

感情評定値付きオンラインゲーム音声チャットコーパス (OGVC) 使用説明書

有本泰子 河津宏美

Oct., 2013

ver. 1.1

1 概要

本コーパスは独立行政法人情報処理推進機構（IPA）の 2007 年度第 II 期末踏ソフトウェア創造事業（末踏ユース）「オンラインゲームにおける匿名性を有した音声チャットの開発」（代表者：河津宏美）のために収集された音声と感情の評定値を取めている。

本コーパスは 2 種類の感情音声で構成される。ひとつは対話中に自然に表出した感情を含んだ自発対話音声であり、もうひとつは従来の感情音声研究で用いられてきた演技による感情音声である。

自発対話音声では、オンラインゲームを対話収録に導入することで自発的で生き活きとした感情を誘発した。さらに、ゲーム中のプレイヤー同士に音声チャットによるコミュニケーションを行なわせることで、誘発した感情が音声に反映されるように工夫した。本手法により、自然な対話を 6 対話（約 1 万発話）収録し、さらに個々の発話に対して感情評定値を付与した。

演技による感情音声は、感情表現が得意と考えられるプロの俳優に演技させたものである。音声収録時に使用した発話テキストは、自然な対話で収録した発話の転記テキストから選択した。発話単位で表現させる感情とその強度を指定して、プロの俳優 4 名に感情を込めて発声させ、計 2,656 発話を収録した。

本説明書では、自発対話音声の音声データ・転記テキスト・感情種別ラベル・感情強度ラベルおよび演技音声の発話テキスト・音声データの仕様に関して説明を行う。収録環境の詳細や音響分析、ラベリングの統計分析の結果などは 3 章に示す文献を参照のこと。

なお、本コーパスの対話内容には友人同士の日常会話を含むため、話者の個人情報・人権保護の観点から、本コーパスの使用にあたり以下の点を遵守願いたい。

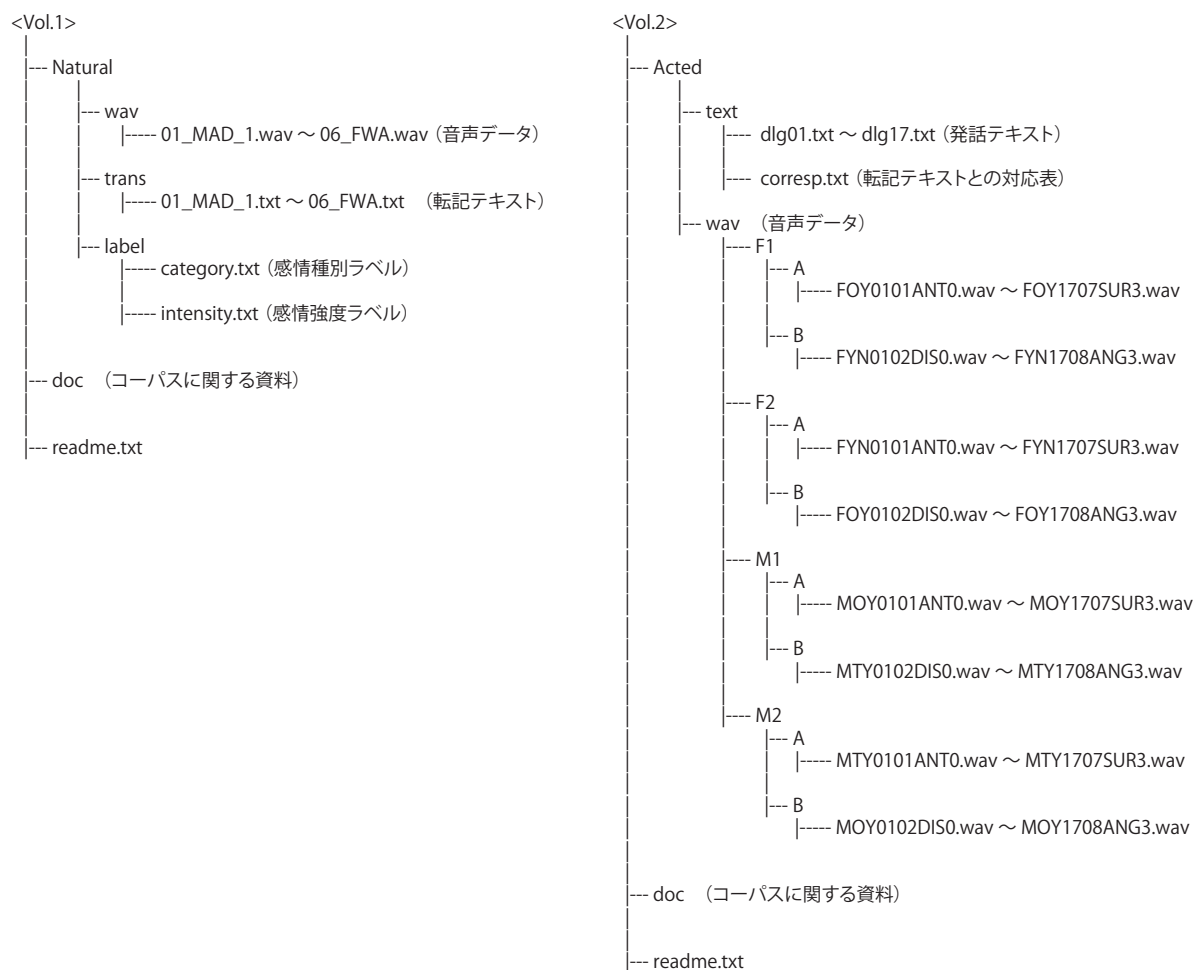


図 1 コーパスのフォルダ構成。（Vol.1 と Vol.2 の doc フォルダ内のファイルおよび readme.txt は同一のファイルである）

- 本コーパスを解析することによって知りえた話者の個人情報積極的に保護すること。
- 研究成果の公表時には本コーパスに記録された以外の話者情報およびそれを取得することのできる情報を公開してはならない。
- 本コーパスに含まれる発話については、事実関係の正誤適否等、発話の内容に関する議論、批判、感想等を公開してはならない。

2 コーパスの構成

このコーパスは自発対話音声と演技音声の 2 種類の感情音声をそれぞれ 1 枚ずつの DVD に格納している（Vol.1：自発対話音声，Vol.2：演技音声）。図 1 に Vol.1 および Vol.2 のフォルダ構成を示す。

表 1 話者の出身地とオンラインゲーム歴およびプレイ時間（／月）.

話者 ID	出身地	ゲーム歴	プレイ時間
01_MAD	長野県	1 年 6 ヶ月	30 時間
01_MMK	長野県	3 年強	75 時間
02_MEM	東京都	4 年 6 ヶ月	1～2 時間
02_MFM	東京都	3 年	30 時間
02_MTN	東京都	5 年 1 ヶ月	24 時間
03_FMA	静岡県	3 年	60 時間
03_FTY	東京都	1 年	2 時間
04_MNN	不明	3 年	8 時間
04_MSY	不明	2 年	2 時間
05_MKK	不明	4 年 2 ヶ月	100 時間
05_MYH	不明	3 年 4 ヶ月	0 時間
06_FTY	神奈川県	2 年 9 ヶ月	80 時間
06_FWA	東京都	2 年 9 ヶ月	60 時間



図 2 話者 ID の構成.

2.1 自発対話音声 (Vol.1)

Vol.1 には, MMORPG (Massively Multiplayer Online Role-Playing Game) と呼ばれるオンラインゲームを行っている間に、2 名もしくは 3 名の話者間で交わした対話音声 6 対話分, 計 9,114 発話を収めている.

音声データは話者ごとにひとつのファイルに保存されており, 各発話の時間情報は別に保存した転記テキストに発話内容とともに記述している. 音声データと転記テキストはファイル名に話者 ID を付して拡張子以外を同一とした. 各発話の感情評定値 (感情種別ラベルと感情強度ラベル) は, 全話者分をまとめてひとつのテキストファイルに保存した.

2.1.1 話者

話者はオンラインゲームの経験がある大学生 13 名 (男性 9 名, 女性 4 名) で, 親近性のある同性同士で参加させた. 各話者の出身地, オンラインゲーム歴, 収録当時の一ヶ月当たりのプレイ時間を表 1 に示す. オンラインゲーム歴は平均 38 ヶ月, 平均プレイ時間は 33 時間／月であった.

表 2 音圧レベルを修正したファイルとその修正値.

音声ファイル	修正値
02_MEM.wav	60dB
02_MFM.wav	60dB
02_MTN.wav	60dB
03_FMA.wav	55dB
03_FTY.wav	55dB
04_MNN.wav	55dB
04_MSX.wav	55dB

話者 ID は、全部で 6 文字で構成されている。話者 ID の構成図を図 2 に示す。6 文字のうち最初の 2 文字は対話 ID であり、そのあとアンダースコア (_) に続いて話者の性別を 1 文字で示している。最後の 2 文字は話者を区別するための任意の文字列である。話者 ID の最初の 2 文字が対話 ID を表しているため、この 2 文字が一致する話者は同じ対話に参加した対話参与者であることが分かる。

2.1.2 音声データ

音声データは話者ごとに WAVE 形式 (48kHz・16bit・monaural) で Natural/wav フォルダに保存している。なお、対話 ID が 01 の話者 (01_MAD と 01_MMK) は対話収録を途中で中断したため、音声ファイルが二つ存在しており、話者 ID に続けてアンダースコア (_) と対話の順序を示す数値 (1 あるいは 2) を付与してある。

また、収録音声の振幅レベルが小さく、発話の聞き取りが困難であった話者の音声ファイルについては、対話全体の音圧レベルの平均が 60dB もしくは 55dB となるように修正を加えている。収録後に音圧レベルの修正を加えた音声ファイルとその修正値を表 2 に示す。これらの中には、修正を加えた影響により雑音レベルが上がっているものもあるのでご注意願いたい。

注意：

今回公開した音声は、音声チャットシステムを通じてインターネット経由で行った対話を、それぞれの収録サイトにて、その収録サイトにいる話者の音声のみを録音したものである。そのため、各話者の音声ファイルを組み合わせれば対話音声として聞くことはできるが、各話者が収録時に聴取した相手話者の音声にはネットワーク遅延を含むため、実際の対話状況を再現することはできない。つまり、話者内の発話の分析には利用可能であるが、話者交代など話者間のインタラクションの分析に用いることはできないことに注意が必要である。

```

001,0.0925,1.4850,B:,組めるよね?、ぎりぎり↓
002,1.9775,2.2100,B:,うん↓
003,3.4625,3.9325,B:,よく↓
004,4.8075,7.7075,B:,よくキュージョーロクまで上げたよね。{笑}↓
005,9.3825,10.8675,B:,全部ソロでしょうしかも?↓
006,14.2375,14.6800,B:,マジ?↓
007,15.7525,17.1950,B:,びっくりするよね。{笑}↓
008,18.2500,19.5575,B:,あ、ん別に↓
009,22.7150,23.2525,B:, {笑} ↓
010,24.0175,24.2100,B:, {笑} ↓
011,26.7825,27.1800,B:, {笑} ↓
012,29.3475,29.7150,B:,うん↓
013,32.2700,32.6850,B:,ああ↓
014,34.9900,35.2525,B:,うん↓
015,40.2075,40.5775,B:,うん↓
016,46.3875,46.7300,B:,うん↓
017,50.4525,51.4675,B:,ロクジョーナナねー。↓
018,51.8675,53.5450,B:,それもアコとは組んでるよね。↓
019,57.6925,58.0700,B:, {笑} ↓
020,65.9075,66.1325,B:,うん↓
021,67.9075,68.1850,B:,うん↓
022,70.4925,72.4425,B:,うんまあそれもありっちゃあります。{笑}↓
023,72.9150,73.0825,B:, {笑} ↓
024,76.3500,78.7475,B:,あー、コンロンでもいいかも。↓

```

図 3 転記ファイルの例。

2.1.3 転記テキスト

音声データに対応する転記テキストを Natural/trans フォルダに収めた。図 3 に例を示す。

転記に関する情報は左から順にカンマ区切りで保存した。発話単位基準を 400ms 以上のポーズによってはさまれた音声の範囲として、発話のセグメンテーションを行っている。転記ファイルに収めた情報は下記のとおりである。ヘッダは設けていない。

- 発話番号
対話内の発話の通し番号。感情種別ラベル・感情強度ラベルファイルの uid (2.1.4 を参照) と対応する。
- 開始時刻
発話の開始時刻。単位は秒。
- 終了時刻
発話の終了時刻。単位は秒。
- 話者番号
対話内の発話者 ID。「A:」、「B:」または「C:」の値を取る。
- 発話内容
発話内容の転記。転記基準については以下の本文を参照のこと。

発話の内容は漢字かな交じりの正書法で記述している。言語的な文末を表す箇所には句点 (。) を付した。文末の基準は日本語話し言葉コーパスの絶対境界 [1] までを対象とした。400ms よりも短いポーズには読点 (、) を用いてその箇所を示した。疑問の終助詞や語尾の上昇などにより疑問形とわかる場合には、転記者の基準で疑問符 (?) を付与した。オンラインゲームに特有の用語

表 3 感情種別ラベルの概要.

分類	記号	説明
喜び	JOY	良いことに会って非常に満足し、うれしい、ありがたいと思う感情
受容	ACC	心がひきつけられ、積極的に受け入れよう、接し続けようとする感情
恐れ	FEA	危害が及ぶことを心配してびくびくし、その人やその物と接することを避けたがる感情
驚き	SUR	意外なことを見聞きして心が強く動揺し、平静を失う、どう判断すべきか戸惑う感情
悲しみ	SAD	不幸なことに会った時など、取り返しのつかない事を思い続けて泣きたくなる感情
嫌悪	DIS	その状態・行為をすんなりと受け入れることができず、避けようとする感情
怒り	ANG	許しがたい事柄に接し、不快感を抑えきれず、いらだった状態の感情
期待	ANT	望ましい事態の実現、好機の到来を心から待つ感情
平静	NEU	まったく感情が表れていない
その他	OTH	ノイズが大きい場合など 8 種の感情に分類不能のもの

(BOT や STRAGE) および数字や数詞はその読みをカタカナで表記した.

笑い声や咳など発話以外の言語音などに対するタグを以下に示す.

- { 笑 }, { 咳 }
笑い声 (笑いながらの発話は除外) と咳.
- (?), (?(コメント))
よく聞き取れない音声, 自信がない場合. (コメント) の箇所にコメントを記述.
- [comment=(コメント)]
作業者のコメントがある場合. コロン以下の (コメント) の箇所にコメントを記述.

2.1.4 感情種別ラベル・感情強度ラベル

収録した 9,114 発話のうち, 収録音声の振幅レベルが小さく評価に使用できないと判断した 2 名 (02_MFM, 03_FMA) の話者の 1,009 発話, および音響的分析に影響を及ぼす転記用タグが付与されている 1,527 発話を除外した 6,578 発話に対して, 感情の種類のリベル付けを行なった. 感情種別ラベルには Pluchik の立体構造モデルの基本 8 感情 [2] と「平静」および「その他」の 10 個のリベルを用いた. 表 3 にラベルの概要を示す.

感情種別ラベルファイルの例を図 4 (左) に示す. 各発話に対して異なる 3 名の評価者が評価を行っている. Natural/eval フォルダ内の category.txt に評価の結果を収めた. category.txt ファイルはカンマ区切りのテキストデータで, 1 行目はヘッダ行となっており, 2 行目以降に各発話の評価結果の情報が記述されている. ファイルに含まれる情報の詳細は以下のとおりである.

- fid
ファイル名. 音声ファイル名および転記ファイル名に対応する.
- uid
発話番号. 転記ファイル内の最初の列に示される発話番号 (3 ケタの数値) と対応する.

fid,uid,ans1,ans2,ans3↓	fid,uid,emotion,EF01,EF02,EF03,EF04,EF06,EM01,EM02,EM03
01_MAD_1,001,NEU,JOY,NEU↓	01_MAD_1,002,DIS,1,4,2,1,4,2,3,1,3,2,2,2,1,2,3,1,4,2↓
01_MAD_1,002,DIS,JOY,DIS↓	01_MAD_1,003,JOY,2,4,5,4,1,5,4,4,5,3,3,4,4,3,4,3,3,4↓
01_MAD_1,003,ANG,JOY,JOY↓	01_MAD_1,005,SUR,5,4,2,5,5,3,4,4,5,4,4,4,3,4,5,3,4,4↓
01_MAD_1,005,SUR,SUR,SUR↓	01_MAD_1,014,ACC,3,3,2,2,1,2,3,2,3,2,3,2,2,2,2,2,2↓
01_MAD_1,006,NEU,OTH,SAD↓	01_MAD_1,015,ACC,3,4,1,2,1,2,2,3,2,3,2,3,3,3,2,2,2↓
01_MAD_1,008,NEU,ACC,DIS↓	01_MAD_1,025,DIS,1,4,3,1,4,1,2,2,4,3,2,3,3,2,2,4,4,3↓
01_MAD_1,012,SAD,ACC,NEU↓	01_MAD_1,027,ANT,2,3,2,4,2,3,4,3,3,1,3,3,3,3,4,2,3,3↓
01_MAD_1,013,NEU,DIS,ACC↓	01_MAD_1,028,SAD,1,5,4,3,4,3,1,3,4,3,3,3,5,4,3,4,5,3↓
01_MAD_1,014,SAD,ACC,ACC↓	01_MAD_1,029,ANG,1,3,2,4,3,4,4,4,3,4,3,4,5,4,4,4,4,2↓
01_MAD_1,015,DIS,ACC,ACC↓	01_MAD_1,032,SUR,2,4,1,2,3,4,3,3,2,1,3,3,2,3,3,3,3,2↓
01_MAD_1,016,SAD,NEU,NEU↓	01_MAD_1,038,ANG,1,4,2,2,3,1,2,2,2,4,3,3,4,4,2,2,4,2↓
01_MAD_1,017,FEA,NEU,SAD↓	01_MAD_1,040,DIS,1,3,3,3,3,2,3,3,2,3,2,3,2,3,2,3,2,4,2↓
01_MAD_1,018,NEU,NEU,NEU↓	01_MAD_1,043,ACC,4,5,2,3,2,2,4,3,2,2,3,3,3,3,4,3,2,3↓

図4 感情種別ラベルファイル（左）・感情強度ラベルファイル（右）の例。

- ans1～ans3

3名の評価者による評定値。評定値は表3の記号欄に記載した記号で表記している。

続いて、感情種別ラベルを付与した結果、2名以上の評価者の評定値が一致したもののうち、「平静」と「その他」を除く2,847発話に対して、感情強度ラベルを付与した。評価者は18名の男女（男性13名、女性5名）である。感情強度ラベルファイルの例を図4（右）に示す。感情の強度の評価は1（弱）－3（中）－5（強）の5段階からひとつを選択させている。Natural/evalフォルダ内のintensity.txtに評定の結果を収めた。ファイルの書式はcategory.txtと同様である。ファイルに含まれる情報の詳細は以下のとおりである。

- fid

ファイル名。音声ファイル名および転記ファイル名に対応する。

- uid

発話番号。転記ファイル内の最初の列に示される発話番号（3ケタの数値）と対応する。

- emotion

評価対象となる感情。評定値は表3の記号欄に記載した記号で表記している。

- EF01～EM15

感情の強度評定値（1（弱）～5（強）の値が入る）。ヘッダの記号は評価者を表している（評価者記号（E）＋性別記号（[F/M]）＋任意の2ケタの数値（xx））。

2.2 演技音声（Vol.2）

Vol.2には、4名のプロの俳優（男性2名、女性2名）に8つの感情状態を演じさせた計2,656発話を収めている。収録は同性同士の俳優にペアを組ませて、感情を含まない平静状態（強度0）と弱（1）－中（2）－強（3）の三段階の感情強度で17対話を表現させて、各発話音声を収録した。この時、対話者間で発話が重複しないように指示を出している。発話量を話者間で同一にするため、役を交代させて、さらに対話の収録を行った。

```

01,A:,ANT,属性は？↓
02,B:,DIS,属性はね↓
03,B:,ANG,水↓
04,A:,FEA,水、水武器を持ってけばいいの？↓
05,B:,ACC,そうだね↓
06,A:,FEA,持ってたっけな↓
07,B:,ANT,水武器じゃないや。風武器↓
08,A:,SUR,風武器？↓
09,B:,DIS,ああ↓
10,A:,SAD,風ダメあったかな？↓
11,A:,ANG,ちょっと待って。今狩ってる場合じゃないんだ↓
12,A:,ANG,準備時間なんだよ↓

```

図5 発話テキストの例.

```

dlgid,uid,sid,fd,uid,stime,etime,sid,fd,uid,stime,etime,sid↓
01,01,A:,01_MMK_1,096,324.1325,324.7025,A:↓
01,02,B:,01_MAD_1,098,324.4250,325.3925,B:↓
01,03,B:,01_MAD_1,099,326.2250,326.5325,B:↓
01,04,A:,01_MMK_1,097,327.7475,330.5850,A:↓
01,05,B:,01_MAD_1,101,332.7425,333.2700,B:↓
01,06,A:,01_MMK_1,098,333.3325,334.2300,A:↓
01,07,B:,01_MAD_1,102,334.8825,336.3775,B:↓
01,08,A:,01_MMK_1,099,336.6000,337.1825,A:↓
01,09,B:,01_MAD_1,103,338.9350,339.1850,B:↓
01,10,A:,01_MMK_1,100,339.6850,340.8100,A:↓
01,11,A:,01_MMK_1,101,342.0225,343.6800,A:↓
01,12,A:,01_MMK_1,102,344.2850,345.3050,A:↓

```

図6 発話テキストと自発対話音声の対応表.

2.2.1 発話テキスト

Vol.1に収録されている自発対話音声から、17の対話シーケンス（計166発話）を抜粋し、読み上げ対象となる発話テキストを作成した。発話テキストは10前後の発話で構成されており、発話ごとに演じる感情が指定されている。

発話テキストの例を図5に示す。発話テキストは対話シーケンスごとにActed/textフォルダに保存した。カンマ区切りのテキストデータで、ヘッダは設けていない。ファイルに収めた情報は下記のとおりである。

- 発話番号
対話内の発話の通し番号。発話者に依存せず番号をふっている。
- 話者番号
対話内の発話に対する発話者情報。「A:」または「B:」の値を取る。
- 感情
表現対象となる感情。自発対話音声で評価された感情を概ね踏襲しているが、一部異なる場合がある。
- 発話内容
読み上げる発話内容。読みやすさのため、言いよどみなどを削除しているので、自発対話音

声の転記内容と一部異なる場合がある。また、自発対話音声の 2 発話を 1 つの発話とした場合と、逆に自発対話音声の 1 発話を 2 発話に分割した場合がある。

また、これらの発話テキストに対応する自発対話音声の発話番号を Acted/text フォルダ内の corresp.txt に収めた。例を図 6 に示す。カンマ区切りのテキストデータで全発話テキスト分を記載しており、1 行目はヘッダ行となっている。ファイルに含まれる情報の詳細は、左から順に以下のとおりである。

- 発話テキスト情報
 - dlgid
テキスト番号。発話テキストのファイル名の数値に対応する。
 - uid
発話番号。発話テキスト内の最初の列に示される発話番号（2 ケタの数値）に対応する。
 - sid
発話者情報。「A:」または「B:」の値を取る。
- 対応する自発対話音声情報（1 行に 2 セット存在する場合がある）
 - fid
ファイル名。音声ファイル名および転記ファイル名に対応する。
 - uid
発話番号。転記ファイル内の最初の列に示される発話番号（3 ケタの数値）に対応する。
 - stime
当該発話の開始時刻。単位は秒。
 - etime
当該発話の終了時刻。単位は秒。
 - sid
当該発話の発話者情報。「A:」または「B:」の値を取る。

2.2.2 音声データ

音声データは発話ごとに WAVE 形式（44.1kHz・16bit・monaural）で Acted/wav フォルダに保存している。発話の前後には 400ms の無音区間を挿入してファイルに保存している。役を変えて、俳優一人当たり 2 回の対話収録を行っているため、1 回目の収録と 2 回目の収録でフォルダを分け（F1/F2/M1/M2）、さらにその下の階層に配役でフォルダを分けて（A/B）、音声ファイルを保存した（図 1）。つまり、1 回目の対話収録で A の役を演じた女性の俳優の音声は、Acted/wav/F1/A フォルダと Acted/wav/F2/B フォルダに保存されていることになる。

F	O	Y	0	1	0	1	A	N	T	0	.	w	a	v
①	②	③	④	⑤	⑥									

① 性別 (F: 女性, M: 男性), ② 任意の 2 文字, ③ 対話番号, ④ 発話番号
 ⑤ 感情の種類 (表 2 参照), ⑥ 感情の強度 (0 ~ 3)

図 7 演技音声のファイル名の構成.

音声ファイルと発話テキストおよび演じた感情の種類と強度との対応付けを分かりやすくするため, 音声ファイル名に対話番号と発話番号, および感情の種類と強度も記載している. 図 7 に演技音声のファイル名の構成を示す.

3 関連する文献

コーパスの概要および統計的分析結果等の詳細については, 下記の文献を参照のこと. 各文献は PDF ファイルで doc フォルダ内に格納している.

- ASJ2013.pdf
有本泰子, 河津宏美, “音声チャットを利用したオンラインゲーム感情音声コーパス,”
日本音響学会 2013 年秋季研究発表会講演論文集, 1-P-46a, pp. 385-388, 2013.
- AST2012.pdf
Y. Arimoto, H. Kawatsu, S. Ohno, and H. Iida, “Naturalistic emotional speech collection paradigm with online game and its psychological and acoustical assessment,”
Acoustical Science and Technology, vol. 33, no. 6, pp. 359-369, 2012.
- AST2011.pdf
Y. Arimoto, S. Ohno, and H. Iida, “Assessment of spontaneous emotional speech database toward emotion recognition: Intensity and similarity of perceived emotion from spontaneously expressed emotional speech,”
Acoustical Science and Technology, vol. 32, no. 1, pp. 26-29, 2011.
- INTERSPEECH2008.pdf
Y. Arimoto, H. Kawatsu, S. Ohno, and H. Iida, “Emotion recognition in spontaneous emotional speech for anonymity-protected voice chat systems,”
in Proceedings of Interspeech 2008, pp. 322-325, 2008.
- IEICE2008.pdf
河津宏美, 有本泰子, 大野澄雄, “感情情報を伝達する匿名音声チャットの開発,”
2008 年電子情報通信学会総合大会情報・システム講演論文集 1, D-14-3, pp. 176, 2008.

- SLP2008.pdf

有本泰子, 河津宏美, 大野澄雄, 飯田仁, “感情音声のコーパス構築と音響的特徴の分析-MMORPG における音声チャットを利用した対話中に表れた感情の識別 -,”
情報処理学会第 70 回音声言語情報処理研究会 (SIG-SLP) 研究報告, pp. 133-138, 2008.

4 問合せ先

有本泰子

email: ar@mf.teu.ac.jp

website: <https://sites.google.com/site/ogcorpus/>

更新情報

2013/10/28 演技音声のサンプリングレートを修正 (48kHz ⇒ 44.1kHz)

参考文献

- [1] 丸山岳彦, 高梨克也, 内元清貴. 節単位情報. 報告書『日本語話し言葉コーパスの構築法』, 第 5 章, pp. 255–322. 国立国語研究所, 2006.
- [2] Robert Plutchik. *Emotions: A psychoevolutionary synthesis*. Harper & Row, 1980.