ATR地域別中国語音声データベースⅡ

概要

本データベースは、中国語話者の居住地域多様性をカバーするため、おもに中国沿岸部を南北に網羅した地域における中国語母語話者による標準中国語(普通話)を収録した音声データベースです。

(注)各地域の方言発話ではありませんので、ご注意ください。

特徴

- 中国の北方地域(東北/北京/山東/河南)と南方地域(広東/江蘇/浙江/福建/台湾/その他南方)による母語話者による音声を収録[1]
- 読み上げテキストと、話者毎の音声ファイルにより音声認識開発(モデル学習/評価)に利用可能
- 男女比、年代分布をバランスさせた1,500名(北方:500名、南方:1,000名)の話者による収録時間が総計513時間におよぶ大規模な音声を収録

[1] 東北とは、吉林、遼寧、黒竜江を表し、広東とは、広東と香港を表します。　またその他の南方とは、湖南、青海、貴州、広西を表します。話者地域分布は、16歳以前で最長期間居住地によって分類しています。

話者数/発話文リスト/発話量

- 話者数：1,500名(男女比 1：1)

地域		話者数　(男性/女性)
北方	北京	200　(100/100)
	山東	149　(75/74)
	河南	50　(25/25)
	東北	101　(50/51)
	計	500　(250/250)
南方	広東	312　(154/158)
	江蘇	262　(134/128)
	浙江	160　(84/76)
	福建	155　(95/60)
	台湾	105　(31/74)
	その他南方	6　(2/4)
	計	1,000　(500/500)
合計		1,500　(750/750)

- 発話文リスト
  - ・旅行会話文(134文/話者)
  - ・音素バランス文(66文/話者)
- 発話時間
  - ・約 513.3時間
  - (北方：約171.9時間、南方：約341.4時間)
  - ※発話前後の無音区間含む
- 総発話数
  - ・300,000発話(200/話者)
  - (北方：100,000　南方：200,000)

- 年代別話者数内訳

	年代	話者数　(男性/女性)
北方	18～20	77 (41/36)
	21～30	307 (158/149)
	31～40	72 (39/33)
	41～50	28 (9/19)
	51～60	16 (3/13)
	計	500 (250/250)
南方	18～20	374 (210/164)
	21～30	415 (215/200)
	31～40	127 (51/76)
	41～50	46 (15/31)
	51～60	38 (9/29)
	計	1000 (500/500)
合計	18～20	451 (251/200)
	21～30	722 (373/349)
	31～40	199 (90/109)
	41～50	74 (24/50)
	51～60	54 (12/42)
		1500 (750/750)

北京
広東
江蘇
福建

収録条件

- 収録環境：残響の少ない静音環境
- マイクロフォン：接話型ヘッドセット(ATM73a/AUDIO TECHNICA)
- サンプリング周波数：48kHz
- 量子化ビット数：16bit
- 無音区間：発話の前後に0.5～3.0秒の無音区間あり

データ構成

- 発話音声：WAV形式、48kHz、16bit、モノラル
- 書き起こしデータ
  - ファイル形式　TSV(タブ区切り)形式
  - 文字コード　UTF-8(BOM無し)
  - 改行コード　LF
  - 発音(ピンイン)表記含む
- 話者情報：PDF形式 (性別、年代、出身地など)

納品物

- 構成：データ構成物、マニュアル
- 納品メディア：BD又はHDD

(北方：約56.9GB、南方：約113.0GB、北方+南方：約169.9GB)

価格 (消費税抜)

商用利用：(北方) ¥2,700,000	研究利用：(北方) ¥1,62,000
(南方) ¥5,400,000	(南方) ¥3,240,000
(北方+南方) ¥8,000,000	(北方+南方) ¥4,800,000

※　地域（北京など）毎のサブセットもご用意しています。詳しくはお問合せ下さい。

※　アカデミック価格についてはお問合せ下さい。

ご利用方法

株式会社ATR-Promotionsより用途に応じた条件によるライセンス(利用許諾)をいたします。

ホームページまたは下記よりご連絡ください。

●本データベースの仕様/価格は予告なしに変更されることがあります。

株式会社 ATR-Promotions
https://www.atr-p.com

〒619-0288　京都府相楽郡精華町光台2-2-2
TEL：0774-95-1300　FAX：0774-95-1191
E-mail：voice@atr-p.com