

音声からのストレス状態検出システムの開発

松尾直司^{†1} 早川昭二^{†1} 原田将治^{†1} 武田一哉^{†2} 降籬喜和男^{†3}

音声信号処理を用いた見守り技術として、会話音声からのストレス状態の検出技術を開発し、更に、パソコンや普及が進んでいるスマートフォンなどを利用した端末-センター連携システムのプロトタイプを開発した。本開発は、振り込み詐欺誘引通話検出といった安心安全のための独居高齢者の見守りや、顧客満足度向上のための顧客と従業員間のトラブルの早期検出や、走行中のドライバーの見守りなどの実現を目指している。本ストレス状態検出技術では、人がストレスを受けた時にのどが渇いて声の調子の変化、すなわち、音声のピッチとレベルが通常の状態から変わることを基に、ストレス状態らしさを推定する。更に、通常の会話からストレス状態を検出しないように、詐欺やトラブルなどに関連する特徴的なキーワードの検出を行い、これらの結果を統合して最終判定を行う。被験者にストレスを与える模擬会話を行って検出精度を評価した結果、90%以上の精度でストレス状態を検出できることが分かった。また、主に端末で検出処理を行う場合と、端末とセンターで連携して検出処理を行う場合のプロトタイプシステムを開発して動作確認を行った。本稿では、音声のピッチとレベルからストレス状態を検出する技術と、開発した2種類のプロトタイプシステムの構成について述べる。

Development of the System for Stress Situation Detection based on the Voice Information

NAOSHI MATSUO^{†1} SHOJI HAYAKA^{†1} SHOUJI HARADA^{†1}
KAZUYA TAKEDA^{†2} KIWAO FURIHATA^{†3}

1. はじめに

一般家庭、事務所、病院、車室内などの様々な場所や状況における見守りサービスの実現を目指している。例えば、2015年には独居高齢者世帯数が600万を超え、そのときの高齢者の見守りサービスは116億円の市場に拡大すると予測されている[1][2]。また、警察庁の統計資料によると、2013年の振り込み詐欺発生件数は9,223件(2012年:6,348件)であり、実質約260億円(同約160億円)の被害が出ている[3]。深刻な社会問題であり、その抑止が課題である。このような背景の中、日常生活や業務の中で対象者に意識させずに見守りを行う手段の一つとして、電話における会話音声の分析技術を開発している。ここで、迷惑電話や振り込み詐欺誘引通話を受けた場合や、顧客とのトラブル発生時などには、批判や強い要求などの、通話相手から与えられる好ましくない言い方や情報によりストレスを受けた状態になる。この状況を素早く解決するためには、図1に示すように、音声からストレス状態を検出し、周囲の冷静な第三者の注意を喚起してサポートにつなげることが有効と考えている。例えば振り込み詐欺の場合、詐欺の手段の研究結果[4][5]を調べたり、誘引通話の録音データを

試聴したりした結果、最初に考察能力が低下するストレス状態、次に状況認識が誤る錯誤状態に誘引して、振り込みなどの行動に導くことが分かった。したがって、被害者がストレス状態になっていることを自動検出し、家族などの冷静な第三者によるサポートなどに繋げることが被害抑止に必要であると考えている。

著者らは、技術開発に必要なストレス状態の評価データの収録を行い[6]、これらの評価データを用いて音声のピッチとレベルからストレス状態らしさを推定する分析技術を研究している[7]。また、通常の会話からのストレス状態検出を抑えるために、ワードスポッティング型音声認識技

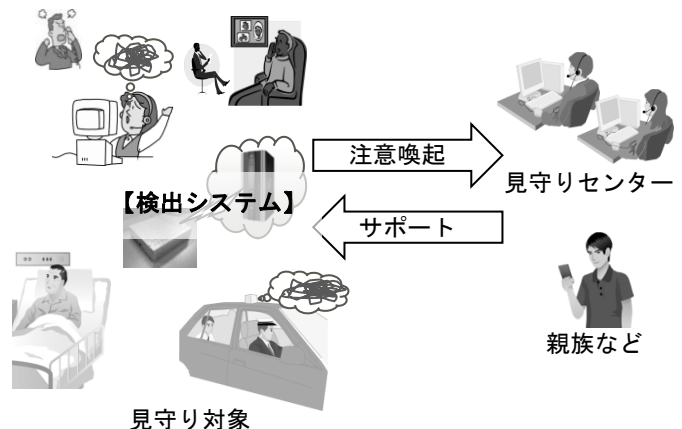


図1 見守りサービス
Figure 1 Monitoring Service.

^{†1} 富士通(株)
FUJITSU LIMITED.
^{†2} 名古屋大学
NAGOYA UNIVERSITY
^{†3} 警察大学校
NATIONAL POLICE ACADEMY

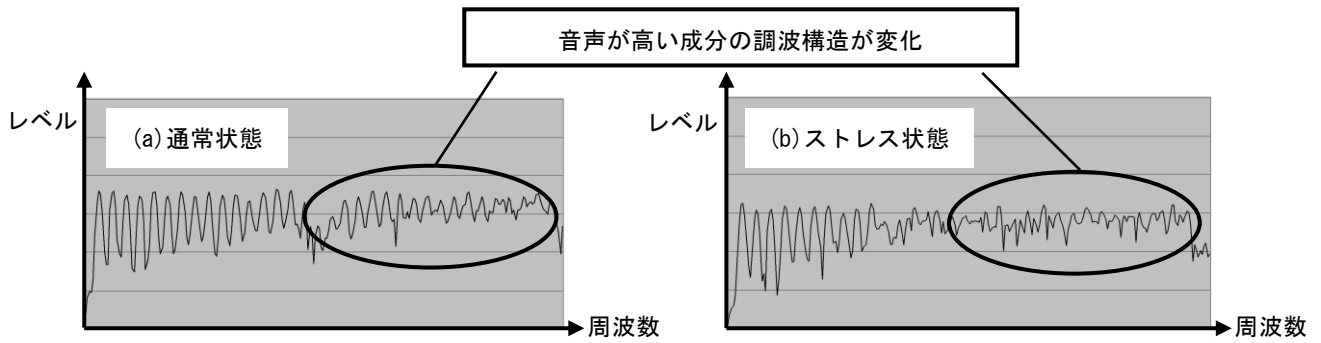


図 2 パワースペクトルとストレス状態の関係

Figure 2 Relation between Power Spectrum and Stress Situation.

術[8]を用いて詐欺やトラブルなどに関連する特徴的なキーワードを検出する技術を統合して、これらの状況におけるストレス状態を検出する技術を研究している[9]。また、研究結果の実証実験として、固定電話に接続して振り込み詐欺誘引通話を検出するプロトタイプシステムを開発して、警察や銀行と協力し、岡山県で一般家庭を対象とした実証実験を行った[10][11][12][13]。更に、パソコンやスマートフォンを用い、端末-センター連携のプロトタイプシステムを開発し、通話音声からストレス状態を検出する実験を行っている。

本稿では、ストレス状態の検出技術と、端末-センター連携のプロトタイプシステムの構成について述べる。

2. ストレス状態における音声の特徴

ストレスを受けた場合、発声器官である声帯に関する筋肉の動きや気道液の粘性が影響を受けるといわれている[14][15][16][17]。したがって、声帯の挙動に依存する音声のピッチやレベルが、通常状態とストレス状態では異なると考えた。

図 2 に、友人と通話時の(a)通常状態と、トラブルを模擬した通話時の(b)ストレス状態の、音声のピッチとレベルに関するパワースペクトルを示す。ここでは、「あ」を発声している区間を分析した結果である。なお、声帯の挙動の影響を分かりやすくするために、線形予測分析によるフォルマントのピーク抑制を行っている。この結果から、通常状態に比べてストレス状態では、声帯の振動によって生じるレベルの山谷の変化、すなわち調波構造が高域で不明瞭になっていて、声帯の挙動がストレス状態の場合に弱くなることが分かった。

3. ストレス状態の検出技術

3.1 音声のピッチとレベルからのストレス状態らしさの推定

図 2 の結果から、ストレス状態が声帯の挙動に影響を与えることが分かった。このことからストレス状態らしさの推定のために、声帯の挙動と強い関連がある音声のピッチの分散とレベルの分散を特徴量として用いた。

分散を計算するためには、最初にピッチとレベルの平均を通話単位で算出し、次に、ピッチとレベルの値の各平均からの距離の 2 乗和を計算し、更に通話単位の平均として分散を算出する。したがって、分散はピッチやレベルの分布が平均値からどれだけ散乱しているかを示す値であり、分散の値が大きいほどピッチやレベルの値の変化幅は大きく、声帯の挙動が強いことを示している。一方で、分散の値が小さい場合は、ピッチやレベルの値の変化幅、すなわち声の調子の変化が小さく、声帯の挙動が弱いことを示している。

男声と女声の模擬通話データから通話単位で算出した、ピッチの分散とレベルの分散およびストレス状態の関係を図 3 に示す。これらの結果から、●印の通常状態と、■印のストレス状態とでは、ピッチの分散とレベルの分散の偏りに違いがあり、ストレス状態の場合に分散が小さくなる傾向であることが分かった。これは、声帯の挙動がストレス状態の場合に弱くなることを示していて、図 2 の結果とも一致する。

したがって、音声のピッチの分散とレベルの分散から、図 4 に示すようにストレス状態らしさや通常状態らしさの推定を行った。

- ・ピッチの分散とレベルの分散の両方の値が小:
ストレス状態らしさが高いと推測。
- ・ピッチの分散とレベルの分散のいずれかの値が大:
ストレス状態らしさが低い(=通常状態らしさが高い)と推測。

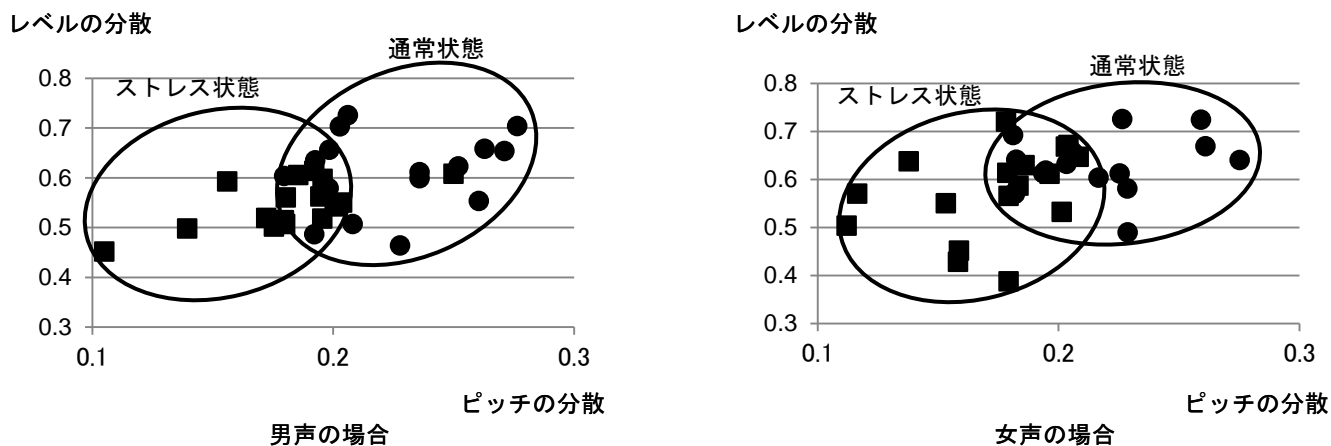


図 3 ピッチ・レベルの分散とストレス状態の関係
 Figure 3 Relation between Variances of both Pitch and Level of Voice, and Stress.

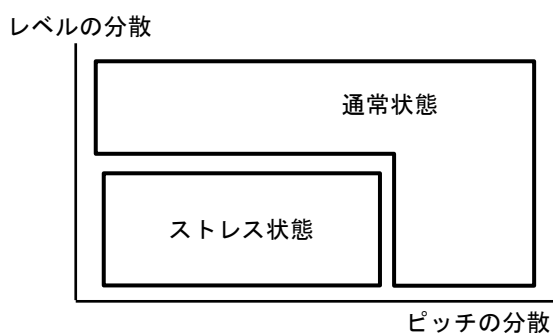


図 4 ストレス状態の検出範囲
 Figure 4 Range of Detection for Stress Situation.

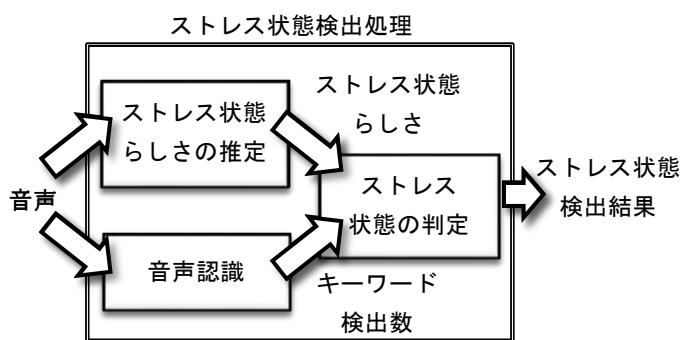


図 5 ストレス状態の検出処理
 Figure 5 Detection Process of Stress Situation.

3.2 音声認識による特徴的なキーワードの検出

3.1 節で示した、図 1 左側の見守り対象者の音声からのストレス状態らしさの推定技術について、複数の迷惑電話や振り込め詐欺誘引通話の模擬通話データを用いた評価を行った。その結果、例えば、インターネット契約の説明を受けるなど、日頃馴染みの薄い内容を扱う通話を行った場合にストレス状態と誤検出する機会が多いことが分かった。これは、不慣れな内容の会話を行うことに、被験者がストレスを感じたためと考えている。従って、ストレス状態らしさの推定結果と統合するために、会話相手の発声に頻出する、詐欺やトラブルに関連する特徴的なキーワードの検出を行った。

迷惑電話や振り込め詐欺誘引通話などのストレス状態につながる通話の事例を調べると[3][18]、金銭関連や不祥事に関する特徴的なキーワードが会話に多く含まれることが分かった。したがって、人同士の自然な会話が対象であることから、ワードスポッティング型音声認識技術を用いて会話音声に含まれる特徴的なキーワードを検出し、キーワード検出数をストレス状態と関係がある値とした。

なお、実際の見守りサービスでは、従業員向けや独居高齢者向けなど、サービス内容によって登録するキーワードのリストを変更する。

3.3 ストレス状態の検出

図 5 に、ストレス状態の検出処理を示す。

3.1 節と 3.2 節で述べたように、音声のピッチとレベルの分析結果からストレス状態らしさを推定する。また、並行して、ワードスポッティング型音声認識により、会話に含まれる特徴的なキーワードの数を検出する。最後に、重み付け加算によってこれらの結果を統合し、しきい値と比較することによりストレス状態か通常状態かを判定する構成とした。

なお、ストレス状態らしさ推定処理の調整と音声認識処理の調整、それぞれの結果を統合する際の重み付けの調整、および統合後にストレス状態か通常状態かを判定する際に用いるしきい値の調整は、事前に調整用データを用いて行う。

4. 評価実験

4.1 シミュレーション結果

被験者を心理的にストレス状態とするための迷惑電話の模擬通話と、通常状態とするための友人との通話を複数回行ったときの会話データを収録して評価実験を行った。

ここで、被験者とその友人は一般から募集し、年齢層と性別が極端に偏らないようにした。また、通話相手は、俳優の男性1名である。なお、迷惑電話の模擬通話については、被験者に事前に通話の詳細を提示せず、不愉快に感じた場合には途中で電話を切るように指示を与えておいた。

今回は、図5の処理の調整に、次に示す通話のデータを用いた。

- ・ 模擬通話: 38 データ
被験者: 25 名
- ・ 友人との通話: 62 データ
被験者: 25 組 (50 名)

調整後に、次に示す別のデータセットを用いて、迷惑電話の模擬通話からのストレス状態の検出率と、友人との通話からのストレス状態の誤検出率を計算機シミュレーションにより評価した。

- ・ 模擬通話: 133 データ
被験者: 73 名
- ・ 友人との通話: 278 データ
被験者: 73 組 (146 名)

評価結果は表1の通りとなり、会話音声を基にしたストレス状態検出の基本技術実現の目処が付いた。

表1 ストレス状態検出の評価結果

Table 1 Evaluation Result of Stress Situation Detection.

検出率	約 91 %
誤検出率	約 0 %

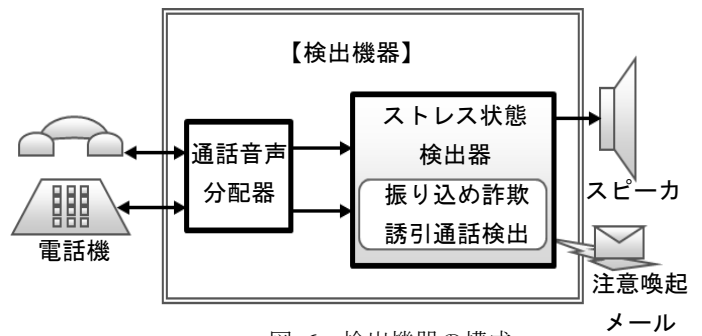


図6 検出機器の構成

Figure 6 Configuration of Detection Device.

4.2 実証実験

開発したストレス状態検出技術を振り込め詐欺誘引通話検出に適用し、一般家庭を対象とした実験を行った。

図6に、モニター宅に設置した検出機器の構成を示す。電話機の本体とハンドセットの間に検出機器を設置し、通話音声分配器を用いて会話音声を取り出し、ストレス状態検出器において図5の処理を行う。また、振り込め詐欺誘引通話を検出した場合、モニターの注意を喚起するための音声メッセージをスピーカから出力する。また、同時に、関係者に所要の行動を促すための注意喚起メールを送信する。

実証実験開始後、約2週間の調整期間が必要だったが、その後は、通常通話を振り込め詐欺と誤検出することを抑え、正常に動作することを確認した。

5. ストレス状態検出システム

開発したストレス状態検出技術の実用化を検討するために、端末-センター連携のプロトタイプシステムを開発した。

実証実験の経験から、見守りサービスを実施するにあたり、ユーザーの普段の生活や業務を妨げずに、音声認識処理のキーワードリスト変更やストレス状態か通常状態かの判定しきい値の調整などのソフトウェア更新を行うことが課題であることが分かった。この課題を解決して多数のユーザーを見守るためには、端末をセンターから管理する仕組みが必要である。また、見守りサービスの種類やユー

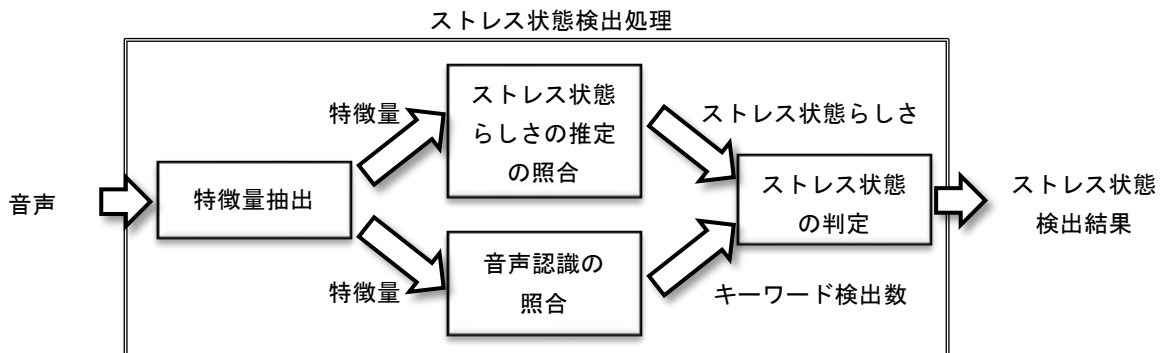


図7 ストレス状態の検出処理

Figure 7 Detection Process of Stress Situation.

ザーの手元にある端末の処理能力に応じ、ストレス状態検出処理を端末とセンターで分担する必要がある。

5.1 ソフトウェアの機能ブロック

様々な構成でストレス状態検出処理を端末とセンターで分担できるように、図 7 に示すようにソフトウェアを機能ブロックに分け、各ブロックの処理と API、およびブロック間でやり取りする情報のフォーマットを定義して実装した。

・特徴量抽出

例えば FFT(Fast Fourier Transform)といった変換処理を用い、音声信号からスペクトルをベースとする特徴量を算出する。ここでは、ストレス状態らしさの推定処理用と音声認識処理用の特徴量をそれぞれ算出する。

・ストレス状態らしさ推定の照合

音声のピッチとレベルを基にし、ストレス状態らしさを推定する。

・音声認識の照合

ワードスポッティング型音声認識により、リストに登録されたキーワードの発声を検出する。

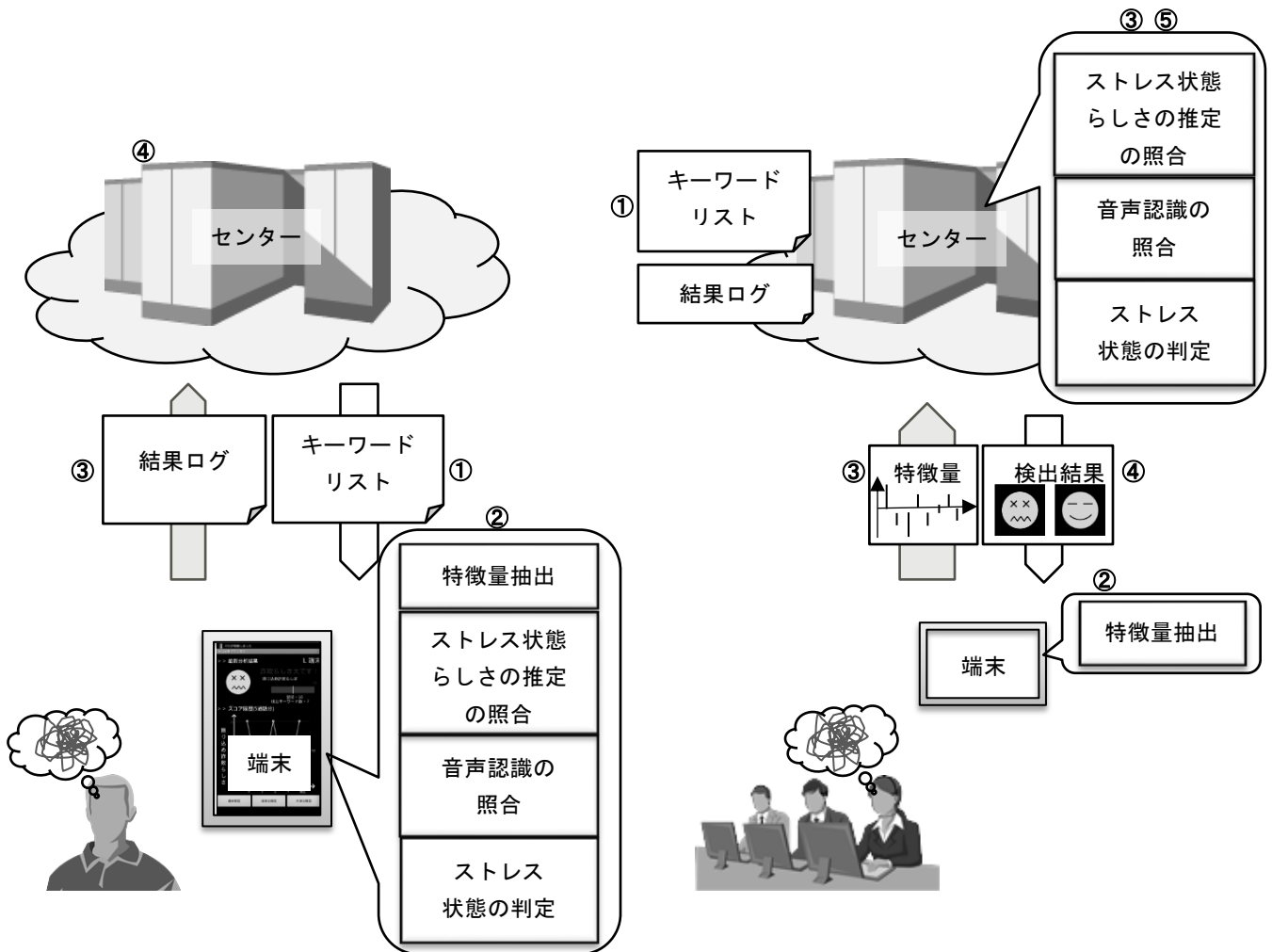
・ストレス状態の判定

ストレス状態らしさの推定結果と、検出したキーワードの発声数を統合し、ストレス状態か通常状態かを判定する。

5.2 プロトタイプシステムの構成

5.1 節で述べたソフトウェアの機能ブロックによって、10 種類のシステム構成が可能である。この中で今回は特徴が分かりやすい例として、図 8 に示すように、(1)ストレス状態検出処理を主に端末で行う構成と、(2)ストレス状態検出処理を端末とセンターで分担する構成のプロトタイプシステムを開発した。

(1)の構成は、例えば家庭向け見守りといった、多数のユーザーの端末からセンターに接続する形態のサービスを実現するのに適している。これは、(2)の構成に比べ、キーワードリストや判定しきい値などの調整結果の反映に時間



(1) ストレス検出処理を主に端末で行う構成

(2) ストレス検出処理を端末とセンターで分担する構成

図 8 ストレス状態検出システムの構成例

Figure 8 Example of Configuration of Stress Situation Detection System.

がかかる場合があるが、端末 1 台あたりの通信量が少なく、また、センターの負荷が軽いためである。

一方、(2)の構成は、例えば従業員の見守りといった業務用端末からセンターに接続する形態のサービスを実現するのに適している。これは、(1)の構成に比べて、端末 1 台あたりの通信量が多く、センターの負荷も重い、音声の特徴量を基にしたセンターでのストレス状態検出状況の分析が可能となるためである。これにより、例えばコールセンターの顧客満足度向上を目指したオペレータ教育への活用や、個別の検出状況に合わせたソフトウェアの調整が可能である。また、(1)の構成に比べて、端末の処理負荷が軽いため、端末の導入コストを抑えられる。

各構成における処理の流れは、次の通りである。

(1) ストレス状態検出処理を主に端末で行う構成

①センターから音声認識で使用するキーワードリストや、ストレス状態検出の判定に用いるしきい値などのパラメーターを端末にダウンロードする。ここで、金銭関連や不祥事などに関するキーワードのリストはサービス内容ごとに設定する。

また、ダウンロードはセンターにアクセスして行う。ここで、複数の端末からのアクセスがセンターに集中しないように、アクセス頻度は 1 回/日程度とし、アクセス時刻は端末ごとにランダムに設定する。

②通話が開始されると端末において、最初に、音声信号の高さと強さから成る特徴量の抽出処理、次に、特徴量からのストレス状態らしさの推測処理と音声認識処理、最後に、ストレス状態の判定処理を行う。ストレス状態と判定した場合は、図 1 の見守りセンターや親族などの冷静な第三者に通知してサポートに繋げる。

③端末で行ったストレス状態らしさの推測結果、音声認識結果、および最終のストレス状態判定結果のログファイルをセンターにアップロードして、ストレス状態検出の記録として残す。

④人手による分析が必要になるが、この記録を基に、出現頻度の低いキーワードの入替えや判定しきい値の変更などの、検出精度向上のための調整を行う。この調整結果は、次のキーワードリストやパラメーターのダウンロード時に、端末のソフトウェアに反映される。

(2) ストレス状態検出処理を端末とセンターで分担する構成

①センターにおいて、音声認識で使用するキーワードリストや、ストレス状態検出の判定に用いるしきい値などのパラメーターを設定する。

②通話が開始されると端末で音声の高さと強さから成る特徴量を抽出する。

③端末で抽出した特徴量をセンターにアップロードし、

センターにおいて、特徴量を基に、ストレス状態らしさの推測処理、音声認識処理、およびストレス状態の判定処理を行う。

④ストレス状態の判定結果を端末にダウンロードする。また、ストレス状態と判定した場合は、(1)の構成と同様に上司や管理者などの冷静な第三者に通知してサポートに繋げる。

⑤センターで行ったストレス状態らしさの推測結果と音声認識結果、および最終のストレス状態判定結果のログファイルを、対応する特徴量と共にストレス状態検出の記録として残す。

また、(1)の構成と同様に、この記録を基に、出現頻度の低いキーワードの入替えや判定しきい値の変更などの、検出精度向上のための調整を行う。なお、(1)の構成と異なり、検出処理の基となる特徴量も記録として残すため、ストレス状態となった状況の分析と、調整結果の効果確認が可能である。また、センターにおける調整結果をすぐに上記③の処理に反映可能である。

5.3 プロトタイプシステムの動作確認

5.2 節で述べたプロトタイプシステムの端末の処理をスマートフォン (CPU のクロック周波数 1GHz, RAM 容量 512M バイト)、また、センターの処理をパソコン上で実装して動作を確認した。

通話の録音データを端末に入力して評価実験を行った結果、(1)の構成と(2)の構成の両方において、シミュレーションと同等のストレス状態検出精度を実現できることを確認した。

また、実際の使用時を想定して、通話しながらストレス状態検出処理を行った場合のスマートフォンの使用リソースを調べた結果は表 2 の通りであった。

表 2 端末 (スマートフォン) の処理リソース

Table 2 Required Resource for Processing on the Terminal(Smart Phone).

	ストレス状態検出処理の構成	
	主に端末で処理	端末とセンターで分担処理
CPU 使用率	約 52%	約 16%
RAM 使用量	約 22M バイト	約 15M バイト

今回の結果から、見守りサービスの種類や端末の処理能力に応じたシステム構成にできることがわかった。

6. おわりに

会話音声からの見守りサービスの実現を目指して、音声信号からのストレス状態検出技術を開発し、更にプロトタイプシステムの動作確認を行った。

また、画像や振動などのほかのセンサーからの情報を統合することで、例えば、高速走行中の車室内といった雑音環境下でも使用可能な技術の開発を進める。更に、今回開発した音声からのストレス状態検出技術を、生活音や機械の動作音などの異常検出に応用し、家庭内や工場内などの見守りに展開していく。

謝辞 本研究の一部は、独立行政法人 科学技術振興機構(JST)の戦略的創造研究推進事業(CREST)の支援により行った。

ストレスが声帯の挙動に影響を与える件について、東京ボイスクリニック院長楠山敏行先生から助言いただいた。

参考文献

[1]シード・プランニング：見守りサービスの市場規模
<http://www.seedplanning.co.jp/press/2011/2011092801.html>

[2]内閣府：独居高齢者の人数予測
http://www8.cao.go.jp/kourei/whitepaper/w-2013/zenbun/pdf/1s2s_1.pdf

[3]警察庁，“警察庁振り込め詐欺対策 HP, 被害発生状況・被害額”，
http://www.npa.go.jp/safetylife/seianki31/1_hurikome.htm

[4]鈴木護：“防災基礎講座 振り込め詐欺の理解と予防に向けて”，
予防時 241 号,日本損害保険協会 pp.8-14,2010

[5]永岑ら：“振り込め詐欺への神経科学からのアプローチ”，
社会技術研究論文集 Vol.6, pp.177-186, 2009

[6]釜野，鷺尾，原田，松尾，“非言語情報を基にした心理的抑圧
検出の検討”，信学技報 SP2010-64, pp.107-110, 2010.

[7]松尾，鷺尾，原田，釜野，早川，武田，“非言語情報を基にした
ストレス状態検出の検討”，信学技報 SP2011-35, pp.29-33, 2011.

[8]PRESS RELEASE 株式会社アニモ，富士通研究所，
“日本初のキーワード型 音声検索ソフトウェア
『VoiceTracking/KeywordFinder』 開発・販売”，
<http://pr.fujitsu.com/jp/news/2009/02/20-1.html>

[9]PRESS RELEASE 国立大学法人名古屋大学，富士通，独立行政
法人 科学技術振興機構，“世界初！過信状態を声の高さと大きさを
基に検出する技術を開発”，
<http://pr.fujitsu.com/jp/news/2012/03/19.html>

[10]PRESS RELEASE 国立大学法人名古屋大学，富士通，“岡山県
で一般家庭を対象にした振り込め詐欺誘引通話検出技術の実証
実験を開始”，
<http://pr.fujitsu.com/jp/news/2012/08/3.html>

[11]松尾，早川，原田，武田，降旗，“非言語情報を基にした過信
状態の検出”，信学技報 SP2012-130, pp.57-62, 2013.

[12]早川，原田，松尾，武田，降旗，“振り込め詐欺誘引通話検出技
術の開発”，電子情報通信学会 2013 年総合大会，情報・システム
講演論文集, vol.1, pp.155, March 2013.

[13]原田，早川，松尾，武田，降旗，“振り込め詐欺誘引通話の検出
システムの実証実験”，情報処理学会第 75 回全国大会講演論文集，
pp.4-511, 4-512, March, 2013.

[14]G. Zhou et al : Nonlinear feature based classification of speech
under stress. IEEE Trans. on Speech and Audio Processing, vol.9, no.3,
p.201-216, 2001

[15]北村達也，“情動による声道形状変化の MRI 観測.”，
第 23 回日本音声学会全国大会予稿集, p.45-50, 2009.

[16]H. Nakagawa et al. : Lubrication mechanism of the larynx during
phonation: an experiment in excised canine larynges. Folia Phoniatr
Logop Vol.50, No.4, p.183-194(1998).

[17]Rachel E. Witt et al. : Effects of surface dehydration on mucosal
wave amplitude and frequency in excised canine larynges. Otolaryngol
Head Neck Surg, Vol.144, No.1, p.108-113(2011).

[18]独立行政法人国民生活センター，見守り情報
(高齢者・障がい者・子どものトラブル防止)，
<http://www.kokusen.go.jp/mimamori/index.html>