

「RWCP 音声対話データベース（96 年版）」解説書

1. データ収録環境

1.1 収録場所

収録は音声関連実験室内の防音室で実施した。図 1 に概略の見取図を示す。

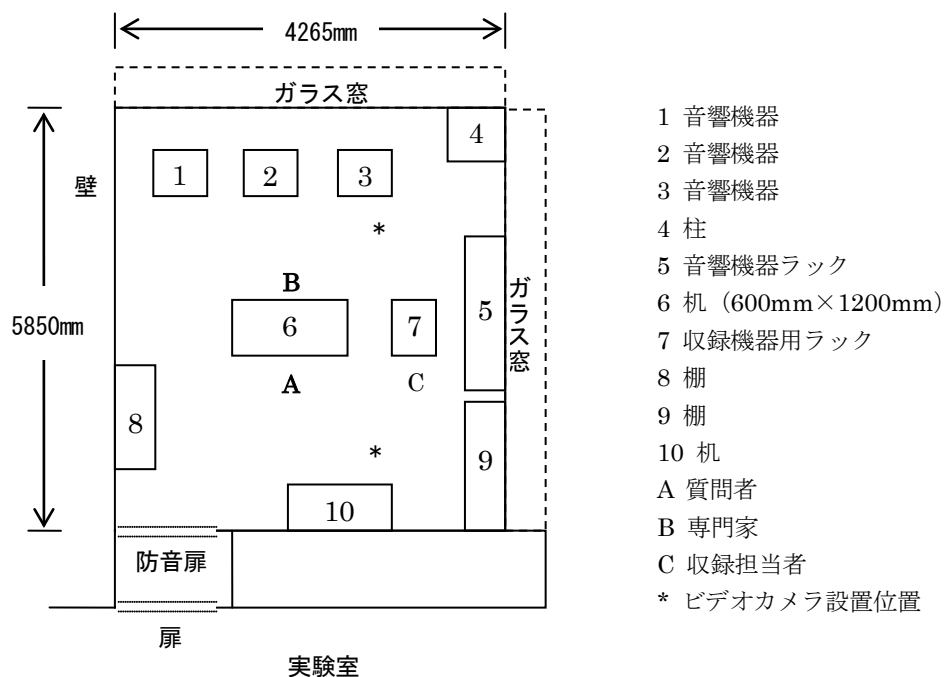


図 1. 収録場所

1.2 収録場所騒音レベル

表 1 に示すとおりである。

表 1. 収録環境騒音レベル（単位[dB]）

	等価騒音レベル	単発騒音暴露レベル	騒音レベルの最大値
A 特性	27.5	37.5	35.0
C 特性	47.7	57.7	51.0
FLAT 特性	52.8	62.8	57.6

2. データ収録機器

表 2 に示すとおりである。

表 2. 使用収録機器

機器	数	メーカー	型名
スタンドマイク	2	SONY	C-38B
スタンドマイク用 DAT デッキ	1	STUDER	D780
接話型マイク	2	SENHEISER	HDM410
接話型マイク用マイクアンプ	1	SHURE	FP32 MIXER
接話型マイク用 DAT デッキ	1	TASCAM	DA-30
8mm ビデオカメラ	2	SONY	CCD-TR650
8mm ビデオデッキ	2	SONY	WV-H1

表 2 の機器を以下のとおり接続した。

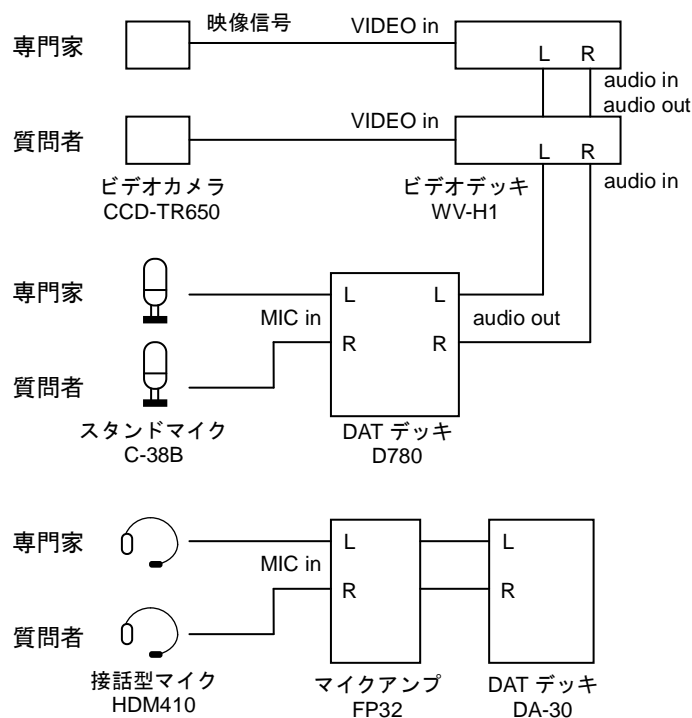


図 2. 収録機器の接続

3. 話者

3.1 専門家

(1) 車の購入

男性話者 1 名 (12 対話) : 現在外車販売会社勤務

女性話者 1 名 (12 対話) : 外車販売会社勤務経験 7 年、退職後約 1 年
(男性話者と同一の会社)

(2) 海外旅行の計画

男性話者 1 名 (12 対話) : 現在旅行代理店勤務

女性話者 1 名 (12 対話) : 現在旅行代理店勤務
(男性話者と同一の会社)

専門家役の話者は、ともに音声対話収録の経験はない。話者はいずれもデータ収集に対しての報酬があることを予め知らされ、自ら望んで収録に参加した者である。

話者の年齢、出身地等プロフィールは別途ファイルに記載されている (5.5 節を参照のこと)。

3.2 質問者

(1) 車の購入

男性話者 12 名、女性話者 12 名 (各話者 1 対話)

(2) 海外旅行の計画

男性話者 12 名、女性話者 12 名 (各話者 1 対話)

質問者役の話者のうち、音声データ収録の経験があるのは車の購入タスクの男性話者 2 名、旅行タスクの男性話者 1 名。他の話者は経験がない。また、質問者役の話者は特定の条件を設けず選択したため、一部の話者を除き、車の購入／海外旅行の計画を現在実際の生活の上で検討してはいない。

なお、車の購入タスクの男性話者 2 名（前記収録経験者）、旅行タスクの男性話者 3 名と女性話者 3 名（収録実施機関の社員）以外は、収録に対しての報酬があることを予め知らされ、自ら望んで応募した者である。

話者の年齢、出身地等プロフィールは別途ファイルに記載されている（5.5 節を参照のこと）。

4. 収録の方法

4.1 話者への指示

専門家に対しては以下の指示を行なった。

- a. 研究用のデータ収録であること。
- b. 基本的には通常のショールーム（車の購入）、カウンター（旅行の計画）に来たお客と同等の話をすること。
- c. 対話収録中に利用できる資料は限られるため、話の内容そのもの（例えば価格など）が現実と異なることがあっても構わない。
- d. 質問者との対話があまり長くなると、データ収集に支障をきたすので、時計を見る必要はないが長くなった場合には、通常の業務で他のお客が待っているような場合を想像し、話を完結させる方向に導いてよい。
- e. スタンドマイクの使用は緊張などしないようなるべく無視してほしい。ただし、マイクから極端に遠ざかる姿勢を取ったり、マイクにカタログをぶつけたりということがないようにすること。

質問者には実験当日はじめてタスクを知らせた上で、以下のような指示を与えた。

- a. 研究用のデータ収録であること。
- b. 車の購入：車の購入を計画し、外車の販売のショールームに来ていることを想像すること。現在の状態は、「はじめて話を聞きにきた場面」「カタログなどを既に見て何らかの話をしにきた場面」「既に車種などをきめ、購入のための詳細な話を行ないに来た場面」など想像しやすい状況を各自考える。車のことすなわち、性能、価格、スタイル、居住性、安全性、など話題は何でも構わない。

旅行計画：海外旅行を計画し、旅行代理店のカウンターに来ていることを想像すること。現在の状況は、「はじめて話を聞きにきた場面」「旅行についての詳細な希望を決めた上で、計画を行なう、あるいはパッケージツアーを決めるような場面」など想像しやすい状況を各自考えること。ただし、実際のカウンターではなく資料は限られているため、無い資料（パッケージツアーのカタログ）を請求したりはしないこと、費用、目的地、時期等の希望は予め詳細に決めておいても構わないし、漠然としたもの（例えば今年の夏休み）でも構わない。
- c. 場面の想定については、実生活と異なる設定（家族、予算）をしても構わないが、矛盾のないように考えること。また、ディーラー／旅行代理店社員と既知であると言う設定は行なってはならない。

- d. 対話の長さに特に規定はない。ショールーム、あるいは代理店に足を運んだ目的が達成できるまでと考えること。ただし、ひとりのひとが余り長くなると、他の質問者の方に長く待ってもらうことになるので、極端に長くならないようにしてほしい。
- e. スタンドマイクの使用は緊張などしないようなるべく無視してほしい。ただし、マイクから極端に遠ざかる姿勢を取ったり、マイクにカタログをぶつけたりということがないようにすること。

4.2 説明から収録まで

質問者役の話者を3名ずつ集めて上記の説明を約10分程度行ない、その後、車のカタログ、旅行パンフレットを提示、場面の想定ができた質問者から、収録を行なう。他の質問者が収録中はカタログ・パンフレットを見て対話の状況を検討する。このとき収録実験室とは離れた部屋で待機し、質問者は収録時にはじめて専門家と相対する。

また、収録が終了した質問者が未実施の質問者と収録内容について話をすることがないようにする。

4.3 収録

話者は着席し、収録スタッフの合図で、収録を開始する。収録は1回のみ行なわれ、また収録中にスタッフは一切の指示を行なわない。

5. データファイル

5.1 ファイル名規則とフォルダ構成

音声データ、書き起こしテキストならびに発声データファイルのファイル名は、

対話番号. ファイル識別子

の形式である。対話番号の形式は以下のとおりである。

[タスクコード]_[専門家性別コード]_[シリアル番号]

[タスクコード] C1 … 車の購入

 T1 … 旅行計画

[専門家性別コード] M … 男性

 F … 女性

[シリアル番号] 01 から始まる数字2桁

なお、各データファイルは、

[タスクコード]_[専門家性別コード]

からなるフォルダ（C1_M, C1_F, T1_M, T1_F）に、該当する12対話分ずつ格納されている。

5.2 音声データ

音声データは、収録したDATをDAT-Link+で16kHzにダウンサンプリングしたものである。16bitのサンプルが、leftチャンネル、rightチャンネル交互に記録され、対話開始から終了までを1ファイルとしている。音声ファイルはWAV形式、ファイル識別子は.wavである。

例 C1_F_04.wav

なお、接話型マイクで同時に収録した音声ファイル（専門家 1 名あたり 3 対話）は、対話番号 H.wav をファイル名とする。

例 C1_F_04H.wav

5.3 書き起こしテキスト

書き起こしテキストは、仮名漢字 (UTF-8) で記述されている。ファイル識別子は .txt である。発話単位で記述され、各発話の先頭には以下がマークされる。

A : … 質問者

B : … 専門家

例

B : で、お子様ですとここから、[えー] 三万円 {ええ。} 引かれますので。

A : [あ] 三万円引かれるんですね。

B : ええ、[えーっと] お子様はお年はおいくつですか。

A : [えーと] ニ才、になりますが。

B : [あ] ニ才ですか。

A : ええ。

発話中の無意味語（えー、あの一など）は [] で囲み、発話中の相手のあいづち、あるいはオーバーラップした発声は { } で囲む。また、言い直しの区間は () で囲む。

例 B : アメリカ {はい} ですと、そうですね。(ひが) 西海岸で....

※「はい」は話者 B の発話中の話者 A のあいづち

仮名漢字での記述は、無意味語ではなるべく発音を正しく表現するように記述するが、有意味部では、発声の曖昧さなどは記述しない。

書き起こしテキストで読みが複数存在し曖昧な部分は、<>で囲みローマ字記述してある。

例 B : そうですね。日本<nihon>の車も確かに....

なお「,」は言語的な意味と、音響的な区切りの両方の意味に用い、短いポーズが有る場合にも「,」で示している。

例 A : ベンツ, ですか。

5.4 発声データファイル

音声データを以下の規則により発声単位に切り出し、そのデータを記述したファイルである。

- A. 発話内の言語的に文と認められる区間は単位とする。
- B. 発話内で言語的に文と認められる発声の一部で、500msec 以上のポーズが存在する場合、ポーズ前後で別単位とする。
- C. 発話内の無意味語、冗長語は 1 単位とする。
- D. 発話内の言い直しの区間は 1 単位とする。

無意味語、冗長語とは、「えー」、「あの一」などの発声である。相手の発話に対しての肯定を示す「そうですね」以外の、単に発話の間をうめるための「そうですねー」も無意味語としている。

言い直しとは、

「その場合は、よや、ご予約いただきまして」

における「よや」の部分等である。ただし、これはその部分の発声が言語的に意味をなさない場合に言い直しと判定し、

「やはり、ハワイ、あっ、バリも人気有りますよ」

における「ハワイ」の区間は、言い直しとはしていない。

ファイルには、個々の発声単位に対して以下のデータが記述されている。以降、各データの詳細を述べる。

シーケンシャル番号

フラグ

話者

開始時刻

終了時刻

仮名漢字記述

ローマ字記述（表音記述）

（ターミネータ）

例

```
0059
2A
A
71890
74160
そうですね、向こうで、レンタカーを借りて
so-desune,muko-de,ren'taka-okarite
#
0060
4B
A
76060
77020
まわりたいなど。
mawaritainato.
#
0061
5C
B
74880
75520
そうですね,
so-desune,
```

シーケンシャル番号

各発声単位に 1 から順に付けられた番号。ファイル中では 0 を付して、4 桁で記述してある。

例 0001

なお、この順番は、発話開始時刻でソーティングされたものではなく、1 発話のデータがなるべく連続するように順序付けられたものである。

例

専門家	[1] ----	[2] ----	[3] -----		[8] ---	[9] -----	[10] -----
質問者		[4] ---	[5] ---	[6] -----	[7] -----		[11] ---

→時間

したがって、あいづちは、相手発話のデータの後に記述されている。

フラグ

発声単位の、言語的、音響的特徴を示す 2 文字のデータ。

【1 文字目】1 から 7 の数字で、言語的特徴を表す。

- 1 その発声が言語的にまとまった単位

例 「はい」

「ハワイのツアーの場合はもう、航空会社とホテルの選択なんですね」

(必ずしも文法的な文ではないが、以下の説明では簡単のため文と呼ぶ)

- 2 文の一部で、文の先頭に位置する単位
3 文の一部で、文の中間に位置する単位
4 文の一部で、文の末尾に位置する単位
5 冗長語、無意味語
6 言い直し部分
7 非言語 (笑いごえ、おおきなブレス)

【2 文字目】A から E までのアルファベットで、音響的特徴を表す。

- A ノイズ、相手の声等が重畳していない比較的良好に収録されている音声
B 相手の声が重畳している音声
C ノイズが重畳している音声 (おもに、パンフレットをめくる音)
D 相手の声と、ノイズがともに重畳している音声
E 極めて不明瞭な音声 (極めて声が小さいなど、相手の声が重畳しているものも含まれている)

なお、A から E の判定は基本的に、切り出した音声単位の聴取によって行なった。したがって、

例

専門家	[1] -----		[4] -----
質問者		[2] ---	[3] -----

→時間

のような場合で、[2]の発声が偶然、[1]の発声の短いポーズに有ると

[1] … B (相手の発声の重畳あり)

[2] … A (相手の発声の重畳なし)

となる場合が有り得る。また、レベルに著しい相違があり、3の発声が極めて弱い場合、

[3] … B (相手の発声の重畳あり)

[4] … A (相手の発声の重畳なし)

となる場合も有り得る。

話者

発声者を示すアルファベット 1 文字。

A … 質問者

B … 専門家

開始・終了時刻

ファイルの先頭から msec 単位で表した発声単位の始端、終端。

始端、終端の検出は、切り出し音声の聴取、ならびに音声データのパワー外形、波形の観察によって行なった。ただし、2 話者の発声の重畳など始終端の決定が困難であることも考慮し、10ms 単位で記述している。

仮名漢字記述

前述の書き起こしテキストを参照のこと。

ただしここでは、書き起こしテキストで記述した、{ }、() 等の記号、< >による読み記述はなく、単にその発声単位について、話者の発声を仮名漢字で記述している。

ローマ字記述

次ページの音節記号（電子協表記法による）を用いて、発声を記述したもの。

ここでは、なるべく音声に忠実に表記することを目的としている。

例 車の購入を考えているんですが

kurumanoko-nyu-okan'gaeteirun'desuga

文節内の長母音は「-」で記述するが、文節をまたがる場合は「-」は用いない。

例 アメリカあたりでは

amerikaataridewa

ア	a	イ	i	ウ	u	エ	e	オ	o
カ	ka	キ	ki	ク	ku	ケ	ke	コ	ko
サ	sa	シ	shi	ス	su	セ	se	ソ	so
タ	ta	チ	chi	ツ	tsu	テ	te	ト	to
ナ	na	ニ	ni	ヌ	nu	ネ	ne	ノ	no
ハ	ha	ヒ	hi	フ	fu	ヘ	he	ホ	ho
マ	ma	ミ	mi	ム	mu	メ	me	モ	mo
ヤ	ya			ユ	yu			ヨ	yo
ラ	ra	リ	ri	ル	ru	レ	re	ロ	ro
ワ	wa								
ガ	ga	ギ	gi	グ	gu	ゲ	ge	ゴ	go
ザ	za	ジ	ji	ズ	zu	ゼ	ze	ゾ	zo
ダ	da					デ	de	ド	do
バ	ba	ビ	bi	ブ	bu	ベ	be	ボ	bo
パ	pa	ピ	pi	プ	pu	ペ	pe	ポ	po
キャ	kya			キュ	kyu	キエ	kye	キョ	kyo
シャ	sha			シュ	shu	シェ	she	ショ	sho
チャ	cha			チュ	chu	チェ	che	チョ	cho
ニャ	nya			ニュ	nyu	ニエ	nye	ニョ	nyo
ヒャ	hya			ヒュ	hyu	ヒエ	hye	ヒョ	hyo
ミャ	mya			ミュ	myu	ミエ	mye	ミョ	myo
リャ	rya			リュ	ryu	リエ	rye	リョ	ryo
ギャ	gya			ギュ	gyu	ギエ	gye	ギョ	gyo
ビャ	bya			ビュ	byu	ビエ	bye	ビョ	byo
ピャ	pya			ピュ	pyu	ピエ	pye	ピョ	pyo
		スイ	si						
		ティ	ti	トゥ	tu	イエ	ye		
						ウエ	we	ウオ	wo
ファ	fa	ウィ	wi	ドウ	du	フェ	fe	フォ	fo
ツァ	tsa	ディ	di			ツエ	tse	ツォ	tso
		フィ	fi						
ジャ	ja	ツイ	tsi	ジュ	ju	ジェ	je	ジョ	jo
ヴァ	va	ズイ	zi	ヴ	vu	ヴェ	ve	ヴォ	vo
		ヴィ	vi						
ン	n'								
ツ	q								
ー	—								

5.5 話者プロフィール

話者のプロフィールは、以下のファイル名でタスクと専門家の性別ごとにカンマ区切りファイルにまとめられている。

[タスクコード]_[専門家性別コード]_speaker.csv

各行に記述されている内容は以下のとおりである。

話者コード
話者イニシャル
年齢
性別（男性 M／女性 F で表記）
職業
現住所
出身地
父親の出身地
母親の出身地

話者コードは次のように記述されている。専門家は各タスクで共通であるため、対話番号のシリアル番号を ?? としている。

対話番号 A … 質問者プロフィール （例 C1_F_01A）

対話番号 B … 専門家プロフィール （例 C1_F_??B）

職業は本人が記述した通り。

出身地については、小学生時代を過ごした主な土地を記述させている。現住所、出身地、両親の出身地は、半角スペースを挟み以下の形式で記述されている。

xxxx [都道府県] xxxx [市区郡町村]

両親の出身地が不明な場合（県までしか分からない場合）は、「?」を付している。