Vol.2020-SLP-133 No.6
2020/10/22

超高齢者音声コーパスEARSの構築と音声認識への利用の予備検討

福田 芽衣子[1,a)]　入部　百合絵[2]　西崎　博光[3]　山本　一公[4]　西村　良太[1]　北岡　教英[5]

概要：高齢者の音声は一般成人と異なる複数の特徴を持つため，その認識精度は現在のところ不十分と言わざるを得ない．その精度向上には大量の高齢者音声データが必要であり，大規模な高齢者音声コーパスとして，話者の平均年齢 67.6 歳の S-JNAS が汎用されている．しかし現在の日本の平均寿命との間に大きな年齢差が生じていることから，我々は超高齢者を対象とした音声コーパス（EARS: Elderly Adults Read Speech）の構築を開始した．コーパスのデザインは S-JNAS を参考にし，現在までに 121 名（平均年齢：83.4 歳）の音声を収集・データベース化した．本報告ではその仕様について述べるとともに，本コーパスを用いた高齢者音声の音響モデルの予備的検討についても報告する．

キーワード：超高齢者，音声コーパス，音声認識

Meiko Fukuda[1,a)]　Yurie Iribe[2]　Hiromistu Nishizaki[3]　Kazumasa Yamamoto[4]
Ryota Nishimura[1]　Norihide Kitaoka[5]

1. はじめに

高齢者の間にも電子機器の普及が進んでいる昨近，街中で高齢者がスマートフォンを操作する姿を見かけることが珍しくない．しかしながら，加齢や疾患により視力や手指の運動機能が低下してしまうと，タッチパネルなどの小さなデバイスの操作が煩雑な作業になることは想像に難くない．電子機器を日常生活の中で使う機会として，例えば，従来は店舗や施設に直接出向いて行ってきた様々な手続きやサービスが，現在は自宅にいながらウェブ上で行えるようになっている．高齢者が音声認識技術によりこれらを気軽に使えるようすることで，外出が難しい状態の方などの生活の手助けができるだろう．また，経済産業省および厚生労働省の高齢者介護支援ロボット開発促進事業の一環として，高齢者とコミュニケーションをするロボットが重点課題に追加された[*1]．これらに音声認識は不可欠の要素である．こうしたことから我々は，高齢者を対象にした音声認識精度の向上が急務であると考えている．

しかし，従来の音声認識器は成人音声から得られる音響モデルを用いており，高齢者音声の認識精度は不十分と言わざるを得えない [1], [2], [3], [4]．老化に伴い，高齢者の音声には基本周波数やフォルマント周波数の変化，喉頭雑音の増加，発音の不明瞭化，話速の低下など多種多様な特徴が顕著になることが報告されていることから [5], [6], [7], [8], [9]，大規模な高齢者音声コーパスがこの認識精度の向上に不可欠であり，既に新聞読み上げ高齢者音声コーパス（S-JNAS）が広く研究に用いられている [10]．しかしその話者の平均年齢は 67.6 歳と，日本の平均寿命（男性：81.41 歳，女性：87.45 歳）とは開きがある．

そこで，我々は S-JNAS よりも高い年齢層の音声コーパス構築の必要性を感じ，全国各地にて音声の録音を実施しデータベース化を行ってきた．本稿では，この高齢者音声コーパス EARS(Elderly Adults Read Speech) について説明する．高齢者が読み上げた音声を，その書き起こしなどを含めた，音声認識の音響モデル学習に適したコーパスとなっている．

また我々の収集した音声ならびに既存の音声コーパス（JNAS[11]，S-JNAS および CSJ[12]）を用いた高齢者音声

[1] 徳島大学，Tokushima university
[2] 愛知県立大学，Aichi Prefectural University
[3] 山梨大学，University of Yamanashi
[4] 中部大学，Chubu university
[5] 豊橋技術科学大学，Toyohashi university of Technology
[a)] fukuda.meiko@tokushima-u.ac.jp
[*1] https://www.meti.go.jp/press/2017/10/20171012001/20171012001.html

表 1　コーパスの概要

録音地域	愛知，徳島，三重，千葉
話者	121 人，平均年齢：83.4 歳
話者の選択基準	出来るだけ高齢，かつ収録が心身の負担にならない程度に健康な方（認知症傾向者を含む）
学習データ	ATR 音素バランス文 1 人につき 50 または 53 文
テストデータ	JNAS 収載新聞記事文 一人につき 10 文
音声データ	16kHz，16bit，wav ファイル
コンテンツ	a. 発話毎に区切りファイル化 b. テキストファイル 3 種 （漢字かな混り，ひらがな，カタカナ） c. 話者の情報（年齢，性別，録音地域，読み上げ文セット名）

表 2　EARS　話者の年齢層構成

年齢	男性	女性	合計
70-74	1	9	10
75-79	11	14	25
80-84	9	22	31
85-89	5	30	35
90-94	5	10	15
95-99	2	3	5
合計	33	88	121

認識の音響モデルの作成方法を検討した．さらにそれらの認識率と年齢の相関から，加齢の音声認識への影響を分析した．

2. 超高齢者音声コーパス EARS の構築

2.1 音声データの収集

本コーパスの概要を表 1 に示す.

2.1.1 話者の選定基準

本コーパスの目的上，話者はできるだけ高齢の，特に 80 歳以上の方の参加を協力施設にお願いした．他の選定基準は一つのみで，我々による音声の録音が参加者の身体的・精神的な負担とならない健康状態の方のみに参加して頂くこととした．より多くのデータを収集するために，参加可能な方であれば性別，身体的状況（認知症傾向の有無，視力，聴力，義歯など）や，生活状況（自宅または介護施設に居住）などの項目に基準は設けていない．なお，実際にはご協力いただく高齢者施設の職員から，ご本人およびご家族に録音の趣旨を事前に説明いただき，本コーパスの趣旨に賛同された方のみに参加をお願いし，さらに当日にも参加の可否を確認，同意書に署名頂いた上で録音を行った．

2.2 話者の人数，性別，年齢，録音地域分布

現在までに愛知県名古屋市，徳島県徳島市および周辺の市町村，千葉県木更津市，三重県鈴鹿市および四日市市にて録音を行い，読み上げ文の音読が可能であった 121 名による音素バランス文（ATR503 文）のうち各人 50 文程度読み上げた音声をデータベース化した．話者の年齢は 70〜98 歳，平均 83.4 歳（男性：83.3 歳，女性：83.4 歳）であった．表 2 に年齢層別の構成を，表 3 に地域別の話者の人数および平均年齢を示す．千葉県，三重県をはじめ，今後も全国各地で収録を重ねる予定である．

表 3　EARS　地域別の話者の人数

地域	男性	女性	合計（人）	平均年齢
愛知	16	50	66	81.9
徳島	13	25	38	85.9
千葉	1	5	6	81.0
三重	3	8	11	85.5

テストデータの新聞記事文については，徳島で 5 名，千葉ならびに三重では参加者全員の音声を収録した．名古屋については当初の計画の不備でテストデータ収録は行っていない．

また，前報 [13] に記載の長崎および山形の音声は民間企業との共同研究にて収集していた為，音声使用・公開の権利の関係上公開が不可となり，後述の音声認識実験にのみ使用し，超高齢者音声コーパス（EARS）の構成からは除外した．

2.3 日本語読み上げ文

本コーパスはコーパスのデザインを S-JNAS に倣い，日本語読み上げ文として ATR503 文 [14] と J-NAS 新聞記事文 [11] を採用した．ATR503 文は，日本語の 2 音素連鎖 402 種と 3 音素連鎖 223 種，計 625 種をバランス良く含む文 503 文のセットで，音声認識や分析に汎用されている．1 セット 50 文または 53 文から成るテキストセットが 10 セット用意されている（セット名：A〜J）．S-JNAS では ATR503 文を一人につき 2 セット（約 100 文）読み上げた音声が収録されている．また，JNAS 新聞記事文は，毎日新聞記事より抜粋した 16,176 文を，1 セットにつき約 100 文に分割したテキストセットで，新聞記事読み上げ音声コーパス（JNAS）および S-JNAS に収載されている．

S-JNAS では話者一人につき 1 セット（100 文）分の音声を収録しているが，本コーパスは超高齢者や認知症傾向の方などに参加いただいている都合上，参加者への負担を掛けないよう読み上げ文の量を削減する必要があった．そこで，限られた中でも日本語の音素を出来る限り多く収集する為に，学習用データを ATR 音素バランス文とするが，その量は一人につき 1 セット（約 50 文）のみとした．（セット別の発話数を表 4 に示す．）新聞記事文は ATR 音素バランス文より読み上げるのが難しい文が含まれているため，抜粋した文を 1 セット 10 文とし，5 セット（セット名：T1〜T5）準備し，一人に付き 1 セットにとどめ，本コーパスではテストデータに位置付けた．

表 4　EARS　学習データ読み上げセット数

セット名（発話数）	男性	女性	合計
Set A (50)	3 (150)	10 (500)	13(650)
Set B (50)	3 (150)	9 (450)	12 (600)
Set C (50)	4 (200)	7 (350)	11(550)
Set D (50)	2 (100)	10 (500)	12 (600)
Set E (50)	6 (300)	7 (350)	13(650)
Set F (50)	3 (150)	12 (600)	15(750)
Set G (50)	4 (200)	10 (500)	14(700)
Set H (50)	3 (150)	7 (350)	10 (500)
Set I (50)	3 (150)	9 (450)	12 (600)
Set J (53)	2 (106)	7 (371)	9 (477)
総発話数	1656	4421	6077
[話者数]	[33]	[88]	[121]

2.4　音声収録実施の手順

始めに話者に手順を説明し，次に平仮名のルビ付きの読み上げ原稿を渡して練習をしていただいてから，収録を開始した．途中，話者の体調に応じ適宜休憩を取り，録音を続けられるか話者に確認を取りながら行った．

読み上げ後に，認知症簡易テストとして HDS-R（長谷川式簡易認知評価スケール）にも回答して頂いた．また収録には協力施設の職員に同席して頂き，職員からみた話者の感情的な様子を記録用紙に記入してもらった．ただし，これらの付加情報は公開されるデータベースには含まれない．

2.5　収録使用機器および収録環境

名古屋での録音では卓上タイプのマイク（Audio-Technica 社 AT9930）とレコーダー（TASCAM 社 DR-05VERION2）を用いた．徳島，千葉および三重では上記卓上マイクと共にピンマイク（SONY, ECM-88B）を，レコーダーは TASCAM 社 DR680MKⅡ を用いた．

録音は各高齢者施設内の部屋をお借りして行ったので，室外からの話者以外の音声や，空調などの生活雑音が含まれている場合があるが，おおよそ 40〜45dB の騒音レベルであった．

2.6　音声データ詳細，テキストファイル，話者情報

現在までにコーパスに収載する音声データは，学習用データとして ATR503 バランス文を合計 6,077 文，テストデータ用として新聞記事文を合計 220 文収録した．各セットの読み上げ人数を表 4 および表 5 に示す．（音声認識実験のみに用いた音声については後述する．）

音声データのサンプリング周波数は 16kHz，量子化ビット数は 16 ビットで，文単位でファイル化し，各ファイルの音声区間の前後には可能な限り約 150ms の無音空間を含めた．

テキストファイルについては，話者の発音や言い間違い等を忠実に書き起こし，漢字かな混じり，ひらがな及びカタカナ表記の 3 種類を作成した．テキストは，基本的には読み上げた文であるが，高齢のため読み間違いも多い．その場合は，可能な限り，実際に発音した内容に忠実にテキストを修正している．話者個人の情報としては，話者の年齢，性別，収録地，読み上げ文セット名を収載した．

表 5　EARS　テストデータ読み上げセット数

セット名（発話数）	男性	女性	合計
T1 (10)	2 (20)	3 (30)	5(50)
T2 (10)	0 (0)	5 (50)	5(50)
T3 (10)	0 (0)	4 (40)	4(40)
T4 (10)	4 (40)	4 (40)	6 (60)
T5 (10)	1 (10)	2 (20)	3 (30)
総発話数	60	160	220
[話者数]	[6]	[16]	[22]

表 6　音響モデル作成に使用した高齢者音声

地域	男性	女性	合計（人）	平均年齢
愛知	15	49	64	81.8
徳島	13	26	39	85.5
長崎	32	76	108	76.0
山形	10	0	10	73.4
4 地域合計	70	151	221	79.2

3.　音声認識利用への予備的検討

高齢者音声認識に用いる音響モデルの作成と適応化を予備的に検討した結果を報告する．

3.1　実験設定

3.1.1　音声認識実験に使用した高齢者音声

本稿の実験では．2.2 節に述べた愛知および徳島にて録音した音声と，公開データに含まれていない長崎および山形にて収録した音声の両者を合わせて「高齢者音声」と呼ぶこととする．これらの総録音時間は 21.7 時間であった．話者の人数と平均年齢を表 6 に，学習データ読み上げセット数を表 7 に，テストデータ読み上げセット数を表 8 に示す．なお，実験で用いた徳島および名古屋収録の人数と EARS 収載の人数が若干名異なる．これは，EARS の構築目的を鑑み，公開データから 70 歳未満の話者を除外し，加えて 70 歳以上の新たな話者をデータベース化したためである．

3.2　音響モデル作成

我々の作成した高齢者音声データは総録音時間が 21.7 時間と，このデータのみで音響モデルを作成するには十分なデータ量ではない．そこで，既存の大規模音声コーパスと我々の高齢者音声を共に音響モデル学習に用いることで，高齢者に適した音響モデルが作成できないか試みた．

既存の音声コーパスからは，読み上げ音声コーパスであ

表 7　テストデータ読み上げセット数

セット名（発話数）	男性	女性	合計
T1 (10)	5 (50)	3 (30)	8(80)
T2 (10)	1 (10)	2 (20)	3(30)
T3 (10)	0 (0)	2 (20)	2(20)
T4 (10)	2 (20)	1 (10)	3 (30)
T5 (10)	4 (40)	0 (0)	4 (40)
総発話数	120	8	200
[話者数]	[12]	[8]	[20]

表 8　既存音声コーパス音響モデル及び BCWJ 言語モデルならびに高齢者音声適応化音声認識結果 WERs(%)

	JNAS	S-JNAS	CSJ
ベースライン	25.5	21.9	27.3
適応後	21.6	20.2	17.2

る JNAS あるいは S-JNAS と，自発性音声コーパスである CSJ[12] を選択した．CSJ は総録音時間 113.4 時間の，学会や模擬講演などの自発性音声のコーパスである．この CSJ を読み上げ音声である高齢者音声の音響モデルに用いた理由については，実際に超高齢者に定型文の読み上げをして頂いたところ，読み間違いや言い淀み，フィラー，発音の休止区間の多発や延長がかなりの頻度みられ，JNAS 及び S-JNAS の流暢な読み上げ音声とは発話スタイルが同質とは言い難いことが判明した．そこで，フィラーや言い淀みなどを呈する自発的発話の CSJ が高齢者音声に適する可能性があると考えたからである．

3.3　音声認識実験

今回も前報 [13] と同様に，各音響モデルの作成は Kaldi の CSJ レシピ [8] に基づき DNN-HMM を用いた．DNN には simple feed-forward network(nnet1) を使用した．言語モデルも前報と同様に現代日本語書き言葉均衡コーパス（BCCWJ）を用いた．BCCWJ は現代日本語の書き言葉のテキストコーパスとして最大規模のコーパスである [15]．

3.3.1　JNAS，S-JNAS および CSJ を用いたベースライン音響モデルの実験結果

まず，JNAS，S-JNAS および CSJ を各々単独で学習データとして用いた音響モデルをベースラインとし，これらを高齢者音声へ適応させた後に，高齢者音声の認識実験を行った．ここで，適応には適応データを用いてバックプロパゲーションによる追加学習を用いた．この結果を表 8 に示す．

ベースラインでは，予想よりと異なり S-JNAS 音響モデルの WER が最も低い値を示した．しかし各モデルを高齢者音声に適応させると，CSJ 音響モデルが最も低い値を示した．CSJ の自由発話の現象と，高齢者音声への適応が相乗効果を示し，高齢者音声への CSJ の有効性が示された．

表 9　’mixed’ あるいは’高齢者音声なし’音響モデルおよび BCCWJ 言語モデル音声認識結果，ならびに高齢者音声適応化の認識結果 WERs(%)

	ベースライン	適応後
mixed	13.4	14.9
高齢者音声なし	14.4	14.3

表 10　’高齢者音声なし’音響モデルおよび BCCWJ 言語モデルならびに各地域の高齢者音声適応化音声認識結果 WERs(%)

	高齢者音声なし	適応後 長崎	徳島	山形
長崎	11.6	**11.0**	13.7	116
徳島	31.0	32.2	**28.0**	36.2
山形	5.6	6.4	7.9	**4.9**
計	14.4	**12.4***		

*各地域の音声をその地域音声へ適応して得られた値（太字）を平均した値

3.3.2　既存コーパスならびに高齢者音声を用いた音響モデルの実験結果

次に，ベースラインの音響モデルに大量の音声データを用いることで精度が向上するか検討した．具体的には，前述の既存の 3 つの音声コーパス（JNAS，S-JNAS および CSJ）と我々の高齢者音声データをまとめて一つの音響モデル作成に用いた（以後，これを‘mixed’と呼ぶ）．また，高齢者音声の有効性を確認する為に，既存の 3 コーパスのみの音響モデルも作成し（以後，‘高齢者音声無し’），音声認識実験を行い’mixed’と比較した（表 9）．その結果，’mixed’ 音響モデルの WER が 13.4%とこれまでで最も低い WER が得られ，大量の音声コーパスとともに高齢者音声を用いた音響モデル作成の有用性が示された．

また，‘mixed’および‘高齢者音声無し’について話者の年齢と WER の相関について散布図を作成し，音声認識への年齢の影響を分析した（図 1 および図 2）．相関性を検討するには話者数が十分では無いものの，‘mixed’および‘高齢者音声無し’の両者に年齢と共に WER が上昇する傾向がみられた．両モデルの近似線をみると，その傾きは‘mixed’の方が緩やかであり，高齢者音声を音響モデルに用いることで年齢に伴う認識率の低下を補う傾向がみられた．今後，特に 80 歳以上の話者を増やしたのちにこの傾向について再度確認したいと考えている．

更なる認識精度向上のために，この二つの音響モデルを DNN のバックプロパゲーションによる追加学習で高齢者音声に更に適応させた（表 9）．しかしながら，‘mixed’の WER が 14.9%，‘高齢者音声無し’は 14.3%と，適応前の‘mixed’の 13.2%以上の精度は得られなかった．適応の方法については今後の課題である．

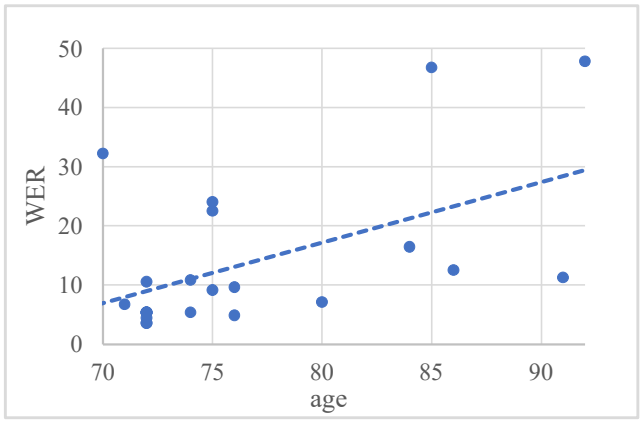

図 1 ’高齢者音声無し’モデルにおける年齢と WER の相関 (corr.=0.32)

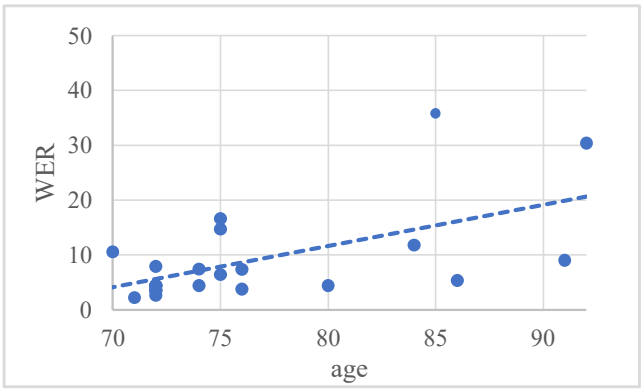

図 2 ’mixed’ モデルにおける年齢と WER の相関 (corr.=0.53)

3.3.3 方言の音声認識への影響

また，各地域の方言の音声認識への影響をみるために，長崎，徳島および山形音声の‘高齢者音声無し’音響モデルを各地域の音声で追加学習した（表 10）．適応方法は前節の実験と同様である．実験の結果は，例えば，徳島音声の‘高齢者音声無し’の追加学習前は WER が 31.0%だったが，これを徳島音声で追加学習すると WER が 28.0%に，山形音声で追加学習後は 36.2%と上昇した．長崎および山形音声でも同様の傾向がみられた（’mixed’ については適応の効果がみられなかったので結果は示していない）．今回のモデルに関しては，３地域の方言は音声認識に影響を与えたと言える．より効果的な適応法を用いることで，地域適応の効果がさらに明確になる可能性があるだろう．

なお，３地域の‘mixed’および‘高齢者音声無し’モデルの認識結果を比較すると，徳島は他の２地域より WER 値が顕著に高い．これについては，徳島の話者の平均年齢が 85.5 歳と長崎や山形より 10 歳ほど高齢であること（表 10），非常に高齢の参加者につき原稿の読み間違いが多くみられてこれに対応するテキストを修正したものの，不明瞭な発話が多くなってしまったことなどが大きな要因と考えられる．

4. まとめ

我々は，これまでに全国４県にて超高齢者の音声を収集・データベース化を行い，超高齢者音声コーパス（EARS: Elderly Adults Read Speech）を構築してきた．本コーパスは，S-JNAS よりも高齢の年齢層を対象としていて，高齢者が読み上げた音声を，その書き起こしなどを含めた，音声認識の音響モデル学習に適したコーパスとなっている．今後も日本各地で収録を行い，コーパスの規模を拡大し，一般公開を目指している．

本稿では，高齢者音声認識の音響モデル作成の予備的検討として，既存の音声コーパス（JNAS，S-JNAS および CSJ）と我々の構築した高齢者音声データを用いた．高齢者音声データ量が少量であるにも関わらず，高齢者音声の認識精度が若干向上した．また，年齢に伴う WER の低下傾向が示されたので，今後も話者数を増やして検討を続けたい．さらに，各地域の方言が音響モデルの追加学習時に認識精度に影響を与えることが示唆された．

今後の課題としては，コーパスの規模を拡大することで音響モデルの更なる高精度化をはかること，また今回得られなかった音響モデルの有効な追加学習方法について検討したい．

5. 謝辞

本研究は JSPS 科研費 17H01977，19H0112 および 19K12022，2019 年度国立情報学研究所公募型共同研究（19S0403）の助成を受けたものです．

参考文献

[1] Anderson, S., Liberman, N., Bernstein, E., Foster, S., Cate, E., Levin, B. and Hudson, R.: Recognition of elderly speech and voice-driven document retrieval, *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No. 99CH36258)*, Vol. 1, IEEE, pp. 145–148 (1999).

[2] Vipperla, R., Renals, S. and Frankel, J.: Longitudinal study of ASR performance on ageing voices (2008).

[3] Wilpon, J. G. and Jacobsen, C. N.: A study of speech recognition for children and the elderly, *1996 IEEE international conference on acoustics, speech, and signal processing conference proceedings*, Vol. 1, IEEE, pp. 349–352 (1996).

[4] Pellegrini, T., Trancoso, I., Hämäläinen, A., Calado, A., Dias, M. S. and Braga, D.: Impact of age in ASR for the elderly: preliminary experiments in European Portuguese, *Advances in Speech and Language Technologies for Iberian Languages*, Springer, pp. 139–147 (2012).

[5] Winkler, R., Brückl, M. and Sendlmeier, W.: The aging voice: an acoustic, electroglottographic and perceptive analysis of male and female voices, *Proc. of ICPhS*, Vol. 3, pp. 2869–2872 (2003).

[6] Miyazaki, T., Mizumachi, M. and Niyada, K.: Acoustic Analysis of Breathy and Rough Voice Characterizing Elderly Speech., *JACIII*, Vol. 14, No. 2, pp. 135–141

(2010).
[7] Eichhorn, J. T., Kent, R. D., Austin, D. and Vorperian, H. K.: Effects of aging on vocal fundamental frequency and vowel formants in men and women, *Journal of Voice*, Vol. 32, No. 5, pp. 644–e1 (2018).
[8] 濱崎健太：高齢者の 「めりはりの無い声」 を表す物理量に関する考察，音講論 (春), 2010 (2010).
[9] 児嶋久剛：高齢者と気管食道科 高齢者の喉頭 (発声) 機能，日本気管食道科学会会報，Vol. 45, No. 5, pp. 360–364 (1994).
[10] Baba, A., Yoshizawa, S., Yamada, M., Lee, A. and Shikano, K.: Elderly acoustic model for large vocabulary continuous speech recognition (2001).
[11] Itou, K., Yamamoto, M., Takeda, K., Takezawa, T., Matsuoka, T., Kobayashi, T., Shikano, K. and Itahashi, S.: JNAS: Japanese speech corpus for large vocabulary continuous speech recognition research, *Journal of the Acoustical Society of Japan (E)*, Vol. 20, No. 3, pp. 199–206 (1999).
[12] Furui, S., Maekawa, K. and Isahara, H.: CSJ 文献 A Japanese national project on spontaneous speech corpus and processing technology，*ASR2000-Automatic Speech Recognition: Challenges for the new Millenium ISCA Tutorial and Research Workshop (ITRW)* (2000).
[13] Fukuda, M., Nishizaki, H., Iribe, Y., Nishimura, R. and Kitaoka, N.: Improving Speech Recognition for the Elderly: A New Corpus of Elderly Japanese Speech and Investigation of Acoustic Modeling for Speech Recognition, *Proceedings of The 12th Language Resources and Evaluation Conference*, pp. 6578–6585 (2020).
[14] Kurematsu, A., Takeda, K., Sagisaka, Y., Katagiri, S., Kuwabara, H. and Shikano, K.: ATR Japanese speech database as a tool of speech recognition and synthesis, *Speech communication*, Vol. 9, No. 4, pp. 357–363 (1990).
[15] Maekawa, K., Yamazaki, M., Ogiso, T., Maruyama, T., Ogura, H., Kashino, W., Koiso, H., Yamaguchi, M., Tanaka, M. and Den, Y.: Balanced corpus of contemporary written Japanese, *Language resources and evaluation*, Vol. 48, No. 2, pp. 345–371 (2014).