

平成 23 年 6 月 3 日現在

機関番号：33907
 研究種目：若手研究（B）
 研究期間：2007～2010
 課題番号：19700172
 研究課題名（和文） 長・短期間の音声変動の解明および音声変動に頑健な音声認識手法に関する研究
 研究課題名（英文） Study of specific speaker's speech variability over long and short time periods for speech recognition
 研究代表者
 柘植 覚 (TSUGE SATORU)
 大同大学・情報学部・准教授
 研究者番号：00325250

研究成果の概要（和文）：

同一発声者が特定の音声認識（話者依存音声認識）システムを使用した場合においても必ず毎回同一の音声認識結果とならない。このような音声認識結果の異なりは、「さっき認識したのに、なんで今回認識しないのだろうか？」というような音声認識システムへの不信感につながり、音声認識が広く一般に普及しない原因の一つとなっている。音声は個人性により大きく変動することもさることながら、個々人が同一内容を発声したとしても、気分や発声環境により変動（音声変動）する。これが認識結果の変動を引き起こす原因となっている。

本研究では、特定話者ならば長期間に渡り、安定した高精度の音声認識ができる音声認識手法の確立を目標とする。申請期間において、科学的側面から特定話者（個々人）の音声が発声時期、時間、体調、気分、環境などの違いでどのように変動するかを明確にし、これらの音声変動に頑健な特定話者音声認識手法を確立する。

研究成果の概要（英文）：

Even if a speaker uses a speaker-dependent speech recognition system, speech recognition performance varies. However, the relationships between intra-speaker's speech variability and speech recognition performance are not clear. In this study, I investigate relationships between intra-speaker's speech variability and the speech recognition performance.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2007年度	1,300,000	0	1,300,000
2008年度	800,000	240,000	1,040,000
2009年度	600,000	180,000	780,000
2010年度	600,000	180,000	780,000
年度			
総計	3,300,000	600,000	3,900,000

研究分野：総合領域

科研費の分科・細目：情報学・知覚情報処理・知能ロボティクス

キーワード：特定話者、話者内音声変動、音声認識、音声分析

1. 研究開始当初の背景

現在、カーナビゲーションや携帯電話などに音声認識システムが搭載されているが、利用者が積極的に使用しているの話はあまり聞かず、実際には音声認識システムが普及しているとは言い難い。この原因の一つとして、不特定話者の連続音声認識を行うシステムはもちろん、特定話者の認識においても認識精度が十分でないことが挙げられる。さらに、同一発声者が特定の音声認識（話者依存音声認識）システムを使用した場合においても必ず毎回同一の音声認識結果とならないことが挙げられる。このような音声認識結果の異なりは、音声認識システムへの不信感につながり、音声認識が広く一般に普及しない原因の一つとなっていると思われる。音声は発声者の異なりによる変動もさることながら、個々人が同一内容を発声したとしても、気分や発声環境により変動（音声変動）する。これが認識結果の変動を引き起こす原因となっている。

2. 研究の目的

本研究では、特定話者ならば長期間に渡り、安定した高精度の音声認識ができる音声認識手法を確立する。不特定話者ではなく特定話者に着目し研究を進め、今後家庭内に音声認識技術が付与された情報家電などが普及した場合、不特定話者の発声をそれなりの精度で認識できる技術より、特定話者（家族など）の発声を確実に認識する音声認識技術が重要になってくると考えられるからである。そこで、科学的側面から特定話者（個々人）の音声は時期、時間、体調、気分、環境などの違いでどのように変動するかを明確にし、これらの音声変動に頑健な特定話者音声認識手法を確立する。

また、長期間の音声変動を調べるためには、同一話者が一定間隔で同一内容を長期間に渡り発声した音声データベースが必要である。そこで、本申請期間の4年間で話者内の音声変動を明確にするための音声データベース構築を行う。またこのデータベースを用い、話者内の音声変動を明確にし、時間経過や体調などによる話者内音声変動に頑健な音声認識手法の確立を目指す。さらに、今までに収録されていない音声認識が使用される環

境下を想定した発声（横になった姿勢や立った姿勢での発声、雑音・音楽を試聴しながらの発声、運動後の発声など）を収録し、日常生活で音声認識システムが使用される環境を想定し、それらの環境における音声変動を明確にする。このデータベースが完成すれば世界に唯一無二のデータベースとなる。

3. 研究の方法

●長期間特定話者音声データベースの構築

申請期間である4年間、毎週1日、1日3回、特定話者3名（男性2名、女性1名：内1名は申請者である柘植）の音声収録し、長期間特定話者音声データベースを構築する。各収録時には音声変動の要因を調査するため、発声者へのアンケート（発声時の体調情報や環境情報）を実施する。また、音声データベースの構築時の人的な労力軽減のため、本申請期間中に音声データベース構築支援ツール（自動発声ファイリングツール）の研究開発を行う。

●特定話者の音声変動の解明および音声変動と音声認識精度の関連

発声時期や発声時間帯が異なることによる音声変動を明確にし、この音声変動が生じる要因を明確にする。具体的には発声する日や時間帯による音声変動を調査し、時期（発声日）による音声変動の影響や時間帯（一日内で発声する時間帯）による影響を詳細に調査分析する。この際に、発声者の気分や体調などの内的要因や発声時の室温湿度などの環境的要因（外的要因）が発声変動に及ぼす影響を調査する。また、横になりながらの発声や運動後の発声などの音声認識装置が使用される可能性のある環境下での音声と通常時（椅子に座りながら落ち着いて発声）の音声との比較を行い、音声はどのように変動しているかを詳細に分析し、音声変動を明確にする。

●発声変動に頑健な音声認識手法の確立

発声変動と音声認識精度との関連性を明確にした後、音声認識で用いる音響特徴空間上ではどのように変動している

かを明確にする。音響特徴空間において変動を最小にする超空間を探し、その空間へ音響特徴空間を射影し正規化する手法を提案する。また、音声変動に逐次追従可能な音響モデルの適応手法の検討を行い、音声変動に対し音声認識精度が変動しない頑健な音声認識手法を確立する。

4. 研究成果

長期間にわたる音声収録により、大量の特定話者音声データの収集およびそれらを整備し、データベースとして使用できるようにした。

この音声データベースを利用して、音声認識実験を行い、特定話者に対する音声認識を行う際には、時間帯の異なりよりも認識対象発声にできるだけ近い日の音声により適応を行うことにより認識性能が向上することが分かった。

さらに、話者認識においては日によって変化しない音響空間を見つけ、その空間を利用することにより、話者認識精度が向上することを確認した。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 件)

[学会発表] (計 18 件)

1. Satoru Tsuge, Keiji Seida, Masami Shishibori, Kenji Kita, Fuji Ren, Kenji Kita, Minoru Fukumi, Shingo Kuroiwa, Analysis of Variation on Intra-Speakers Speech Recognition Performances, IEEE NLP-KE2007, 2007.8
2. Mitsuhiro Ozawa, Satoru Tsuge, Masami Shishibori, Kenji Kita, Minoru Fukumi, Fuji Ren, Shingo Kuroiwa, Automatic Utterance Segmentation Tool for Speech Corpus, IEEE NLP-KE2007, 2007.8
3. Katsuyuki Oobayashi, Satoru Tsuge, et.al, A Study of speaker identification using phoneme-information, NCSP'2008, 2008.3
4. 柘植 覚 他, 日本語大規模話者認識用データベースを用いた話者識別実験, 日本音響学会 春季研究発表会, 2007年9月
5. 喜多 雅彦, 柘植 覚 他, 大規模話者骨導音声データベースを用いたテキスト独立型話者照合実験, 情報処理学会音言言語情報処理, 2007年12月
6. 柘植 覚 他, 骨導音と気導音を併用した話者識別手法, 日本音響学会 秋季研究発表, 2008年3月
7. Satoru Tsuge, et.al, Combination method of Bone-conduction Speech and Air-conduction Speech for Speaker Recognition, Interspeech 2008, 2008.9
8. Satoru Tsuge et.al, Specific speakers' speech corpus over long and short time period, Oriental COCODSA, 2008.11
9. Yasunori Kashihara, Satoru Tsuge, et. al, Non-Stationary Noise Robust Speech Recognition Method using Repetitive Phrase, NCSP2009, 2009.3
10. 柏原 康徳, 柘植 覚 他, 繰り返し発声を用いた突発性雑音に頑健な音声認識に関する研究, 平成21年度電気学会電子情報システム部門大会, 2009年9月
11. 中尾 昌史, 柘植 覚 他, 気導音と骨導音を併用した話者ベクトルを用いた話者識別, 平成21年度電気学会電子情報システム部門大会, 2009年9月
12. Satoru Tsuge, et. al, Speaker verification method using bone-conduction and air-conduction speech, ISPACS 2009, 2009.12
13. Haruka Okamoto, Satoru Tsuge, et. al, Text-Independent Speaker Verification Using Rank Threshold in Large Number of Speaker Models, InterSpeech 2009, 2009.9
14. 森 健太郎, 柘植 覚 他, 話者依存音声認識のための発音辞書・音響モデル適応手法, 平成22年電気学会 電子・情報・システム部門大会, 2010年9月
15. Wenbin Zhang, Haoze Lu, Yasuo Horiuchi, Satoru Tsuge, Kenji Kita, Shingo Kuroiwa Text-Independent Speaker Identification Based on Reducing Inter-Session Variability of Speech Feature Using PCA Transformation, 2011 International Workshop on Nonlinear Circuits, Communication and Signal Processing, 2010.3
16. Masato Miyoshi, Satoru Tsuge, et. al, Music Impression Detection Method for User Independent Music Retrieval System, 14th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, 2010.9
17. 柘植 覚 他, フィルタバンク特徴量と Earth Mover's Distanceを用いた音楽検索, 第89回音楽情報科学研究会プログラム, 2011年2月
18. 三好 真人, 柘植 覚 他, 音楽検索のための楽曲印象値の自動付与手法, 第89

回音楽情報科学研究会プログラム, 2011
年 2 月

〔図書〕(計 件)

〔産業財産権〕

○出願状況 (計 0 件)

名称:

発明者:

権利者:

種類:

番号:

出願年月日:

国内外の別:

○取得状況 (計◇件)

名称:

発明者:

権利者:

種類:

番号:

取得年月日:

国内外の別:

〔その他〕

ホームページ等

6. 研究組織

(1) 研究代表者

柘植 覚 (TSUGE SATORU)

大同大学・情報学部・准教授

研究者番号: 00325250

(2) 研究分担者 なし

(3) 連携研究者 なし