音声対話コーパスにおける画像情報のアノテーション手法の検討

前田　真季子、堀内　靖雄、市川　熹
千葉大学

概要

本論文では、日本語地図課題対話コーパスで用いられた音声の書き起こし手法を拡張し、ジェスチャー情報を音声現象との時間的関係において表現するアノテーション手法を提案する。また、データの交換用として、それらを TEI 形式で表現する手法を提案し、記述する。さらに、予備的検討として、うなづき、視線一致などが数多く生じることが期待されるディベート課題による対話と共通の話題による自由対話の収録を行ない、本アノテーション手法を適用した。

Annotation of Gestures and Transcription for Speech Dialogues

Makiko Maeda, Yasuo Horiuchi, Akira Ichikawa
University of Chiba

abstract

In this paper, we will introduce an annotation method for describing gestures made by speakers in speech dialogues in terms of their relationship in time. This method is based on the transcribing method in Japanese Map Task Dialogue Corpus. Equivalent information to the annotation of transcription and gestures represented by this method can be also represented in the TEI P4 format for data exchange between other researchers. As a preliminary investigation, we annotated the two new recorded dialogues between a pair of good friends; one is a debate about the theme which they have different opinions and another is a free talk about the event when they met first.

１．はじめに

音声に関する様々な分野においての研究が進む中、これまでにも多種多様な音声対話コーパスが作成されてきた。そしてそれに伴い、各コーパスに対して様々なタグ付け方法が考案されている[１]。そんな中、人間のジェスチャーが対話にどのように関係するかといった研究が多く見られるようになり、タグ付けとして画像情報を付加するものをよく目にする[２][３][４][５][６]。しかし、それらのジェスチャーのタグは、発話の言語情報と詳しく同期されていないのがほとんどである。

そこで、本論文では日本語地図課題対話コーパス[７]で用いられた書き起こし手法を拡張し、画像情報における動作のタイミングがそのときの言語情報との時間的な関係の中で明確に表現可能な音声・画像アノテーション手法を提案する。そのとき、同期関係の測定には、タグ付け作業者の負担を軽減するため、自動音素境界検出技術［８］［９］の利用を検討する。また、この音声・画像コーパス情報を交換するための表現方法として、TEI 形式を採用する。さらに、これらの提

案をもとに、実際に対話の収録を行ない、その実験データの書き起こし作業を行なった。

2. 日本語地図課題対話コーパスにおけるアノテーション手法

日本語地図課題対話コーパスの実現にあたり、音声対話のアノテーション手法についてはすでに提案を行なった[7]。本章ではそこで採用された音声の書き起こしのアノテーション手法について概説する。書き起こしテキストのアノテーション方式には互いに等価な内容を持つ二つの形式(ブラウジング形式、交換用形式)が存在する[１０]。

2.1 ブラウジング形式

ブラウジング形式は、研究者が音声対話の分析を行なう際などに、話者交替や重複現象が視覚的に良く分かるように配慮された形式となっている（図１参照)。発話の書き起こし情報以外に、基本的情報タグとして、以下のものが表現されている。

- 発話単位の話者識別子、開始時刻、終了時刻
- 発話内の無音区間の時間長
- 発話の重複位置、ラッチング発話

書き起こしテキストを作成する場合、その単位は重要である。従来、文や発話という単位が用いられることがあったが、これらは、具体的かつ客観的に決定できる単位ではない。そこで、日本語地図課題対話コーパスでは、転記の単位として、「自己発話内の 400 ミリ秒以上の無音区間によって区切られた音声的連続」という発話単位を提案、採用した[7]。さらに、時間的に前後する２つの発話単位について、その時間的関係を以下の 4 種類に分類し、表現している。

・継続
ある話者の発話単位が終了後、相手話者が次の発話単位を開始する前にその話者が次の発話単位を開始する場合。

・交替
ある話者の発話単位が終了後、その話者が次の発話単位を開始する前に相手話者が発話単位を開始する場合。

・準重複(ラッチング)
ある話者の発話単位の終了直前(100 ミリ秒以内)に相手話者が発話単位を開始する場合。

・重複
準重複以外の状態で 2 人の話者の発話単位が重複する場合。

上記のうち、上の 3 つは先行発話を１行で記述後、空行を１行挿入し、続けて後続発話を１行で記述する。ただし、準重複では先行発話の最後と後続発話の最初に`+'を記入する。また、重複については、空行をあけずに 2 つの発話単位を記述し、先行発話と後続発話の同期位置を文字単位(モーラ単位)であわせ、重複が開始する場所にそれぞれ`＊'(アスタリスク)を記入する。

```
00:29:168-00:31:456 G:でだいたいしたからやく

00:32:144-00:33:792 G:いっせんちぐらいのところのところで

00:33:952-00:34:144 F:はい

00:34:768-00:36:448 G:うせ<128>えとみぎにまがっ<112>てください+

00:36:448-00:36:832 F:+あはい

00:37:024-00:37:680 G:みぎに<208>*はい
00:37:520-00:38:736 F:                *はい<240>はいみぎにいきました
```

図１　転記テキストの例

2.2 TEI 形式(交換用形式)

ブラウジング形式は人間が見やすいという利点があるが、計算機で扱うにはそのデータフォーマットが適切ではない。そこで、研究者間のデータの交換を目的として、日本語地図課題対話コーパスでは、TEI 形式を採用した[７]。TEI とは SGML アプリケーションの一つとして開発された方式であり、様々な種類の文書の形式を定義することが可能である[１１]。ブラウジング形式の転記において、発話レベルでの表現形式は以下のようなタグに翻訳されている。

発話単位：<u>タグ

属性：id,who,start,end

重　　複：重複位置を<anchor>タグでマーク

準 重 複：<u>タグの属性として latching

このうち、anchor には属性である id を与え、その id を参照するリンクをはる。重複の際、そのリンクでは同期の関係のデータベースが記述されていることになる。また、属性の who は発話者、start は発話開始時刻、end は発話終了時刻を表す（図２参照)。

```
<u id=j1n1u00190k who=j1na1 start=j1n1u0190s end=j1n1u0190e>
でだいたいしたからだいたいやく</u>
<u id=j1n1u00200k who=j1na1 start=j1n1u0200s end=j1n1u0200e>
いっせんちぐらいのところで</u>
<u id=j1n1u00210k who=j1na2 start=j1n1u0210s end=j1n1u0210e>
はい</u>
<u id=j1n1u00220k who=j1na1 start=j1n1u0220s end=j1n1u0220e>
うせ
<pause id=j1n1p00300k dur='128'>
えとみぎにまがっ
<pause id=j1n1p00310k dur='112'>
てください</u>
<u id=j1n1u00230k who=j1na2 start=j1n1u0230s end=j1n1u0230e trans=latching>
あはい</u>
<u id=j1n1u00240k who=j1na1 start=j1n1u0240s end=j1n1u0240e>
みぎに
<pause id=j1n1p00320k dur='208'>
<anchor id=j1n1a005gk>
はい</u>
<u id=j1n1u00250k who=j1na2 start=j1n1u0250s end=j1n1u0250e>
<anchor id=j1n1a005fk>
はい
<pause id=j1n1p00330k dur='240'>
はいみぎいきました</u>

<!--END OF THE FRAGMENT-->

<!--XPTRs to a digital sound file -->

<xptr id=j1n1u0190s doc=j1n1sound from='space (1d) (00:29:168)'>
<xptr id=j1n1u0190e doc=j1n1sound from='space (1d) (00:31:456)'>
                    <！――中略――>
<xptr id=j1n1u0250s doc=j1n1sound from='space (1d) (00:37:520)'>
<xptr id=j1n1u0250e doc=j1n1sound from='space (1d) (00:38:736)'>

<!--Temporal Synchronization-->

<linkGrp type='temporal synchronization'
        targetType='anchor anchor'
        domains='j1n1k j1n1k'>
<sync id=j1n1sync0001 targets='j1n1a005gk j1n1a005fk'></linkGrp>
```

図2　TEI 形式テキストの例

2.3 改善点

準重複は発話が交替していると考え、空行を挿入して表現したが、物理的には重複現象であるため、本研究では、準重複を重複と同様に扱うこととした。重複との表現上の唯一の違いは、記号が‘＊’ではなく‘＋’を用いる点だけとする。

3. ジェスチャーのアノテーション手法の提案

音声対話は二人の話者により行なわれるため、二つの音声現象が同時に起こりうる。前章では、それらを書き起こす手法を概説した。本研究では、ジェスチャーを音声現象と同様に、対話と同時に発生するイベント情報であると考え、前章で解説した手法を拡張することにより、ジェスチャーイベントを音声イベントとともに、統一的にアノテートする手法について提案する。

3.1 ジェスチャーイベントの時間的情報

音声対話における諸現象を分析するという立場から考えると、ジェスチャーイベントについては、その発生時刻がわかるだけでは不十分であり、その動作が音声のどのタイミングで発生したのかを知ることが重要である。よって、音声の重複現象と同様、モーラ単位でその発生位置を書き起こすことを検討した。しかし、動作の発生が発話単位中のどの位置かを手作業で調べていくのは時間がかかり、大変な作業である。そこで今回、手作業で行なうのは、通常の書き起こしテキスト（ブラウジング形式）と各ジェスチャーイベントの時刻情報を記録したファイルの作成のみとし、両者の同期関係の計測は自動音素境界推定手法[８][９]を利用して実現することとした。

なお、ジェスチャーイベントの時刻情報は、フレーム単位であるため、小数点以下を四捨五入し音声の単位に合わせた。

3.2 ジェスチャー表記方法(ブラウジング形式)

各ジェスチャーについて、記述すべき情報は以下のものが挙げられる。

・　動作者
・　ジェスチャーの種類
・　イベント発生時刻
（・　イベント終了時刻　）

ここで、今回イベント終了時刻はオプショナルとして考えることとし、イベントの継続に意味がある場合のみ記述する。

転記方法としては、音声の発話単位と同様、先頭から開始時刻(終了時刻)、動作者、ジェスチャーを表す表記記号を一行に記録することとした。

3.2.1 表記記号の種類

本研究では、目の動きと顔の動きを対象とし、各動作について、ジェスチャーを表わす表記記号を以下のように決定した。なお、これらの分類は現時点で暫定的に決定したものであり、研究の必要性などを考慮し、再検討する必要がある。

＜eye＞

・look（目を見る）　・avert（目をそらす）

＜head＞

・nod（うなづく）　・left（左を向く）
・right（右を向く）　・up（上を向く）
・down（下を向く）　・tilt（首をかしげる）
・shake（首をふる）　・normal（元に戻す）

なお、元に戻すという動作は、首をかしげたり、横を向いた場合の、元の状態に戻そうとする動きすべてをいう。また、表記記号を記録する際には、eye-look、head-nod のように表記する。さらに、発話との対応関係を表わす際には、‘ ＊ ’のかわりに‘ ^ ’でマークする。

3.3 TEI 形式での表現方法

本章では、上述のブラウジング形式の音声・ジェスチャー書き起こしファイルの研究者間での交換を目的とし、TEI 形式で表現する手法について提案する。ジェスチャーは発話タグ <u> と同様に <g> タグを用意する。<g>タグがもつ属性を以下に示す。

who: 動作者
object:ジェスチャーの対象部位（目や顔）
type: 動きの種類（見る、そらす・・）
start: イベント発生時刻
（ end: イベント終了時刻）

また、時刻の同期関係は、重複と同様に <anchor> タグを用いてリンクを張って表現する。なお、object、type に記述する表記記号は転記ファイルに対応した記号である。

ブラウジング形式、および、TEI 形式の具体例は第５章で示す。

4. 対話・ジェスチャー収録

本研究では自然対話におけるジェスチャーを分析するため、自然対話の動画像による収録を検討している。本章では予備的に行なった収録について概説し、次章で第 3 章の表現方法を用いて実際に書き起こした例を示す。

4.1 収録設備

図３のような防音室を使用した。防音室は中央の仕切りによって区切られ、それぞれが独立した防音室になっている。仕切りの一部は窓ガラスが設置されており相手の顔や動きを見ることができる。部屋にはそれぞれ椅子が用意されており、二人の被験者は窓越しに向かい合って座った。自然な対話が行なえるように二人の間には窓ガラス以外、小型カメラを置く最低限の台を置いただけで、互いの距離ができる限り近くなるようにした。

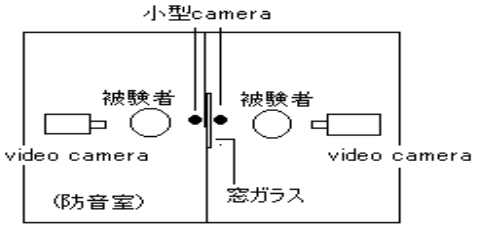

図３　防音室の収録設備

4.2 音声情報の収録

今回の実験では、お互いのいる部屋を別々に仕

切り、相手の音声がマイクに回りこむのを防ぐことで、二人の被験者の音声を完全に分離して録音した。音声のやり取りはイヤホンとピンマイクを用いた。ヘッドセットの使用も検討されたが、ヘッドセットでは頭にある程度の重さのものをかぶることになるので、自然対話の妨げになり、自然な顔の表情が出にくくなる恐れがあると考えた。イヤホンは片耳だけの装着とし、相手の声だけが聞こえてくるようにした。

4.3 画像情報の収録

アイコンタクトやうなずきといった顔の情報を記録するため、それぞれの被験者の正面にカメラを設置した。ビデオカメラの存在が発話の自然さの妨げにならないようにカメラは非常に小さな CCD カメラを用いた。さらに、顔の表情だけでなく、手や体の動きも観察するために、カメラを被験者の後方に設置し、窓ガラスをはさんで向こう側にいる被験者の腰から上の映像を記録した。このとき、窓ガラスをはさんでの録画のため、向かい側に座っている被験者の姿がガラスに映りこんでしまう傾向が見られた。これを少しでもなくすために、被験者にはできる限り黒い服を着てもらった。

画像情報は全部で 4 つあるが、画面を 4 分割しそれぞれの映像を同時に録画した。これにより、各動きを同時に観察する事が可能となった。

画像、音声はともにデジタルビデオデッキで記録された。画像は NTSC、音声は量子化ビット数 16bit、サンプリング周波数 48kHz である。

4.4 実験参加者

今回協力してもらった被験者は、千葉大学の男子学生２名である。この２名は日頃から仲のよい友人同士であり、収録という環境においても、ある程度自然な対話を行なうことが期待される。

4.5 収録実験の内容

ディベート形式の対話と、共通のテーマを話題にした自由対話の 2 対話を収録。両方とも１０分間の時間制限を設け、１０分経過後に実験者が対話を打ち切った。

4.5.1 ディベート形式

前もって、被験者二人に意見が分かれそうな２者択一の質問に答えてもらった。この時、互いに相談することは禁止し、きちんと理由を考えて答えを選んでもらうように指示した。そして、二つのアンケートを比較し、意見が分かれたものに対して二人で討論してもらうようにした。今回ディベート形式を試みたのは、テーマを与えることによって、会話が途切れず、自分の考えを相手に訴えかける姿勢や、相手の意見を聞く姿勢が現れ、視線一致やうなずき、身振り手振りの動作が数多く表出されることが期待できるという意図からである。

4.5.2 共通話題の自由対話

二人の被験者が仲のよい友達同士ということから、どのようなつながりで知り合いになったのかを聞き、そのきっかけをテーマにして、対話をしてもらった。これにより、普段行なう会話に近い自然な対話が期待できる。

例：サークル、バイト、学科、etc・・・

4.6 実験結果

2 種類の対話収録とも、期待していたような自然な対話が行われた。ディベート対話では、前もって行なったアンケート結果から、「男と女どちらが得か？」をディベートのテーマとした。

収録の結果、目の前の小型カメラにより、視線一致やうなずきなど、必要な情報を得ることができた。しかし、体全体をガラス越しに撮影する映像では、相手の姿が反射して映りこんでいるものが多く、体の動きなどの、ジェスチャーの抽出が困難であった。今後、どのようにするかを検討していく必要がある。

5. ジェスチャーのアノテーション例

今回予備収録を行なった実験データについて

書き起こしテキスト、ジェスチャーイベントファイルを作成し、自動音素境界検出により音声対話とジェスチャーの同期をとった。それぞれのファイル、および、音声・画像書き起こしテキスト(ブラウジング形式、TEI 形式)の一部を以下に示す。

5.1 書き起こしテキスト,ジェスチャーイベントファイル

```
03:28:850-03:29:820 G:ってかんがえるとねやっ<130>ぱ

03:30:630-03:31:440 F:おとこの*ほうが
03:31:100-03:32:350 G:          *おとこのほうとくなんちゃうん

03:32:800-03:34:450 G:しごとめんでも*こううえにいける
03:33:580-03:33:840 F:                *ああ

03:35:240-03:36:060 G:うえへうえへいける

03:36:240-03:37:310 F:しょうらいをみすえるとね

03:37:860-03:38:120 G:*うん
03:37:890-03:39:080 F:*いまは<120>たかだ<150>からばいと
```

図４ 書き起こしテキスト(ブラウジング形式)

```
03:28:09 A:head-nod
03:28:22 A:head-tilt
03:28:27 A:head-normal
03:29:12 A:eye-look
03:29:12 B:eye-avert
03:30:15 B:head-nod
03:30:27 A:eye-avert
03:30:29 A:head-tilt
03:31:22 A:head-normal
03:32:02 A:eye-look
03:32:04 A:head-tilt
03:32:16 A:head-normal
03:32:21 A:eye-avert
03:32:21 A:head-tilt
03:32:27 A:head-normal
03:33:00 B:eye-look
03:33:05 A:head-tilt
03:33:21 B:eye-avert
03:33:25 A:head-normal
03:34:01 A:eye-look
03:35:17 B:eye-look
03:36:03 B:eye-avert
03:37:00 B:eye-look
03:37:03 B:head-nod
03:37:16 A:head-nod
03:37:19 A:eye-avert
03:37:21 B:eye-avert
03:37:26 B:head-left
03:38:03 A:eye-look
03:38:13 B:head-normal
03:38:23 B:head-nod
03:39:02 A:head-nod
```

図５ ジェスチャーイベントファイル(フレーム単位)

5.2 ブラウジング形式の対話・ジェスチャー書き起こしファイル

```
03:28:300          A:head-nod

03:28:733          A:head-tilt

03:28:850-03:29:820  A:ってかん^がえるとねやっ^^<130>ぱ
03:28:900          A:        ^head-normal
03:29:400          A:                  ^eye-look
03:29:400          B:                  ^eye-avert

03:30:500          B:head-nod

03:30:630-03:31:440  B:^^おとこの*ほうが
03:31:100-03:32:350  A:        *おとこのほうとくな^んちゃう^^ん
03:30:900          A:^eye-avert
03:30:967          A: ^head-tilt
03:31:733          A:                  ^head-normal
03:32:067          A:                  ^eye-look
03:32:133          A:                  ^head-tilt

03:32:533          A:head-normal

03:32:700          A:eye-avert

03:32:700          A:head-tilt

03:32:800-03:34:450  A:し^ご^と^めんでも*こうう^えに^いけ^る
03:33:580-03:33:840  B:              *あ   ^   ^あ
03:32:900          A:  ^head-normal
03:33:000          B:    ^eye-look
03:33:167          A:        ^head-tilt
03:33:700          B:                  ^^eye-avert
03:33:833          A:                  ^^head-normal
03:34:033          A:                  ^eye-look

03:35:240-03:36:060  A:うえへうえ^へいける
03:35:567          B:        ^eye-look

03:36:100          B:eye-avert

03:36:240-03:37:310  B:しょうらいをみすえると^^ね
03:37:000          B:                  ^eye-look
03:37:100          B:                  ^head-nod

03:37:533          A:head-nod

03:37:633          A:eye-avert

03:37:700          B:eye-avert

03:37:860-03:38:120  A:*^う^ん
03:37:890-03:39:080  B:* い^まは<120>たかだ<150>か^らばい^^と
03:37:867          B:  ^head-left
03:38:100          A:   ^^eye-look
03:38:433          B:                  ^head-normal
03:38:767          B:                  ^head-nod
03:39:067          A:                  ^head-nod
```

図６ 対話,ジェスチャー書き起こしファイル

5.3 TEI 形式ファイル

```
<g id=jgp001g0013 who=jgp001a object=head type=nod start=jgp001g0013s>
head-nod</g>
<g id=jgp001g0014 who=jgp001a object=head type=tilt start=jgp001g0014s>
head-tilt</g>
<u id=jgp001u0021 who=jgp001a start=jgp001u0021s end=jgp001u0021e>
ってかん
<anchor id=jgp001h0001a>
がえるとねやっ
<anchor id=jgp001h0002a>
<anchor id=jgp001h0003a>
<pause id=jgp001p0001 dur=130>
ぱ</u>
<g id=jgp001g0015 who=jgp001a object=head type=normal start=jgp001g0015s>
<anchor id=jgp001h0001g>
head-normal</g>
<g id=jgp001g0016 object=eye type=look who=jgp001a start=jgp001g0016s>
<anchor id=jgp001h0002g>
eye-look</g>
<g id=jgp001g0017 who=jgp001b object=eye type=avert start=jgp001g0017s>
<anchor id=jgp001h0003g>
eye-avert</g>
<g id=jgp001g0018 who=jgp001b object=head type=nod start=jgp001g0018s>
head-nod</g>
<u id=jgp001u0022 who=jgp001b start=jgp001u0022s end=jgp001u0022e>
<anchor id=jgp001h0004b>
<anchor id=jgp001h0005b>
おとこの
<anchor id=jgp001a0003b>
ほうが</u>
<g id=jgp001g0019 who=jgp001a object=eye type=avert start=jgp001g0019s>
<anchor id=jgp001h0004g>
eye-avert</g>
<g id=jgp001g0020 who=jgp001a object=head type=tilt start=jgp001g0020s>
<anchor id=jgp001h0005g>
head-tilt</g>
<u id=jgp001u0023 who=jgp001a start=jgp001u0023s end=jgp001u0023e>
<anchor id=jgp001a0003a>
おとこのほうとくな
<anchor id=jgp001g0006a>
んちゃう
<anchor id=jgp001h0007a>
<anchor id=jgp001h0008a>
ん</u>
<g id=jgp001g0021 who=jgp001a object=head type=normal start=jgp001g0021s>
<anchor id=jgp001h0006g>
head-normal</g>
<g id=jgp001g0022 who=jgp001a object=eye type=look start=jgp001g0022s>
<anchor id=jgp001h0007g>
eye-look</g>
<g id=jgp001g0023 who=jgp001a object=head type=tilt start=jgp001g0023s>
<anchor id=jgp001h0008g>
head-tilt</g>
<g id=jgp001g0024 who=jgp001a object=head type=normal start=jgp001g0024s>
head-normal</g>
<g id=jgp001g0025 who=jgp001a object=eye type=avert start=jgp001g0025s>
eye-avert</g>
<g id=jgp001g0026 who=jgp001a object=head type=tilt start=jgp001g0026s>
head-tilt</g>
<u id=jgp001u0024 who=jgp001a start=jgp001u0024s end=jgp001u0024e>
し
<anchor id=jgp001h0009a>
ご
<anchor id=jgp001h0010a>
と
<anchor id=jgp001h0011a>
めんでも
<anchor id=jgp001a0004a>
こうう
<anchor id=jgp001h0012a>
えに
<anchor id=jgp001h0013a>
いけ
<anchor id=jgp001h0014a>
る</u>
<u id=jgp001u0025 who=jgp001b start=jgp001u0025s end=jgp001u0025e>
<anchor id=jgp001a0004b>
あ
<anchor id=jgp001h0012b>
<anchor id=jgp001h0013b>
あ</u>
<g id=jgp001g0027 who=jgp001a object=head type=normal start=jgp001g0027s>
<anchor id=jgp001h0009g>
head-normal</g>
<g id=jgp001g0028 who=jgp001b object=eye type=look start=jgp001g0028s>
<anchor id=jgp001h0010g>
eye-look</g>
<g id=jgp001g0029 who=jgp001a object=head type=tilt start=jgp001g0029s>
<anchor id=jgp001h0011g>
head-tilt</g>
<g id=jgp001g0030 who=jgp001b object=eye type=avert start=jgp001g0030s>
<anchor id=jgp001h0012g>
eye-avert</g>
<g id=jgp001g0031 who=jgp001a object=head type=normal start=jgp001g0031s>
<anchor id=jgp001h0013g>
head-normal</g>
<g id=jgp001g0032 who=jgp001a object=eye type=look start=jgp001g0032s>
<anchor id=jgp001h0014g>
</g>
<u id=jgp001u0026 who=jgp001b start=jgp001u0026s end=jgp001u0026e>
うえへうえ
<anchor id=jgp001h0015a>
へいける</u>
<g id=jgp001g0033 who=jgp001b object=eye type=look start=jgp001g0033s>
<anchor id=jgp001h0015g>
eye-look</g>
<g id=jgp001g0034 who=jgp001b object=eye type=avert start=jgp001g0034s>
eye-avert</g>
<u id=jgp001u0027 who=jgp001b start=jgp001u0027s end=jgp001u0027e>
しょうらいをみすえると
<anchor id=jgp001h0016b>
<anchor id=jgp001h0017b>
ね</u>
<g id=jgp001g0035 who=jgp001b object=eye type=look start=jgp001g0035s>
<anchor id=jgp001h0016g>
eye-look</g>
<g id=jgp001g0036 who=jgp001b object=head type=nod start=jgp001g0036s>
<anchor id=jgp001h0017g>
head-nod</g>
<g id=jgp001g0037 who=jgp001a object=head type=nod start=jgp001g0037s>
head-nod</g>
<g id=jgp001g0038 who=jgp001a object=eye type=avert start=jgp001g0038s>
eye-avert</g>
<g id=jgp001g0039 who=jgp001b object=eye type=avert start=jgp001g0039s>
eye-avert</g>
<u id=jgp001u0028 who=jgp001a start=jgp001u0028s end=jgp001u0028e>
<anchor id=jgp001a0005a>
<anchor id=jgp001h0018a>
う
<anchor id=jgp001h0019a>
ん</u>
<u id=jgp001u0029 who=jgp001b start=jgp001u0029s end=jgp001u0029e>
<anchor id=jgp001a0005b>
い
<anchor id=jgp001h0019b>
まは
<pause id=jgp001p002 dur=120>
たかだ
<pause id=jgp001p003 dur=150>
か
<anchor id=jgp001h0020b>
らばい
<anchor id=jgp001h0021b>
<anchor id=jgp001h0022b>
と</u>
<g id=jgp001g0040 who=jgp001b object=head type=left start=jgp001g0040s>
<anchor id=jgp001h0018g>
```

```
head-left</g>
<g id=jgp001g0041 who=jgp001a object=eye type=look start=jgp001g0041s>
<anchor id=jgp001h0019g>
eye-look</g>
<g id=jgp001g0042 who=jgp001b object=head type=normal start=jgp001g0042s>
<anchor id=jgp001h0020g>
head-normal</g>
<g id=jgp001g0043 who=jgp001b object=head type=nod start=jgp001g0043s>
<anchor id=jgp001h0021g>
head-nod</g>
<g id=jgp001g0044 who=jgp001a object=head type=nod start=jgp001g0044s>
<anchor id=jgp001h0022g>
head-nod</g>

<!--以下アンカーとリンクグループによる記述-->

<!--オーバーラップ開始の同期情報-->
<linkGrp id='jgplnksyn01'
      type='synchronization'
      domains='jgp001 jgp001'
      targType='anchor anchor'>
  <!--前略--->
  <link id=jgpovlp0001 targets='jgp001a0003a jgp001a0003b'>
  <link id=jgpovlp0002 targets='jgp001a0004a jgp001a0004b'>
  <link id=jgpovlp0003 targets='jgp001a0005a jgp001a0005b'>
  <!--後略---></linkGrp>

<!--ジェスチャーとの同期情報-->
<linkGrp id='jgplnkges01'
      type='gesture'
      domains='jgp001 jgp001'
      targType='anchor anchor'
  <!--前略--->
  <link id=jgpges0001 targets='jgp001h001u jgp001h001g'>
  <link id=jgpges0002 targets='jgp001h002u jgp001h002g'>
        <！――中略――>
  <link id=jgpges0019 targets='jgp001h019a jgp001h019b jgp001h019g'>
  <link id=jgpges0020 targets='jgp001h020b jgp001h020g'>
  <link id=jgpges0021 targets='jgp001h021b jgp001h021g'>
  <link id=jgpges0022 targets='jgp001h022b jgp001h022g'>
 <!--後略--></linkGrp>
```

図７　　TEI 形式ファイル

6．おわりに

本論文では、音声対話におけるジェスチャーを発話の書き起こしとともに記述し、各ジェスチャーを発話との時間的関係の中で表現可能なアノテーション手法を提案した。また、それを TEI 形式でデータ表現する手法について検討を行ない、実際の対話に対し、音声、画像のアノテーションを行なった。今後はさまざまなタイプのジェスチャーを書き起こしながら、本手法の有効性、問題点などについての検討を行なう。

謝辞

TEI 表現に関する議論をしていただいた土屋先生に謝意を表します。

参考文献

[１]土屋俊,堀内靖雄,石崎雅人,前川喜久雄：音声対話コーパスの共有化へ向けて,人工知能学会誌,Vol.14,No.2,P231－241(1999)

[２]坂本憲治,綿貫啓子,外川文雄:マルチモーダル対話解析,人工知能学会研究会資料,SIG-FAI-9401,pp.39-46(1994)

[３]Tony Bigbee, Dan Loehr, Lisa Harpr：Emerging Requirements for Multi-Modal Annotation and Analysis Tools,Eurospeech-2001

[４]Michael Kipp：ANVIL A Generic Annotation Tool for Multimodal Dialogue, Eurospeech-2001

[５]Satoru Hayamizu, Shigeki Nagaya, Keiko Watanuki, Masayuki Nakazawa, Shuichi Nobe Takashi Yoshimura：A MULTIMODAL DATABASE OF GESTURES AND SPEECH,Eurospeech-2001

[６]Ulrich Turk：The Technical Processing in SmartKom Data Collection: a Case Study,Eurospeech-2001

[７]堀内靖雄,中野有紀子,小磯花絵,石崎雅人,鈴木浩之,岡田美智男,沖真紀子,土屋俊,市川熹:日本語地図課題対話コーパスの設計と特徴,人工知能学会誌 Vol.14,No.2,p261－272(1999)

[８]大須賀智子,堀内靖雄,市川熹:音素セグメンテーションの自動化に関する検討,『話し言葉』の科学と工学ワークショップ講演予稿集,pp.143-148(2001)

[９]大須賀智子,堀内靖雄,市川熹:音素境界認識の自動化の検討,音響学会講演論文集,pp.107-108,2001

[10] 堀内靖雄：音声対話コーパス作成、分析ツールについて~日本語地図課題対話コーパス作成の経験から~,人工知能学会研究会資料,SIG-SLUD-9902,pp.23-28

[11] http://www.tei-c.org/P4X/