interpreter practiced handling one or two calls between a researcher and the registrar. These professional interpreters represent a unique population with experience specifically in telephone interpretation. Their participation was designed to enhance both the naturalism and external validity of the study's results, since only interpreters who habitually operate in this modality can be expected to have established a stable interpretation style with respect to performance and discourse patterns.

3.2 Tasks

Twelve conference registration tasks and twelve travel information tasks were constructed, based on materials from a real Japanese conference. All tasks involved simple service-oriented exchanges. A full set of task instructions is listed in Appendix A. The conference registration tasks required making arrangements to obtain and return a registration form, and collecting information about a registration question. Completing this task required the registrar to retrieve information from a "Conference Information Sheet," and required all participants to relay factual information such as names and numbers. For example:

1. Get the form you will need to register yourself for the conference. Ask the registrar to FAX you a copy of the registration form (your FAX #: 408-632-3000). Find out whether you can pay in traveller's checks when you arrive. Get the registrar's FAX number so you will be able to return the completed forms quickly.

Registration payable in traveller's checks upon arrival? _____
FAX # of registrar - _____

In contrast, the travel tasks required collecting information about hotels or travel connections within Japan, and accomplishing this task required solving a problem. The registrar again had to retrieve relevant information from a reference sheet, but now the participants also had to solve a problem, given two constraints, and then relay the solution.¹ For example:

2. Another student and you want to share a room at the conference. Find out from the registrar which conference hotel that has subway access to the conference site has the cheapest twin room. You will need to indicate the name of your hotel selection on the registration form when you receive it.

Name of preferred hotel - _____
(with subway access and cheapest twin room)

¹Subsequent analysis of the transcripts confirmed further basic functional distinctions between the registration task and the travel task. While completing task 1, 82% of subjects' subdialogues involved transferring new information, whereas less than 18% were clarification subdialogues. In contrast, while solving task 2, only 34% of subjects' subdialogues simply transferred new information, whereas 66% now focussed on clarifying the problem and its solution.

During discussion of each task, both the researcher and registrar wrote down information needed to follow through in actually accomplishing it. This information was collected to make the tasks more realistic, and to provide an objective index of task performance. In the examples given above, the researcher filled in information beside the prompts, and the registrar filled out a "Contact Report Form" on which she recorded each caller's name, contact information, their request, and any information provided or follow-through actions required to satisfy their request. In short, the instructions were designed both to encourage and assess successful completion of the tasks, and only dialogues that satisfied this basic performance criterion were subjected to full analysis. The only exception to this was analysis of miscommunications, which was performed on all dialogues irrespective of whether the task was completed successfully or not.

3.3 Design and Procedure

Each of the 12 subjects completed four tasks, one registration and one travel task during their interpreted call, and different registration and travel tasks during their noninterpreted call. Both of these phone calls were completed during a half-hour session. As a result, data on both tasks and communication modalities were collected within-subject. A within-subjects design was selected to minimize linguistic variability due to individual differences.

Subjects placed their noninterpreted call to the English-speaking confederate, and their interpreted call to the Japanese-speaking confederate. For the 12 interpreted calls, four subjects each used one of three interpreters. Three different interpreters were studied so that a range of telephone interpretation styles could be sampled. In addition, the order of placing the interpreted (I) and noninterpreted (NI) calls was counterbalanced across the four subjects assigned to each interpreter, so that half were I/NI and half NI/I. The two sets of tasks were shuffled so that each subject received different registration and travel tasks for their two calls, but the 12 interpreted and 12 noninterpreted calls were completed using the same task sets.

The interpreted calls were conducted as conference calls among the three parties, so the English researcher and Japanese registrar each could hear incomprehensible speech in the other language. During both the interpreted and noninterpreted calls, speakers were separated and visually inaccessible to one another.² For both types of call, the researcher and registrar were located in private offices at SRI. However, for the interpreted calls, the professional telephone interpreter was patched in long-distance from Monterey, California. Interpreters were asked to handle these calls as they would any other during their daily work routine. All participants were aware that their speech was being recorded for later study by the researchers.

²Certain aspects of this study were designed to represent actual trends in the emerging commercial use of telephone interpretation. For example, the interviews with Language Line interpreters indicated that most Japanese-English calls that they handle are service-oriented business ones, that one of the two main speakers usually is a first-time user of phone interpretation services, and that three-person conference calls in which the participants are physically separated is the most common structural arrangement for interpreted calls. These types of features were incorporated in the present research design so that the results would bear relevance to high-frequency interpreted calls, which not only enhances the external validity of the results, but also generates data in support of emerging commercial needs.

3.4 Transcript Preparation and Coding

All calls were tape-recorded, using an in-line taping device. Signal distortion was not measured directly in either modality, although less than 1 in 2000, or 0.05%, of all recorded words were judged undecipherable by the transcriber and experimenter during the noninterpreted calls. In the interpreted calls, less than 1 in 500, or 0.2%, were judged undecipherable during the Japanese and English subdialogues, with less than 1 in 300, or 0.3%, during the English sections upon which most analyses were based.

All English and Japanese speech was transcribed from the tape-recordings, with transcription performed by a native speaker of each language. Transcripts were produced that recorded the contributions of all participants, including the interpreter during interpreted calls. Special attention was paid to recording and representing the precise sequencing of each participant's contributions during the dialogue, including the location of interjected back-channel confirmations from the listener, two- and three-person simultaneous speech, and so forth. Appendix B lists the transcription conventions for marking these features, along with samples of transcription for part of task 1 for subject 11's interpreted and noninterpreted calls. Attention also was paid to accurate transcription of spoken language phenomena such as nonword fillers, repetitions, false starts, self-corrections, and confirmations. That is, the transcribers were instructed not to "clean up" any such spoken language dysfluencies, but rather to record them literally as heard. For both interpreted and noninterpreted calls, 17% of the transcribed language (i.e., both English and Japanese) was double-checked to confirm the accuracy of any linguistic phenomena that were being coded. For all such measures, the transcription match was above .92.

The transcribed Japanese speech then was translated. The translator was instructed to perform a relatively literal translation, although not awkwardly so, rather than a more colloquial and idiomatic one. Like the transcribers, the translator also was instructed not to clean up any spoken language dysfluencies or communication errors, but rather to represent their presence in the translation as faithfully as possible. A second translator also double-checked the translation accuracy of linguistic phenomena that were coded in the Japanese subdialogues. In particular, checks were performed on the correctness of all miscommunication codings, and on codings of all first and third-person pronouns that had been filled in during translation to English from Japanese.

After this preparation, the content of the dialogues was coded and analyzed for linguistic and performance features of interest. The methods for coding the dependent measures are summarized below.

Indices of Interpreters' Management of Task Information- The interpreted calls were coded for instances in which the interpreter: 1) provided extra task-relevant information to a speaker (e.g., offering dialing information about the Japanese country code), without having been asked for it directly, and without first having been given the information by the other speaker; 2) directly answered a speaker's question when the interpreter already knew the needed information (e.g., quoting registration fees; spelling the name of a Japanese service), without first checking with the other speaker for the answer, thereby shortcutting the usual interpretive process; 3) selectively omitted passing information provided by one speaker that

was not central to task solution (e.g., not transmitting the student registration fee when only the general fee was asked for); 4) offered to provide a speaker with a task-relevant action or item of information that had not yet been mentioned (e.g., offering to arrange for the registration form to be sent; offering to find out a hotel cost); and 5) prompted a speaker for task-relevant information not yet raised during the conversation (e.g., asking whether a quoted fare is one-way; asking whether connecting transportation is available right at the airport). A tabulation was made of the total number of times that interpreters engaged in each of the above types of behavior during each call.

Lag in Interpreters' Transmission of Task Information- For the interpreted calls, each task was analyzed into the basic information needed to complete it successfully. For example, the task 1 listed in section 3.2 was divided into six task-critical items of information, all of which were defining elements that needed to be transmitted by the English-speaking researcher: request for a registration form; request that it be FAXed; identification of the researcher's name; identification of the researcher's FAX number; request for the registrar's FAX number; request for information about method of paying registration. Then, for each such information item, the English subdialogue in which it was introduced by the researcher to the interpreter was identified, as well as the Japanese subdialogue in which it was passed by the interpreter. The percentage of time then was calculated that interpreters passed such items during the Japanese subdialogue immediately following its introduction.

Number of New Information Items Per Subdialogue- For each task during interpreted calls, the maximum number of new task-critical information items that was introduced per subdialogue was totalled. Tasks then were categorized into those in which no more than 2 items were passed per subdialogue (i.e., low chunking), versus those in which up to 3 or more were passed (i.e., high chunking).

Interpreters' Use of Noun Phrases and Personal Pronouns- A tabulation was made of the number of times each interpreter used noun phrases and first-person singular or plural pronouns such as "I," "me," or "we" in self-reference while speaking with the two primary speakers (e.g., "Hello, this is Japanese interpreter number 385. May I help you?"). A tabulation also was made of the number of times each interpreter used noun phrases and third-person singular or plural pronouns such as "he," "she," or "they" to refer to the participant not currently engaged in a subdialogue (e.g., "Okay, here's her hotel recommendation."). In some cases, pronouns like "we" or "our" were counted in both of the above categories. These first and third-person counts then were converted to a rate per total words uttered. The total number of instances also was counted in which interpreters used first-person singular or plural pronouns to refer to the waiting participant, as if to literally represent that person. During interpreters' Japanese subdialogues, pronouns that were literally present in the text were distinguished from those filled in during translation to English, with the latter being checked by the second translator for accurate translation as first or third person.

Confirmations and Requests for Confirmations- For both the interpreted and non-interpreted calls, the total number of words that each English-speaking researcher spent eliciting and issuing confirmations was tabulated. Included among the scored confirmations were listener backchannels such as "Mhmm" and "Okay," propositional confirmations such

as "Yes" and "Sure, if they can FAX it," verbatim confirmations such as repeating digits and spelled letters, summary confirmations such as "Okay, now let me repeat that back to you" (followed by the repetition of several information items just discussed), and interpreter affirmations of intent to confirm such as "Let me make sure of that again, Ma'am" (followed by a subdialogue shift, and then the interpreter's return with a confirmation or further clarification). Phrases were scored as confirmations or requests for confirmation only when they did not entail introduction of new information. For both types of call, a calculation was made of the percentage of time that each researcher spent issuing and eliciting confirmations out of their total word production.

Linguistic Indirection- For both the interpreted and noninterpreted calls, a tabulation was made of whether each English-speaking researcher's first attempt to satisfy a task-critical goal was formed as a linguistically direct or indirect request. Only first goal satisfaction attempts for information or action, and not subsequent clarifications, were coded. Examples of requests coded as indirect included "I need to register," "I want to get some information," and "Can you ask the registrar X?" whereas those coded as direct forms included imperatives and "Wh" questions such as "What time is the bullet train?" For both types of call, a calculation was made of the total percentage of these first attempts at goal satisfaction that were presented indirectly.

Length of Conversation- For both the interpreted and noninterpreted calls, the total length of each subject's conversation was recorded separately for tasks 1 and 2. Total time was measured using a digital stopwatch, and rounded to the nearest second. The onset of timing began with the initial greeting between two speakers after all two or three speakers were on the phone together. Timing ended after the researcher's first clear signal of completion, followed by an acknowledgment of acceptance by the interpreter, after all information needed to perform the task had been exchanged. When the interpreter initiated closing the call instead of the researcher, then timing ended with the researcher's acceptance of this closing. When a subject returned to discussing or confirming task 1 information at the end of task 2, that section of conversation was timed separately and added to task 1.

Number and Length of Subdialogues- For the interpreted calls, a tabulation was made of the total number of two-person subdialogues in English and in Japanese. The total duration of each subdialogue was measured using a digital stopwatch, and rounded to the nearest second.

Interpreters' Linguistic Marking of Subdialogue Start and End- A tabulation was made of the proportion of all interpreted subdialogues for which the interpreter provided clear linguistic marking in the appropriate target language: 1) only of its beginning, 2) only of its ending, or 3) both. Interpreters' start cues included phrases like "Hello, Ma'am?" and "Yes, anything else, Ma'am?", whereas end cues typically included comments such as "One moment, please," "Please hold," and "Let me check into that for you."

Miscommunications- All task-relevant miscommunications were tabulated for each interpreted and noninterpreted call. These miscommunications could be initiated by any participant, and those that remained uncorrected were distinguished from ones that were corrected by another speaker. All types of miscommunications were counted, however major or minor,

including mistransmission of letters or digits, terminological errors, and general misunderstanding of a question or task goal. Transcripts were not scored further for dialogue and performance features if they contained any major miscommunications resulting in failure to complete the task correctly. However, the analysis of miscommunications was based on all collected transcripts.

Simultaneous Speech and Third-Party Interruptions- For both the interpreted and noninterpreted phone calls, the average number of overlapped words was computed for the two English speakers, and the rate of simultaneously spoken words per total words uttered then was summarized separately for these two types of call. In addition, for interpreted calls, a tabulation was made of the total number of times that all three parties spoke simultaneously, or that the waiting party interrupted an ongoing subdialogue in the other language.

Speech False Starts- For both the interpreted and noninterpreted calls, the total number of false starts was recorded for each English-speaking researcher, and then converted to a rate per total words uttered. An incomplete noun or verb phrase was counted as a false start if it was followed by a self-correction involving a substitution of concepts and grammatical constructions, or simply a substitution of grammatical constructions, and it was not influenced by another speaker's interjection. For example, "I will, um, would like to, um, get a form in order to register my, um, for myself," was scored as two false starts, and "Okay, um, the name, um, her name is Monique," was scored as one false start. Neither simple repetitions of words or phrases, nor hesitations with nonword fillers were counted as false starts.

Inaudible Speech- For both the interpreted and noninterpreted calls, the total number of inaudible words was counted in English and Japanese. A word was considered to be inaudible if both the transcriber and researcher judged it to be either unclear enough to require guessing or completely inaudible. The percentage of inaudible speech per total words then was summarized separately for English in the noninterpreted calls, English in the interpreted calls, and Japanese in the interpreted calls.

Total Number of Words- The total number of English words was tabulated for all English subdialogues, and then used as a basis for converting other dependent measures into rates or proportions.

Reliability- For both interpreted and noninterpreted calls, second scoring was completed for 17% of the data for each reported dependent measure. Interrater reliability was calculated as the percentage of agreements out of the total number of codings per category. All dependent measures reported in this paper had reliabilities ranging over .84, and 90% of the measures had reliabilities above .97.

4 Results and Discussion

Several analyses were performed in an effort to characterize the basic interpretive approach adopted by professional telephone interpreters. It was hypothesized that telephone interpreters, perhaps in response to the emphasis on task accomplishment and limitations inherent in the telephone modality, would assume the role of independent agents who actively manage information needed to complete the researchers' specified tasks [5]. To address this issue,

the interpreted transcripts were coded for any instances in which the interpreter departed from a strictly literal interpretation by providing extra information, shortcutting, selectively omitting information, or offering to provide or prompting a speaker for information. Of the 12 interpreted calls, all contained examples of these initiating interpreter behaviors, with 86 total instances, or an average frequency of 7.2 instances per transcript. The most common involved the interpreter offering to provide an action or piece of information, and interpreter prompts for information from another speaker (average total frequency per transcript = 4.8), whereas interpreter provision of extra information, shortcutting, and selective omission of information occurred less often (average total frequency per transcript = 2.3). In short, the experienced telephone interpreters in this study clearly played an active role in both directing the content of these service-oriented calls, and in organizing the flow and sequencing of task information. Most of their initiative, however, appeared to focus on the management of information sequencing.

However, interpreters rarely altered the content or chunk size of significant new task information presented by speakers during tasks 1 and 2. That is, when speakers presented task-critical new items of information to the interpreter (e.g., their name, their FAX number), 90% of the time the interpreter immediately passed these items to the registrar during the next Japanese subdialogue. In this respect, interpreters were prompt and literal about passing significant new information as they received it. About 63% of the time, speakers adopted the conservative approach of not presenting more than 1 or 2 new items of information to the interpreter per subdialogue, such that cognitive load on interpreters was kept to a minimum. However, when subjects were divided into those who chunked more new information into a subdialogue (≥ 3 items) versus those who chunked less (≤ 2), it was confirmed that high chunkers completed the task in significantly less time than low chunkers, 271.5 versus 342.5 seconds, respectively, $t = 1.872$ ($df = 10$), $p < .05$, one-tailed. On the other hand, in the few cases in which a speaker presented 5 to 7 new items of information all in one subdialogue, interpreters never passed all newly-received information during the next Japanese subdialogue. That is, at least for the type of tasks studied, optimal efficiency was associated with subject chunking of 3 to 4 new items per subdialogue, while 5 or more items appeared to exceed telephone interpreters' ideal load.

The transcripts also were coded for interpreters' manner of referring to the three participants in the interpreted call. It was hypothesized that telephone interpreters would represent themselves explicitly as an independent agent, rather than adopting a transparent interpretation style in which only the primary speakers are represented. Analyses revealed that interpreters used first-person pronouns and noun phrases to assert themselves as a separate agent an average of 19.2 times per transcript, and they used third-person pronominal or noun phrases in reference to the person not engaged in the current subdialogue an average of 15.0 times per transcript. In fact, all three telephone interpreters adopted the habitual style of explicitly referring to each of the three parties engaged in the call, with "I" and "me" used self-referentially by the interpreter, and third-person pronouns used extensively to designate the waiting party. Furthermore, all interpreters exhibited this third-person referential pattern, independent of whether the two primary speakers did so. These referential features are

consistent with telephone interpreters' assumption of an independent and active managerial role during these calls. Given the lack of visual access in the telephone modality, this explicit referential style may have helped to prevent confusion about which of the three speakers was being represented.

Further analyses revealed that interpreters substantially increased their use of explicit third-party references from their first two to their second two calls in the present study. Whereas interpreters' first-person references remained stable (averaging 1.85 to 1.98 first-person references per 100 words from their first to second two calls, $t < 1$ ($df = 5$)), their third-person references increased an average of 106%, from 1.08 to 2.23 instances per 100 words from their first to second two calls. This represents a significant increase, t (matched) = 4.27 ($df = 5$), $p < .004$, one-tailed. These findings not only indicate that interpreters habitually use explicit self-reference and third-party references during interpreted telephone calls, but that their use of third-party references becomes more entrenched, evidently as they become more familiar with the domain they are handling. That is, even though the present interpreters were experienced in the use of the telephone modality, specific domain experience may be needed before interpreters' explicit reference to all three parties reaches its natural ceiling.

With respect to overall dialogue structure, as had been hypothesized, the interpreted calls were organized into a series of extended subdialogues between the interpreter and each of the two primary speakers in their native language. The transcripts averaged 15 subdialogues apiece, including Japanese and English subdialogues on both tasks 1 and 2. The average length of each subdialogue was 39 seconds. The average length of the Japanese subdialogues did not differ from the English ones (40 versus 38 seconds per subdialogue, respectively, with $t < 1$). In addition, there was no reason to believe that the total number of words differed for the Japanese and English sections of each transcript (averaging 693 English words to 690 English words translated from Japanese, with $t < 1$). However, the English speakers often reported the Japanese turns as taking longer. That is, there appears to have been subjective distortion of the length of turns involving incomprehensible language. Similar subjective distortion also has been reported regarding the length of unfilled pauses by telephone speakers during noninterpreted calls [1].

With respect to basic performance issues, the interpreted calls averaged about 2.5 times longer than noninterpreted calls made by the same English subjects accomplishing the same pool of tasks. The average length of task 1 was 2.4 times longer for the interpreted calls (317 seconds for interpreted calls, as opposed to 133 seconds for noninterpreted calls, with $t = 7.74$ ($df = 11$), $p < .0001$, one-tailed), whereas task 2 averaged 2.7 times longer (224 seconds for interpreted calls, as opposed to 83 seconds for noninterpreted calls, with $t = 6.40$ ($df = 11$), $p < .0001$, one-tailed).

The frequency of all speakers' task-relevant miscommunications, including both those that remained uncorrected and those that were corrected by another speaker, was judged from the transcripts and performance sheets for both the interpreted and noninterpreted calls. Miscommunications were scored from all 15 subjects run during the study, including three subjects whose transcripts were not analyzed further because of inadequate task performance or technical considerations. This rate of miscommunications averaged 1.93 per task set for the

interpreted calls, in comparison with 0.45 for the noninterpreted ones, a significant difference by Wilcoxon's Signed Ranks Test, $T+ = 66$, $p < .0005$, one-tailed. Of these, the rate of errors that remained uncorrected was 0.93 for the interpreted calls, compared to 0.13 for noninterpreted calls. This elevated miscommunication level during interpretation included errors in the transmission of letters and digits, terminology errors, and occasional general misunderstanding of task requests, with 67% of all errors due to mistransmission of one or two spelled letters. Most errors were minor (e.g., "Harry" misspelled as "Hary"), and did not lead to consequential performance problems. During the interpreted calls, interpreters were responsible for initiating 55% of the errors that occurred and, since they uttered 56% of the total words, this rate corresponded with their general participation level. That is, it was neither higher nor lower proportionately than the two primary speakers.

An examination of the English-speaking researchers' speech false starts revealed no significant differences between interpreted and noninterpreted calls for this area of performance. The researchers' rate of speech false starts averaged 0.92 per 100 words during the interpreted calls, and 1.02 per 100 words during the noninterpreted calls, $t < 1$. That is, although overall miscommunications were elevated during the interpreted calls, the English researchers themselves had no apparent special difficulties with speech dysfluencies as they planned and articulated task requests during the interpreted calls.

It was hypothesized that the speakers' concern over being understood during interpreted calls might generate a more conservative communication style, including a higher rate of confirmation language geared toward double- and triple-checking the accuracy of information obtained from the interpreter. In fact, an examination of the English-speaking researchers' requests for confirmation and confirmations of information revealed that their rate of confirmation language per total words spoken increased significantly from 23.5% during the noninterpreted calls to 31.5% during interpreted calls, t (matched) = -2.306 (df = 11), $p < .025$, one-tailed. During the interpreted calls, then, nearly one-third of their language was exclusively concerned with the verification of information. A previous study on noninterpreted telephone dialogues in which the speakers completed an assembly task reported an 18-19% rate of confirmation language [6], in comparison with the 23.5% rate in the present service tasks — both lower than the remarkably high rate found for interpreted speech in the present study. This heavy emphasis on confirmation language did not appear to be sufficient, however, to control the higher rate of miscommunications during interpreted calls.

Consistent with expectations, when first attempting to satisfy their task goals during interpreted calls, the English-speaking researchers produced a higher percentage of indirect linguistic forms to request information or actions. For example, they used indirect requests 92% of the time, rather than imperatives or direct "Wh" questions. These same subjects, completing the same set of tasks, reduced their rate of indirect requests to 60% when speaking to the registrar during noninterpreted calls, a significant decrease by Wilcoxon's Signed Ranks Test, $T+ = 66$, $p < .0005$, one-tailed. They may have used a direct style more often during noninterpreted calls simply because they expected the registrar to know the information requested, whereas they did not have such expectations of the interpreter. During interpreted calls, the researchers instead transmitted their goals to the interpreter, since they expected

the interpreter to broker for them, or to collaborate with the registrar in helping them to solve their problem.

Finally, the data revealed that three-person interpreted calls were no more disorganized than two-person noninterpreted ones with respect to turn shifts between subdialogue partners. Analyses of the English segments indicated that two-person simultaneous speech was no more frequent during interpreted calls than during noninterpreted ones, with 2.04 versus 2.19 words of simultaneous speech per 100 words uttered, respectively, $t < 1$. In addition, three-way simultaneous speech and third-party interruptions of a subdialogue were extremely rare during the interpreted calls in this study, with only two occurrences of each found in all the transcripts. Instead, turn-shifting between the English and Japanese subdialogues was remarkably clean during interpreted calls. Further examination indicated that telephone interpreters took considerable initiative in the management of turn regulation. Of course, they effectively signalled turn shifts by initiating switches between languages, a forceful sign to both speakers of the opening and closing of the communication channel. In addition, interpreters provided clear linguistic marking of the start and end of most subdialogues, with the start of subdialogues indicated over 91% of the time, and both the start and end indicated over 57% of the time.

This exploratory research was designed to begin examining the unique characteristics of interpreted telephone dialogues, including how they differ from noninterpreted ones, and to begin specifying the pivotal role played by skilled telephone interpreters as they attempt to expedite this type of dialogue exchange. To optimize the design of future automatic systems, it is clear that effective strategies will be required for controlling the excessive lengthiness and miscommunications to which this modality is prone. Systems also will need to support effective confirmation feedback and, ideally, they would benefit by aiming to reduce the total proportion of language that is consumed by confirmation exchanges during interpreted communication. In addition, techniques for managing the lengthy lag experienced in this modality by the waiting third participant would be very strategic. One possible key to resolving these and other major design issues for future systems may be the incorporation of an AI interface that models the more effective regulatory behaviors of experienced human telephone interpreters, some of which have been outlined in this report.

The present results should begin to provide a basis for predicting many of the human speech patterns that are likely to be encountered by any automatic telephone interpretation system designed to handle service-oriented tasks, patterns that will need to be accommodated for a system to function in an accurate and robust manner. Furthermore, the results should provide test cases of the more frequent and important linguistic phenomena that theoretical work must be able to handle, as well as a basis for constructing and improving algorithms and system design more generally. In the future, it would be instructive to compare the dialogue patterns obtained in this study with those from users speaking to a prototype of an automatic telephone interpretation system.

References

- [1] B. Butterworth, R. R. Hine, and H. D. Brady. Speech and interaction in sound-only communication channels. *Semiotica*, 20(1-2):81-99, 1977.
- [2] H. H. Clark and S. Brennan. Grounding in communication. Unpublished ms., Psychology Department, Stanford University, 1990.
- [3] H. Iida, K. Kogure, I. Nogaito, and T. Aizawa. Analysis of telephone conversations through an interpreter. Technical Report TR-1-002, ATR Interpreting Telephony Laboratories, Osaka, Japan, May 1987.
- [4] H. Iida, M. Kume, I. Nogaito, and T. Aizawa. Collection of interpreted telephone conversation data. Technical Report TR-1-000X, ATR Interpreting Telephony Laboratories, Osaka, Japan, October 1987.
- [5] S. L. Oviatt. Management of miscommunications: Toward a system for automatic telephone interpretation of Japanese-English dialogues. Technical Report 438, Artificial Intelligence Center, SRI International, Menlo Park, California, May 1988.
- [6] S. L. Oviatt and P. R. Cohen. Discourse structure and performance efficiency in interactive and noninteractive spoken modalities. *Computer Speech and Language*, 1991, in press.
- [7] *Proceedings of the conference on computer-supported cooperative work*, Portland, Oregon, October 1988. ACM Press.
- [8] *Proceedings of the conference on computer-supported cooperative work*, Los Angeles, California, October 1990. ACM Press.

5 Acknowledgments

Sincere thanks to the generous people who volunteered to participate in this research as subjects. Special thanks also to the AT&T Language Line telephone interpreters who participated both as interviewees and interpreters for the main study, and to Dr. J. Sottung and others at AT&T for their support and interest in this project. We gratefully acknowledge the valuable assistance of Ms. T. Miyoshi, Dr. H. Miyoshi, Dr. M. Kameyama, and Mr. G. Kawai during the conduct of the experiment. Thanks also to Ms. M. Akahani for careful preparation of the Japanese transcripts. We are indebted to Dr. M. Kameyama, Dr. A. Kurematsu, Dr. C. R. Perrault, and members of the Natural Language group at SRI for helpful comments during presentations of this research. Finally, we are grateful for financial support from both ATR International and the System Development Foundation.

A Task Instructions

General instructions to English-speaking researchers before beginning tasks 1 and 2 in the interpreted and noninterpreted telephone call conditions

Researchers' Instructions During Interpreted Call-

"You are an American researcher, and you need to make a telephone call to register for the Pacific Rim International Conference on Artificial Intelligence (PRICAI) '90 in Nagoya, Japan. Right now, you will be telephoning the Japanese conference registrar for information. **The conference registrar *only* speaks Japanese**, and does not understand English, so a professional Japanese-English interpreter will be on the line with both of you in a conference call as you make your arrangements. You will be able to speak with the interpreter in English to make your arrangements.

During the phone call, you need to accomplish two separate tasks, which are listed below. You should *finish* the first task before beginning the second. For each task, copy down all the information given to you by the registrar that will be needed to accomplish that task, filling in the lines below. Be sure to get all the information you need, and to copy it accurately. When you are done with both tasks, tell the interpreter "End of Call."

If you have any questions about your task, please ask them now."

1. Get the form you will need to register yourself for the conference. Ask the registrar to FAX you a copy of the registration form (your FAX #: 408-632-3000). Find out whether you can pay in traveller's checks when you arrive. Get the registrar's FAX number so you will be able to return the completed forms quickly.

Registration payable in traveller's checks upon arrival? _____

FAX # of registrar - _____

2. Another student and you want to share a room at the conference. Find out from the registrar which conference hotel that has subway access to the conference site has the cheapest twin room. You will need to indicate the name of your hotel selection on the registration form when you receive it.

Name of preferred hotel - _____

(with subway access and cheapest twin room)

Researchers' Instructions During Noninterpreted Call-

Same as instructions for interpreted calls, except the last two sentences of paragraph one were substituted for:

“The Japanese conference registrar speaks English very well, so you will be able to speak with the registrar directly in English to make your arrangements.”

General instructions, reference information, and contact report form for the registrars in the interpreted and noninterpreted telephone call conditions

Japanese-Speaking Registrar's Instructions and Reference Information for Interpreted Calls³.

“You have been working as the conference registrar for the Pacific Rim International Conference on Artificial Intelligence '90 in Nagoya, Japan. You are representing the PRICAI '90 Secretariat, and are located at that address (Akasaka Yamakatsu Building in Tokyo). The date is today's actual date. You will be receiving phone calls from American researchers who wish to register for the conference. These people may also have other questions about the conference or travel arrangements, and their requests and preferences will all be different. Using the information sheets that the conference organizers have provided you, please help these people with their requests by answering their specific questions and helping them get registration materials. None of these researchers have any information about the conference yet. If any of them ask for information not listed on your sheet, it is okay to say that you don't know or that you'll find out the information and forward it.

The American researchers only speak English, and do not understand Japanese, so a professional Japanese-English interpreter will be on the telephone line to assist the two of you as you make arrangements. The interpreter will do all she can to assist the two of you with your task.

As you assist the American researchers in making arrangements, be sure to write down any information that you will need to follow through in handling each request properly. A “Contact Report Form” is provided for you to record the reason for each call, and any information that you will need to process that person's request.

If you have any questions, please ask them now.”

³This information was presented to the Japanese-speaking registrar in her native Japanese language.

CONFERENCE INFORMATION SHEET

Pacific Rim International Conference on Artificial Intelligence '90
November 14 - 16, 1990
Shiratori Century Plaza
Nagoya, Japan

PRICAI'90 Secretariat:
c/o Inter Group Corp.
Akasaka Yamakatsu Bldg.
8-5-32, Akasaka, Minato-ku
Tokyo, 107 Japan
Phone: 03-479-5535
Fax: 03-479-7433
Telex: 02425181 IGC J

REGISTRATION:

Registration forms are available and can be sent by airmail or facsimile. They should be returned to the Secretariat by September 15, 1990.

FEES:

General Participants:

30,000 yen (through September 15)
40,000 yen (late registration - September 16 or later)

Students (discounted):

10,000 yen (through September 15)
15,000 yen (late registration - September 16 or later)

Banquet:

10,000 yen per person

A copy of the proceedings is included in the fee. All registration fees must be paid in Japanese yen by bank remittance. Personal or traveler's checks will not be accepted, nor will credit cards be accepted for payment of registration fees to the Secretariat. The bank's name and address, and the account number to which registrants should wire the money will be provided on the registration form, along with instructions for prepaying all registration fees. Then a copy of the bank receipt should be returned as remittance to the Secretariat along with the completed registration form. Upon receiving the registration form and payment, a registration card and receipt will be issued. Cancellations must be made in writing to the Secretariat. Refunds will be made as follows:

80% — before September 30
50% — before October 30

HOTELS

No.	Name of Hotel (address)	Hotel Rate (Tax and Service charge included)		Access to Conference site	Childcare Availability	Services
		Twin room	Single room			
1	Nagoya Kanko Hotel 1-19-30, Nishiki, Naka-ku, Nagoya 460 TEL 052-231-7711	Y 15,000	Y 11,000	20 minutes by car (no subway access)	Yes, infants from 6 wks. and older	-restaurant and laundry service -pool and athletic facility -next to large park and playground
2	Hotel Castle Plaza 4-3-25, Meieki, Naka-ku, Nagoya 450 TEL 052-582-2121	Y 11,900	Y 8,700	20 minutes by car or 4 subway stops	Yes, 2 year olds and older	-restaurant and laundry service
3	Nagoya Kanayama Washington Htl. 4-6, Kanayama, Naka-ku, Nagoya 460 TEL 052-322-1111	Y 13,000	Y 7,400	8 minutes by car or 1 subway stop	none	-restaurant

Rates listed are discounted conference rates. They include tax and service charge, but not meals. Reservation requests should be sent to the Secretariat by September 15. After that date, please contact hotels directly for reservations. No prepayment is necessary; all hotel fees may be payed upon check-out. Credit cards are acceptable.

TRANSPORTATION TO CONFERENCE SITE

Japanese Arrival Airport	Connecting Route	Type of Transport	Total Estimated Cost	Total Estimated Travel Time (Airport to Conference Site)
Narita (Tokyo)	Tokyo - Nagoya	bullet train (operates hourly)	Y 8,000	5 hours
Narita (Tokyo)	Tokyo - Nagoya	airplane (propeller planes operate once daily at 1 pm)	Y 12,000	2 hours
Osaka	Osaka - Nagoya	bullet train (operates hourly)	Y 5,000	3 hours

No international flights arrive in Nagoya, so conference attendees are advised to fly into either Osaka or Narita Airport.

Those wishing to participate in the group tour of Kyoto should plan round trips between the U.S. and Osaka airport. More information about both the tour and about travel connections within Japan are available from the Conference Secretariat, and can be airmailed or faxed to participants.

Contact Report Form

Telephone Caller

Reason for Call

Information Needed
to Process Request

Bilingual Registrar's Instructions and Reference Information for Noninterpreted Calls-

Same as general instructions and reference sheets for registrar handling interpreted calls, except that materials were presented in English, and the following was substituted for paragraph two:

"The American researchers only speak English, so you should conduct the entire conversation in English. In other words, you'll be having a direct, two-person call exclusively in English."

General pre-study orientation letter to interpreters, in advance of participating in the interpreted telephone call condition

"This letter is to provide you with some general orientation regarding the SRI International calls that you will be receiving during the next month, all of which will involve English-speaking American researchers who need to get the registration forms for an upcoming conference in Nagoya, Japan. These researchers will be needing your assistance to communicate with the Japanese-speaking conference registrar, who is representing the Conference Secretariat. In addition to needing the conference registration forms, individual researchers also may have other questions about hotel accommodations or travel connections. If so, please help them with any additional questions they might have.

The conference is the Pacific Rim International Conference on Artificial Intelligence '90, to be held at the Shiratori Century Plaza in Nagoya, Japan from November 14-16, 1990. Registration forms for the conference can either be sent out by airmail or facsimile, and completed forms also should be returned to the Conference Secretariat by airmail or FAX before the deadline indicated on the form. The registrar also has information available on the conference hotels and travel connections to Nagoya.

Once you have read this orientation letter, please return the enclosed card in the stamped envelope provided, so that we know you have received it. Thanks in advance for your assistance with these calls."

B Transcription Samples for Interpreted and Noninterpreted Telephone Calls

This appendix includes transcription samples for part of task 1 for subject 11's interpreted and noninterpreted calls. For both types of call, the speech of each of the participants is listed in their respective columns. The conversational time line is represented from the top to bottom of each page so that the sequencing of each speaker's contribution is clear. In places where simultaneous speech occurs between two or more speakers during English subdialogues, it is listed on parallel lines in the transcript. Likewise, the approximate timing of listener backchannel confirmation is indicated by starring its location in the speaker's running speech, and then placing the confirmation in parallel. In transcripts of the interpreted calls, English translations are listed beneath the Japanese. The transcript heading also indicates which of the three interpreters handled the call. To protect anonymity, subject 11's name has been altered.

The following conventions have been used to indicate special transcription features:

- ← is a left arrow in the interpreter's column signifies that the interpreter is beginning an English subdialogue; → is a right arrow signifies shifting to Japanese
- ?XXXXXX? designates speech judged to be unintelligible by both the transcriber and one of the experimenters, with each cluster of capital Xs representing a separate syllable or word
- [] square brackets around parallel sections of both English speakers' dialogue indicate simultaneous speech
- * In English subdialogues, stars indicate the placement of backchannel confirmations from the listener during a pause in the main speaker's speech; in Japanese subdialogues, stars indicate either the placement of listener confirmations during a pause, or the end of a segment of speech that overlaps with the listener's phrase or confirmation (i.e., no precise marking of simultaneous speech is provided for Japanese subdialogues)
- () brackets around translated pronouns in the Japanese subdialogues indicate pronouns that were elided in Japanese, but filled in during translation to English

English Speaker	Interpreter	Japanese Speaker
Hello?	↵	
	Hello, this is Japanese interpreter six-five-four. May I help you?	
Yes, I need to, uh register for the Pacific Rim International Conference * [on] Artificial Intelligence.	Yes, all right. [?XX?]	
	All right. Can I ask your full name, please?	
Harry.		
	Could you spell, please?	
H-A-R-R-Y.		
	Uh huh.	
L.		
	L.	
Shoemaker.		
	Shoemaker. Is this, ah your last name?	
Yes, it is.		
	Okay, Shoemaker, hh?, all right. Then, ah would you like to have the, ah form by, ah FAX or by airmail, or...	
Would you please send the form by FAX?		

English Speaker	Interpreter	Japanese Speaker
And, and let me give you the FAX number.	FAX. okay.	
	Yes, can I have a FAX number, please?	
Four-zero-eight-,	Four-zero-eight-,	
eight-zero-three-,	eight-zero-three-,	
four-five-four-six.	four-five-four-six. Okay, I will repeat. Your name is Harry L. Shoemaker. Then FAX number is four-oh-eight-, eight-oh-three-, four-five-four-six.	
Mhmm.	Okay, just a moment.	
	もしもし。 Hello.	⇒
	はい、またよろしくお願いたします。	はい。 Yes.
	Yes, hello again.	よろしくお願いたします。 Hello.

English Speaker	Interpreter	Japanese Speaker
	あのういまこちらの男性の方が ですね、*あのう用紙を、会議 場に関する用紙を、ファックス で送っていただきたいとおっ しゃっておりますから、*ここ にお名前をいかが、あの、伺い ましたから*お願いします。ま ずお名前のほうが、ラリー、 H-A、 <i>Um, now the man here says, um, (he) would like (you) to please send the forms pertaining to the conference place, by FAX, so, here, would (his) name, um, (I) have requested it, so would (you) please, starting with (his) first name, Harry. H-A-,</i>	はい。 はい、わかりました。 はい。 Yes. Yes. (I) see. Yes. H-A、 H-A-, R-Y、 R-Y. R-Y、 R-Y.
	これ、そしてミドルのイニシャ ル、あの真ん中にあるイニシャ ルがありまして、L、* <i>This, and the middle initial, um there is an initial in the middle, L.</i>	はい。L、 Yes, L.
	はい、そしてご名字が、* シューメイカスベル* いいます けども、S-H-O、 <i>Yes, and (his) last name, Shoemaker, (I)'ll spell it, it's S-H-O-,</i>	はい。 はい。 Yes. Yes. S-H-O、 S-H-O-,

English Speaker	Interpreter	Japanese Speaker
	E、 E-	E、 E-
	M-A、 M-A-	M-A、 M-A-
	K-E-R です。 and K-E-R.	K-E-R ですね。 * あ、わかり ました。はい。 *
	E-R、シューメイカー これが名字ですね。 E-R, Shoemaker. This is (his) last name, okay?	K-E-R, right? Ah, (I) see, yes.
	はい、そしてファックスのナン バーが、よん-ゼロ-はち、 Yes, and the FAX number is four-zero-eight-	よん-ゼロ-はち、 Four-zero-eight-
	はち-ゼロ-さん、 eight-zero-three-	はち-ゼロ-さん、 eight-zero-three-
	よん-ごお-よん-ろくです。 four-five-four-six.	よん-ごお-よん-ろくです ね。 four-five-four-six, right?

Researcher	English Speaking Registrar
Hello?	Hello?
Hi, I'd like to register for the Pacific Rim International Conference on n' Artificial Intelligence * in Nagoya.	Okay.
And I need to, ah have you FAX me a copy of the registration form, if you may.	Okay.
Yes, my name is Harry [L.]	Okay, could you, could I have your name and FAX number?
Yeah, Harry L. Shoemaker, S-H-, O-E-,	[Har'], Harry.
M-A-K-E-R.	S-H-, O-E-?
Okay, and my FAX number * is four-zero-eight-,	M-A-K-E-R.
six-three-zero-,	Yes,
five-four-[four-four].	four-zero-eight-,
Mhmm.	six-three-zero-,
[Okay.]	[five]-four-four-four?
Now I need to find out some information also.	Okay, we'll send it, [eh] immediately.
Can you tell me how much the registration fee is. per person, for those who are <i>not</i> attending the banquet?	Okay.