PASD コーパス — 重点領域模擬対話音声コーパス —

板橋秀一

1　はじめに

音声研究を進める上で音声データが必要なことは言うまでもない。その音声データは、多種多様 (性別・年齢・方言・人数等) であることが求められる。従来は各研究者が、必要に応じて音声データを収録し、保管・利用していた。音声研究が進歩するにつれて処理可能なデータ量が増大し、そのため用意すべきデータ量が大幅に増してきた。最近では、特に統計的手法の発達により大量の音声データがシステムの学習のために必要とされるようになった。[1-4]　一方、音声情報処理システムの研究・開発を行うためには、分析・合成・認識の各種の手法を適切に比較・評価することが必要とされる。これを行う方法としては現在のところ、共通の音声データを用いてこれらの処理を行い、その結果を比較するという方法以外は知られていない。このようなことから、共通利用可能な各種・大量の音声データを収録し、保管・公開することは研究・開発過程での利用および認識システムの性能評価の両面から求められている。このような目的に利用される音声データを一般に音声データベースあるいは音声コーパスと呼んでいる。音声情報処理の分野では　「音声データベース」というときには、データベースシステムよりも「大量の音声データの集積」そのものを指すことが多い。そのため最近では、それを意味する「音声コーパス」を使うようになった。音声コーパスの必要性やその意義については近年広く認められるようになってきた。[5-8]

本重点領域研究を遂行する上でも音声対話コーパスは必須という立場に立ち、コーパスワーキンググループ (以下、 WG とする) が設けられた。本 WG の目的は、これから発展するであろう音声対話研究の基礎的なデータとして活用できる、ある程度の規模のコーパスを構築することである。本 WG のメンバーと活動概要を付録 A に示す。

2　音声対話コーパスの概要

3 年間の活動中に Vol.1 から 4 まで 4 枚の CD–ROM からなる音声対話コーパスを作成した。 Vol.1 は 1994 年度、 Vol.2 ~ 4 は 1995 年度に作成された。内容は、 Vol.1 ~ 3 が模擬対話を収録したもの、 Vol.4 が模擬対話、実際の対話、朗読、講演の音声を収録したものとなっている。

それぞれの対話データは、対話の音声を AD 変換でデジタル化した音声ファイル、その音声の情報をテキストによって表現した書き起こしファイル、さらに説明のためのドキュメントファイルからなっている。また、本 CD–ROM を活用するためのツールとして、京都大学で開発された音声ファイル再生プログラムと書き起こしテキストから音声を容易に再生できるブラウジングツールも入っている。

音声ファイルはいくつかの例外を除き、 7.1.3.2 節で述べるような形式で統一できた。しかし、書き起こしテキストは各機関の研究目的や割ける労力の割合の違いによって統一は困難であった。本 WG で一応の基準とした書き起こしの仕様は、多くの機関で準備可能とするために、比較的容易に書き起こしができるものとした。この仕様の中には書き起こしテキストから容易に元の音声を再生するためのツールと連携が取りやすい仕様も含んでいる。これからの音声対話の研究をより高度かつ厳密に行うためには、実際の音声にあたることは必要不可欠であると思われるが、これを容易にするためのブラウジングツールとの連携を取りやすい仕様は音声対話研究において利用価値の高いものと考えている。また、 TEI[9] のようなより厳密で、表現能力の高い仕様に基づく書き起こしを作成した機関もあり、そのデータも収録されているので、活用されるものと思われる。

3　音声対話コーパスの仕様

初年度にアンケート調査を行い、その回答に基づいて「対話コーパスの作成条件」をまとめた。アンケート項目と作成条件を付録 B に示す。

ここでは、本 WG で採用した基本的な仕様を述べる。各研究機関は原則としてこの基準に基づいてコーパスを作成したが、より先進的な立場に立つグループなどは、それぞれの仕様に任せた。各機関ごとの詳しい仕様は、 CD–ROM を参照されたい。それぞれの CD–ROM の協力機関と対話内容の簡単な説明を付録 C に記す。

3.1 名前の規則

ここでは、対話や収録機関などの名前を定義する。ここで定義された名前は CD–ROM 中のファイル名やデータ中の記述に用いられる。

対話番号：収録機関 3 桁 + 対話連番号 4 桁

cccmmmm　　ccc: 収録機関コード
　　mmmm: 対話連番号

*(例) 筑波大学の 3 番目の対話： tsu0003

話者番号：収録機関 3 桁 + 話者情報 2 桁 + 話者番号 2 桁

cccxyzz　　ccc: 収録機関コード
　　x: 性別 (0: 男,1: 女)
　　y: 年齢 (0:0-9,1:10-19,2:20-29,3:30-39,4:40-49,
　　5:50-59,6:60-69,7:70-79,8:80-89,9:90-99)
　　zz: 話者番号 (00-99)

*(例) 筑波大学の 3 番目の話者 (男性で 20 才)： tsu0203

収録機関コード：

chi	千葉大学
kan	関西大学
kyo	京都大学
osa	大阪大学
siz	静岡大学
tit	東京工業大学
tok	東京大学
tsu	筑波大学
tut	豊橋技術科学大学
uec	電気通信大学
was	早稲田大学

3.2 音声ファイルの仕様

音声データのフォーマットは以下の規則に従う。

- AD 変換は 16kHz、 16bit linear、 IEEE Byte order。
- ヘッダーは付与しない。
- 二人の話者の音声を左右チャネルに分割して収録。
- 一対話の音声は、時間順に重複のない複数のファイルに分割。
- 分割は、左右チャンネル共に無声区間になっている点で、左右同じ場所で切ることを原則とする (分割ファイルを番号順に接続することにより、チャネル同期した元の対話音声ファイルが再現できるようにする)。

音声ファイルに関しては、かなりの機関で統一されている。この仕様を満たしていないのは、自由対話で L と R の発話がかなり重なっており、同じところで切れなかった機関などである。

音声ファイル名は以下の規則に従う。英字に関してはすべて小文字とする。

対話番号 (7 桁)+ チャンネル記号 (1 桁). 分割ファイル番号

cccmmmm[lr].nnn

ここで、

ccc: 収録機関コード

mmmm: 対話連番号

[lr]: チャネル (l: 左,r: 右) の記号

nnn: 分割ファイル番号

である。

(例) 筑波大学の 3 番目の対話中の 2 つ目に切り出した音声の
右チャンネルのデータファイル名： tsu0003r.002

3.3 書き起こしテキストの仕様

書き起こしのテキストは以下の 7 種類が存在する。括弧の中に書いてあるのは書き起こしファイルのファイル名の規則である。

(1) 1 行 1 文書き起こし (*.trn)

「音声の知的処理に関する調査研究」書き起こしテキスト作成基準に基づく。 (おおよそ 1 行 1 文になっている)

(2) 1 行 1 文音声対応書き起こし (*.sp)

上記の「1 行 1 文書き起こし」ファイルに音声ファイルと対応させるための区切り記号と対応音声ファイル名を記述したもの。

(3) 1 行 1 ファイル書き起こし (*.syn)

1 行の書き起こしが正確に音声ファイル 1 つに対応し、かつスペースによるおおよその同期情報を含む書き起こし。

(4) 平仮名書き起こし (*.kan)

平仮名およびカタカナ表記の書き起こし。

(5) ローマ字書き起こし (*.rom)

JEIDA 正書法によるローマ字表記 (大文字) による書き起こし。

(6) TEI 書き起こし (*.tei)

TEI に基づく書き起こし

(7) その他の書き起こし (ファイル名は任意)

その他の書き起こし基準、記述法に基づく書き起こしテキスト

但し、 * は対話番号 (cccmmmm)

CD–ROM に入れる書き起こしファイルの原則は以下の通りである。

- (1) ~ (5) はすべて入れる。ただし、 (3) については同期情報がないものも認める。ただし、 1 行 1 ファイルの原則は必ず満たしていること。
- (6) は努力目標として、作成したならば入れる。
- (7) は情報量が (1) ~ (5) よりも多ければ入れる。

以下、 (1) ~ (5) の各書き起こしの仕様を説明する。

(a) 1 行 1 文書き起こし (*.trn)

「音声の知的処理に関する調査研究」[10] 書き起こしテキスト作成基準に基づいた、おおよそ 1 行 1 文になっている書き起こしテキストである。以下に、 [10] を若干修正して、本 WG で採用した書き起こし基準を以下に述べる。

(1) ヘッダー

ファイルの頭に内容を表す次のようなヘッダを付ける。

模擬対話 (書き起こし)

タスク: 観光・旅行案内

トピック: 筑波案内

対話番号:tsu0001
収録年月日:1991/3/24
A/B:tsu0201/tsu1202
テキスト作成日:1991/5/4

- タスクは「観光・旅行案内」、「学校・会社案内」、「その他」の 3 つから 1 つ選ぶ。
- トピックは内容を表すものを自由に書く。
- 対話番号の後には対応する対話番号を書く。
- 英字は話者ラベル以外は小文字に統一する。
- A/B の後には対応する話者番号を書く。話者ラベルに "A:", "B:" 以外を用いた場合は、それに応じて "A/B" を書き換えること。
- 上記のコードは漢字ひらがら以外はすべて半角とする。

(2) 文

- 一文一行とする。(文の終りで必ず改行をし、文の途中で改行しない。)
- 文の終りに「。」をつける。(疑問文でも？は使わない。)
- 文の認定は以下の基準による。
  - 原則として述語句を一つ以上含む。
  - 省略質問文に関しては述語句を含まなくてもよい。
    (例) お値段とか、具体的にわかれば。
  - 感動詞は単独で文になれる。

(3) 発話者ラベル

- 文の先頭には原則として発話者ラベルをつける。
- 次の文が同じ話者の場合、発話者ラベルは省略してよい。 ただし、インデントは揃える。
- 発話者ラベルは半角の "A:" と "B:" を原則とする。異なるラベルを割り当てる場合は、半角で各話者ラベルの長さが同じであること。
- 発話者ラベル "A:","B:" を用いる場合は、 "A:" が左チャンネル、 "B:" が右チャンネルの音声に対応する。

(4) 言い淀み、言い直し、不要語、あいづち、非言語情報の表現

- 言い直しについては、言い直された区間を、また、言い淀みについては、言い淀んだ部分を () でくくる。
  (例) 一番お薦めなのは (甲州街道) 甲州街道で、
- 冗長語、間投詞 (不要語、無意味語) は [ ] でくくる。
  (例)20 キロぐらい走ったところで、
  [あのー] 鶴川街道という...。

  また、冗長語、間投詞と地の文を区別する。
  (例) 同じ「はい」でも、英語の yes の意味のものは地の文にして、「 [はい] 出てまいりました」のようなものは [ ] に入れる。
- 相槌は { } でくくる。
  (例) もう少し行くと、{ はい。 } 右に看板がでますから...。
- 咳や笑い声などの非言語音声情報は、< > でくくって表す。以下の 3 種類を対象とする。
  < 笑い > : 笑い声
  < 咳 / くしゃみ > : 咳あるいは、くしゃみ
  < 雑音 > : 上記以外の雑音
- これらの記号はすべて全角で書く。

(5) 数詞

- 数詞に現れる数字は全角数字を使う。
- 原則として千の位までは数字をそのまま書き、万以上については「万」を用いる。

  (例) 1,000 円→ 1,000 円
  100,000 円→ 10 万円

  ただし、以下のものは例外的に漢数字を用いる。
  ひとり : 一人、ふたり : 二人
  ひとつ : 一つ、ふたつ : 二つ、みっつ : 三つ、..ここのつ : 九つ
  ついたち : 一日、ふつか : 二日

(6) 読み方の指定

- 読み方の分かりにくい固有名詞や、複数の読みかたが考えられる数詞などには、積極的にローマ字で読みをふる。
- ローマ字は半角アルファベットとし〈 〉で囲む。表記は、電子協方式に従う。大文字であるか小文字であるかは問わない。
- わかち書きする場合は、半角ブランクを用いる。
  (例) 二日目〈futsuka me〉

(7) その他

- アルファベットは使用せず、読みをカタカナで書く。
  (例)JR →ジェーアール
- 短いポーズは「、」、長いポーズは「．．．」で表す。
- 引用は記号は「」を用いて表す。このほかの記号は用いない。
- 「・」「~」「？」などの特殊記号は用いない。
- 同時発話、息つぎ、かすれ声、声の上がり下がり等の情報は重要であるが、一定の基準で安定して付与し難いので、ラベルを付けない。

(b) 1 行 1 文音声対応書き起こし (*.sp)

「1 行 1 文書き起こし」のテキストを対応する音声ファイルごとに区切りを入れ、対応する音声ファイル名を記述したファイルである。

区切り： 10 個以上の半角の”-”を区切りとして入れる。見やすくするための、空行は任意に入れてよい。 header の下にも必ず区切りを入れること。

音声ファイル名：音声ファイル名からチャンネル記号を抜いたファイル名を区切りの 1 つ下の行に入れる。音声ファイルとの対応をとるときに使われる情報であるため、半角で記述し、大文字・小文字に関してもファイル名と統一すること。

(c) 1 行 1 ファイル書き起こし (*.syn)

1 行の書き起こしが正確に音声ファイル 1 つに対応し、かつスペースによるおおよその同期情報を含む書き起こしである。この書き起こしテキストによって、ブラウジングツールによる音声の再生が可能となる。

(1) ヘッダー

「1 行 1 文書き起こし」のヘッダーに、以下の行を加える。
同期情報: あり / なし

- 「同期情報」の後には、書き起こし中のスペースが以下で定義される同期情報の意味を担う場合は「あり」、そうでなければ「なし」とする。

(2) 書き起こしの本体

必ず、左チャンネルの発話者から始まり、各チャンネルの話者ごとに交互に (さらに時間順に)1 行に 1 音声ファイルに対応した発話を書き起こしていく。このため、無音の音声ファイルに対応する発話者の行は発話者ラベルだけの行となる。また、同じ時間に発話された左右両チャンネルの書き起こしは必ず連続する行となり、ペアを形成する。このペアの書き起こしは、発話の重なりを表現する同期情報が (4) で説明する方法で記述される (同期情報がある場合)。

(3) 発話者ラベル

「1 行 1 文書き起こし」と同じ。従えない場合は、それを別のファイルで明記する。

(4) 行の定義

1 行には 1 つの音声ファイルに対応する発話の書き起こしを「1 行 1 文書き起こし」の (4) ~ (7) の基準に準じて書く。また、以下の点は異なる。

- 文の認定と文の終わりの「。」は書いても書かなくてもよい。
- 同期情報を入れる場合は、ペアとなる行の文字の方向を時間軸として、〈半角スペース〉で挟まれた発話の記述開始位置によって、ペアとなる話者の発話との相対的な開始時間をおおよそ表現する。全角を 2 単位、半角を 1 単位の時間が経過したと考える。ただし、絶対的な時間は表現しておらず、あくまでも、各発話 (無音区間で挟まれたもの) の開始位置がペアの発話者の発話のどの位置から始まったかを表現するのみである。

(d) 平仮名書き起こし **(*.kan)**

「1 行 1 文書き起こし」または「1 行 1 ファイル書き起こし」の header 以外の部分を平仮名および片仮名に変換したテキスト。カタカナは外来語などに使用。

(e) ローマ字書き起こし **(*.rom)**

「1 行 1 文書き起こし」または「1 行 1 ファイル書き起こし」の header 以外の部分を JEIDA 正書法によるローマ字表記 (大文字) に変換したテキスト。

(f) その他のテキスト

^_spks: 話者情報

^_dlgs: 対話一覧

^.tex^.ps など: 対話の概要、収録状況、使用した図など

speakers: 収録話者一覧

但し、 ^ は収録機関コード (ccc)

4 ツール

音声対話コーパスの書き起こしテキストは、あくまでも文字であり、音声の細かな特徴 (韻律、調子など) や、発話のタイミングなどを表現することが難しい。この点から、 TEI のようなより表現力のある記述法が提案されているが、このような表現は一般に複雑になりやすく直観的な対話の把握が困難となる。書き起こしテキストでよく分からない特徴などの調査は最終的に音声を聞くことになる。

そこで、本 WG では、京都大学の協力により開発された、音声データの再生ツールと書き起こしデータから容易に対応する音声データを聞くことができるブラウジングツールを CD–ROM に入れた。

音声再生ツールは、 CD–ROM 中にある左右チャンネルに分離している音声ファイルを一般のワークステーションでモノラルあるいはステレオ再生を可能とするものである。

また、ブラウジングツールは、「1 行 1 ファイル書き起こし」テキストから動的に HTML(Hyper Text Markup Language) を生成し、 WWW のブラウザを使うことによって、テキストをマウスでクリックするだけで、その文に対応する音声ファイルをモノラルあるいはステレオで再生できる。また、 Vol.2,3 には au ファイル (8bit, 8kHz, μ-law) に変換された音声ファイルと京大のツールで生成された HTML ファイル (ただし、テキストからのポインタは au ファイルを指している) も同時に収録してある。これは、音が少し悪く、ステレオで両チャンネルを同時に聞くことはできないが、一般のパソコンなどでも WWW のブラウザで、ほとんど準備なしでブラウズできるため、おおよその見当をつけるのに適している。

これらのツールによって、これまでかなりの労力を要した「実際の音を聞いて判断する」という作業が容易になり、研究者はより本質的な考察に時間をかけることが可能となる。

5 音声コーパスの将来

音声コーパスとしては、連続音声、対話音声コーパスの充実が求められており、発話者をかなり多数含む必要がある。また、音声コーパスの設計法を確立することが望まれている。さらに、日本ではあまり作成されていない自由発話音声コーパスを収集することも重要である。従来のものはテキストの読み上げであったり、目的を与えられた模擬対話であったりするので、どうしてもある程度は人工的な影響が残っている。それに対して実際に音声自動認識装置を使う場面では、自由発話音声を想定しなければならない。また、現在は各言語の共通語 (標準語) を対象とするのが普通であるが今後は方言音声を取り扱うことが必要になる。自動翻訳電話の開発等を考えると、多言語コーパスが重要になって来るであろう。

音声データベースで問題になるものにセグメンテーションとラベリングがある。音声信号を時間軸上で言語学的・音声学的単位に区分するセグメンテーションと、その区分された区間に音声学的カテゴリーを割り当てるラベリングは一体の問題である。現在は区分の単位として、音素片、単音、音素、半音節、音節、 VCV 音節などが用いられている。どの単位を採用するか、区間の中心部を示すのか境界を示すのか、境界の定義をどうするか等問題は多い。

音声データベースシステムとしては、まだ確立されたものはないと言って良い。関係モデルをベースにしたものが幾つか試みられている程度である。今後も、しばらくはこの方向で進むものと思われる。将来的にはラベリング等を行わずに音声コーパス自身が集積されていれば、音声認識技術と組み合わせて (その内容は予め分かっているので) 任意の検索ができるようになることが望ましい。

6 むすび

11 の大学の協力により、総計 93 対話、約 450 分の音声対話コーパスを作成した。書き起こしテキストの作成法など、まだ問題点は多いが、この音声対話コーパスを活用して音声対話の研究がさらに発展することを期待している。

付録 A 音声対話コーパス WG のメンバーと活動概要

主査: 板橋秀一 (筑波大学)
市川　熹 (千葉大学)
河原達也 (京都大学)
槫松　明 (電気通信大学)
小林哲則 (早稲田大学)
小林 豊 (京都工芸繊維大学)
白井克彦 (早稲田大学)
土屋 俊 (千葉大学)
鶴丸弘昭 (長崎大学)
新美康永 (京都工芸繊維大学)
松本裕治 (奈良先端科学技術大学院大学)
山下洋一 (大阪大学)
山本幹雄 (豊橋技術科学大学 ('94) 筑波大学 ('95))
横田将生 (福岡工業大学)
編集協力者：高木一幸、保浦直子 ('94)、上田直子 ('95)

上記メンバーで音声対話コーパスの仕様などを作成し、さらにいくつかの研究機関の協力のもとにコーパス構築を進めた。ＷＧは会合を 5 回開いたが、実際の仕様の細かい議論は主に電子メールを介して行った。

付録 B アンケート項目

(1) 音声言語コーパスに関するアンケート

所属＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿ 氏名＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

1. 今後必要なコーパスについて (複数回答可)
(1) どのようななものを希望しますか (例：観光案内、教師と生徒、買物、等)
(2) テキスト、音声、両方、が必要
内容：
媒体、フォーマット：

量：

トランスクリプションは不要、必要 (漢字かな、かなのみ、ローマ字)

(3) 使用目的：

(4) 発話様式 (2 人の自由対話、機械との対話、放送、等)：

(5) タグ情報：

コメント：

2. 現在、手持ちのデータについて (非公開の場合もできるだけ回答願います。)

(1) 音声コーパス：無、有、有るが非公開

コーパス名：

内容：

媒体、フォーマット：

量：

トランスクリプションの有無 (漢字かな、かなのみ、ローマ字)

1) 使用目的：

2) 発話様式 (例： 2 人の自由対話、機械との対話、放送、等)：

3) タグ情報：

4) 利用可能な場合、その方法：

コメント：

(2) テキストコーパス：無、有、有るが非公開

コーパス名：

内容：

媒体、フォーマット：

量：

1) 使用目的：

2) 発話様式 (例： 2 人の自由対話、機械との対話、放送、等)：

3) タグ情報：

4) 利用可能な場合、その方法：

コメント：

3. 重点領域研究で作成すべき音声言語コーパスについての希望

4. その他

回答は、　　1994 年 1 月 14 日 (金)

までに E-mail または Fax でお願い致します。

(2) 対話コーパスの作成条件のまとめ

本 WG では〈 〉の項目を取り上げることとした。

1. 対話環境：〈1〉人間 = 人間〈2〉人間 = 機械 (WoZ 方式を含む)〈3〉電話対話 (4) 放送

2. 対話様式：〈1〉自由対話〈2〉模擬対話〈3〉テキストの読み上げ

3. 対話形態：〈1〉対面〈2〉非対面

4. コーパス形態：〈1〉音声〈2〉テキスト

5. 対話言語：〈1〉共通語 (2) 方言 (3) 外国人との対話 (4) 外国語

6. 非言語情報：〈1〉なし (2) 図表 (地図、時刻表等)

7. 文字転写：〈1〉漢字かな混じり〈2〉かな〈3〉ローマ字

8. タグ情報：(1) 文節 (2) 単語〈3〉間投詞〈4〉冗長語 (あいづち、いい淀み、言い直し)〈5〉割り込み〈6〉発話の重なり〈7〉ポーズ (8) 雑音〈9〉非言語情報 (笑い等)

9. 品詞分類：〈1〉名詞〈2〉動詞〈3〉形容詞〈4〉形容動詞〈5〉副詞〈6〉連体詞〈7〉接続詞〈8〉感動詞〈9〉数詞〈1 0〉助詞〈11〉助動詞〈12〉接頭辞〈13〉接尾辞

10. 音声との対応：〈1〉ポーズ (呼気段落) 〈2〉冗長語〈3〉割り込み〈4〉発話の重なり (5) 非言語情報 (笑い等)

11. 録音環境：〈1〉防音室〈2〉静かな室 (3) 事務室 (4) 計算機室 (騒音環境)

12. 収録媒体：〈1〉 DAT(2)8mmMT(直接 AD 変換)

13. マイクロホン：〈1〉接話型〈2〉卓上 (手持ち) 型

14. 低域フィルタ：遮断周波数、遮断周波数での減衰 (dB)、減衰率 (dB/Oct)、位相特性、等を明記

15. AD 変換：量子化精度 (16 ビット、 12 ビット等) を明記

16. 標本化周波数：〈1〉 16kHz(2)20kHz(3)24kHz(4)22kHz

17. ビデオ情報：〈1〉ビデオ撮影なし (2) ビデオ撮影あり

18. 配布媒体： CD–ROM

19. 対話者の音声はできる限り分離して 2 チャンネルに収録すること

20. 対話の長さ： 3 分程度以上任意

21. 対話の数：任意

22. 対話者数：全体で男女約半数づつが望ましい

23. タスクの例：観光案内、地理案内、空港案内、グルメ案内、交通案内、買いもの、テレフォンショッピング、料理、病気、税金、慶弔、お喋り、日常業務、教師と生徒、質問応答、セミナー、インタビュー、重点領域研究会の発表

24. テキストコーパス： News ネットワーク、 Email 等の利用、卒論、修論、博論、重点領域研究発表資料

付録 C 各 CD–ROM の概要

Vol.1 (1994 年度)

秘書システム (京都大学)：

計算機がユーザの秘書の仕事を行うもので、ユーザのスケジュール管理や他のユーザ (の秘書システム) との連絡、会場の予約等を行う。 (16 対話, 話者 7 名, 181Mbyte)

スケジュール管理 (電気通信大学)：

2 人の話者がそれぞれ自分の予定スケジュールのカレンダーを見ながら、会合の時間や場所を決める対話。 (3 対話, 話者 8 名, 54Mbyte)

クロスワードパズル (早稲田大学)：

縦のカギと横のカギが各対話者に渡されていて、 2 人の話者が協力してクロスワードパズルを解く非対面の対話。 (3 対話, 話者 4 名, 75Mbyte)

地理・旅行案内 (大阪大学)：

目的の建物までの道順を教えたり、観光地の案内をする対話である。 (4 対話, 話者 5 名, 81Mbyte)

テレフォンショッピング (筑波大学)：

注文者がカタログを見ながら品物を注文する際の受付者との対話。接話マイクを使って非対面で収録している。 (4 対話, 話者 3 名, 58Mbyte)

地図課題 (千葉大学)：

1 人が道の記された地図を与えられ、もう 1 人に手元の地図上にどのように道を引くかを教える対話である。 (1 対話, 話者 2 名, 76Mbyte)

Vol.2 (1995 年度)

秘書システム (京都大学)：

計算機がユーザの秘書の仕事を行うもので、グループ会議の設定、複数の予定の登録、会議場選択のタスクを設定した。人間 - 機械 (WOZ) の対話と、ほぼ同じ条件での人間 - 人間の対話の両方を収録した。 (4 対話、話者 4 名、73MB)

スケジュール調整タスク (大阪大学)：

会議などを行うための共通した空き時間を見つけることを目的として対話が行われた。スケジュールなどの状況設定には、 A、 B、 C の三つのパタンが用意され、それぞれ、 3 対話ずつ収録した。 (9 対話、話者 6 名、 137MB)

テレフォンショッピング (筑波大学)：

カタログによるテレフォンショッピングをタスクとした対話。顧客は、通信販売のカタログから予め商品を選び、受付センターに電話注文するという設定。情報入力型・システム主導型の対話である。 (4 対話、話者 4 名、 80MB)

Vol.3 (1995 年度)

アルバイトや旅行に関する会話 (関西大学)：

アルバイトや旅行に関する学生同士の雑談を収録。 48KHz サンプリングのファイルも同時に収録。 (1 対話、話者 2 名、 48MB)

留学生の夏休み (東京工業大学)：

留学生 (インドネシアとシンガポール) 同士による日本での夏休みに関する雑談を収録。 (1 対話、話者 2 名、 24MB)

スケジューリング会話 (電気通信大学)：

ミーティングのスケジュールや会場を決定するための会話を収録。本人同士または本人と相手の秘書との会話である。 (6 対話、話者 6 名、 46MB)

間違い探し (早稲田大学)：

対話者に 6ヶ所の違いがある絵を渡して、相手の絵を見ずにその違いを対話だけによって探し出す間違い探しタスク。協調問題解決型の対話である。 (5 対話、話者 6 名、 143MB)

Vol.4 (1995 年度)

地図課題タスク (千葉大学)：

1 人が道の記された地図を与えられ、もう 1 人に手元の地図上にどのように道を引くかを教える対話である。 (5 対話、話者 4 名、 171MB)

インタビュー会話 (静岡大学)：

ラジオのインタビュー会話を収録。 4 名の話者による模擬対話ではない実際の対話である。 (1 対話、話者 4 名、 21MB)

スキープラン対話の朗読 (東京大学)：

スキープランの予約会話の書き起こし文章を、 2 人で対話のように読み上げたものと、 1 人で朗読したものを収録。 (20 対話 (読み上げ用文章は 2 種類)、話者 10 名、 129MB)

講演 (豊橋技術科学大学)：

学会での OHP を使った講演の音声を収録。 (5 講演、話者 5 名、 94MB)

参考文献

[1] 板橋秀一, “音声デ−タベ−ス”, 信学論, **70**, 4, 433-438(1987)

[2] 武田一哉, 匂坂芳典, 片桐滋, 桑原尚夫, “研究用日本語音声データベースの構築”, 音響誌, **44**, 10, 747-754(1988)

[3] T.Ehara, et al., “ATR Dialogue Database”, Proc. ICSLP-90, 1093-1096(1990)

[4] S. Itahashi, “Creating Speech Corpora for Speech Science and Technology”, IEICE Trans. **E74**, 7, 1906-1910(July 1991)

[5] 板橋秀一 , “小特集にあたって”, 音響誌, **48**, 12, 876-877(1992)

[6] 匂坂, 浦谷, “ATR 音声・言語データベース”, 音響誌, **48** , 12, 878-882(1992)

[7] 小林, 板橋, 速水, 竹沢, “日本音響学会研究用連続音声データベース”, 音響誌, **48**, 12, 888-893(1992)

[8] 板橋秀一, 「音声コーパスの構築と利用」, 重点領域研究「音声対話」チュートリアル予稿集 PASD93-T-1, 1-15(1994)

[9] C.M.S-McQueen and L. Burnard, “Guidelines for Electronic Text Encoding and Interchange(TEI P3)”, Text Encoding Initiative(1994)

[10] 田中和世他, 「音声の知的処理に関する調査研究」, 174-176, (財) 日本情報処理開発協会 (1992 年 6 月)