

# スマートフォンを用いた音声による健康状態のモニタリングシステムの開発

大宮 康宏<sup>†</sup> 萩原 直樹<sup>†</sup> 桐田 賢<sup>††</sup> 篠原 修二<sup>†††</sup> 中村 光晃<sup>†††</sup>  
光吉 俊二<sup>†††</sup> 徳野 慎一<sup>†††</sup>

<sup>†</sup> PST 株式会社 〒231-0023 横浜市中区山下町 2-905

<sup>††</sup> 株式会社 AGI 〒108-0074 東京都港区高輪 4-10-31

<sup>†††</sup> 東京大学 〒113-8655 東京都文京区本郷 7-3-1

E-mail: <sup>†</sup>{omiya,hagiwara}@medical-pst.com, <sup>††</sup>kirita@agi-web.co.jp,

<sup>†††</sup>{shinokan99,m-nakamura,mitsuyoshi,tokuno}@m.u-tokyo.ac.jp

あらまし 多くの先進国ではメンタルヘルス不調が問題となっており、これに対処するため、抑うつ状態やストレスなどを手軽にチェックできるスクリーニング技術が求められている。従来、自記式心理テストやバイオマーカが使用されてきたが、レポーティングバイアスや検査費用、実施者の負担などの問題がある。また、それらは年に1回程度の定期検診に合わせて実施される事が多いため、その間のケアが難しく、検診時以外の期間についてもモニタリングできる技術が求められている。一方、先行研究として、音声データを用いた抑うつ状態やストレス状態を推定する研究があり、音声を用いた分析は、非侵襲である上、特殊な専用装置を必要とせず、手軽にかつ遠隔的に行えるという利点がある。本稿では、日常的に声を発する電話の通話音声に着目し、スマートフォンを用いた音声による健康状態のモニタリングシステムの開発を目的とする。具体的には、実装する上での課題とその実現方法について述べ、そして実際に開発したシステムの有効性について議論する。

キーワード 音声, 通話, スマートフォン, メンタルヘルス状態の推定, モニタリング

## Development of a monitoring system of mental health state using call voice with a smartphone

Yasuhiro OMIYA<sup>†</sup>, Naoki HAGIWARA<sup>†</sup>, Takashi KIRITA<sup>††</sup>, Shuji SHINOHARA<sup>†††</sup>, Mitsuteru NAKAMURA<sup>†††</sup>, Shunji MITSUYOSHI<sup>†††</sup>, and Shinichi TOKUNO<sup>†††</sup>

<sup>†</sup> PST Corporation, Inc. 2 Yamashita-cho, Naka-ku, Yokohama-shi, 231-0023 Japan

<sup>††</sup> AGI Inc. 4-10-31, Takanawa, Minato-ku, Tokyo, 108-0074 Japan

<sup>†††</sup> Faculty of Engineering, The University of Tokyo 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8655 Japan

E-mail: <sup>†</sup>{omiya,hagiwara}@medical-pst.com, <sup>††</sup>kirita@agi-web.co.jp,

<sup>†††</sup>{shinokan99,m-nakamura,mitsuyoshi,tokuno}@m.u-tokyo.ac.jp

**Abstract** In many advanced nations, mental health disorder is important problem. To deal with this problem, we need the screening methods which can detect the depression and stress. Self-administered tests and biomarkers are used as conventional method. However, these methods have problems such as the reporting bias, inspection costs and high cost burdens on the examinees. And also, these screening methods were used at yearly medical examinations. But, It is difficult to care the period not receiving health diagnosis. So that we need method which can monitor of the mental health status. On the other hand, our previous studies which estimate the depression and stress state using the voice. Voice analysis is non-invasive, it does not require a special dedicated device, easily and remotely. In this paper, we develop the monitoring system of mental health state using call voice with a smartphone. Specifically, it describes the challenges and the realization methods, and actually discuss the effectiveness of the developed system.

**Key words** voice, call, smart phone, mental health care, monitoring,

## 1. ま え が き

多くの先進国ではメンタルヘルス不調が問題となっており、治療費や仕事への支障など、失われる経済的なコストは計り知れない [1]。この問題に対処するため、抑うつ状態やストレスなどを手軽にチェックできるスクリーニング技術が求められている。

現在、使用されているメンタルヘルス不全患者のスクリーニング手法として、例えば General Health Questionnaire (GHQ) [2] や Beck Depression Inventory (BDI) [3] などの自記式心理テストがある。その他にも唾液 [4] や血液 [5] などのバイオマーカーを用いたメンタルヘルス状態の評価手法が提案されている。しかしながら、自記式心理テストは早期発見や診断補助に有効であるものの、レポーティングバイアスの問題がある。また、バイオマーカーによる評価では、検査費用や実施者の負担などの問題があり、手軽であるとは言い難い。

さらに、従来の自記式心理テストやバイオマーカーを利用したスクリーニングは、通常、年 1 回の定期検診に合わせて実施される事が多く、その間のケアが難しい。精神疾患の有病率に関する世界精神保健調査 (WMH) [6] における日本での調査結果 [7] によれば、過去 12ヶ月間に何らかの気分障害を経験した人が 9%、大うつ病を経験した人だけでも 2.9% みられたと報告されている。この調査結果は、年 1 回の定期検診時のスクリーニングだけでは不十分であり、検診時以外の期間についてもケアが重要であることを示唆している。

一方、筆者らは先行研究として、音声データを用いた抑うつ状態やストレス状態などの健康状態を推定する研究を進めている [8] [9]。音声を用いた分析は、非侵襲である上、特殊な専用装置を必要とせず、手軽にかつ遠隔的に行えるという利点がある。この技術を活用することにより、日常的に音声から健康状態を推定する仕組みを作ることができれば、従来は困難であった日常的なモニタリングを実現できる可能性がある。

そこで本論文では、日常的に声を発する電話の通話音声に着目し、スマートフォンを用いた音声による健康状態のモニタリングシステムの開発を目的とする。

近年、スマートフォンの高機能化・高性能化が進んでおり、そのスマートフォンを実装対象とすることによって、音声入力から解析、結果の出力までの首尾一貫したシステムとして実現することが可能になると考えた。提案システムでは、まずはじめにスマートフォン上で音声を録音する。次にその音声から健康状態を推定するアルゴリズムを用いて解析する。そして、解析結果を蓄積することによって長期的な健康状態のモニタリングを実現する。

本研究においてスマートフォンを対象としたことによる特長として、まず、通話機能を持つスマートフォンに備わっているマイクロフォンを音声入力デバイスとして利用できる。これは、録音装置を別途用意する手間やコストが掛からないという利点がある。次に、録音した音声から健康状態を推定する演算を行うが、この処理もスマートフォン内に実装できる。その結果、録音と演算を同じデバイス上で実現されるため、録音した音声

ファイルを別の演算処理装置に転送する手間やコストが掛からない。さらに、スマートフォンに備わるディスプレイを用いて、ユーザに対して解析結果を提示することができる。

以降、2 章ではスマートフォンを用いた音声による健康状態のモニタリングシステム開発の準備事項として、目的を達成するための課題およびそれを踏まえた要求仕様を述べ、さらにその実現方法及びシステムの構成要素について述べる。3 章では、その要求を実現する具体的なシステムの設計と実装について述べる。4 章では、2 章で述べられた要求仕様に基づく提案システムの有効性を検討し、最後に 5 章で本論文をまとめる。

## 2. 準備事項

本章では、本研究で目的とするスマートフォンを用いた音声による健康状態のモニタリングシステムの実現のために要請される要求仕様およびその実現方法について述べ、さらに、その具体的なシステムの構成要素について述べる。

### 2.1 要求仕様

#### 2.1.1 音声の録音

提案システムにおいて、音声を入力として利用者の健康状態として抑うつ状態やストレス状態などの健康状態を推定する。そこで用いる入力音声の取得には、音声を電気信号に変換するためのマイクロフォンと、その電気信号を音声データとして保存する録音機能が必要である。

#### 2.1.2 音声を用いた健康状態の推定

録音された音声データから健康状態を推定する演算処理を行う。そのためには、まず、音声データの数値列を取り込み、次にその数値列から健康状態を推定する演算を行い、そして結果を出力する、という一連の処理を行える必要がある。

#### 2.1.3 結果の蓄積

音声を用いた健康状態の推定結果を蓄積する。1 回の測定で得られる結果は、実施したタイミングの健康状態のみである。そのため、日常的なモニタリングを実現するためには、複数回にわたり継続的に健康状態を推定し、その推移を蓄積できる必要がある。

### 2.2 実現方法

#### 2.2.1 音声の録音

従来、録音には IC レコーダーやパソコン、専用機など、小型のものから大掛かりなものまで様々な機器が使用されてきた。他方、スマートフォンには録音機能が備わっており、必要な場合には外部マイクの使用や高音質録音への対応も可能であり、録音装置として問題なく使用することができる。録音装置にスマートフォンを採用することによる大きなメリットの一つとして携帯性がある。それは、スマートフォンが持ち運びが容易なコンパクトサイズであり、電話機能を使用するという性質上、常に身近に所持している事が多いということである。これにより、音声を用いた健康状態のモニタリングのために、意識して別の録音装置を持ち歩く必要も無くなる。

#### 2.2.2 音声を用いた健康状態の推定

通常、演算処理をパソコン上に実装することが多いが、スマートフォンはいわゆる小型の携帯可能なパソコンのようなも

のであり、このスマートフォン上に音声から健康状態を推定する演算処理を実装することが可能である。ここで、スマートフォン上に録音と演算処理の両機能を実装することによって、別の録音装置を用いた場合に必要となる録音された音声データを演算装置に転送するコストも削減される。

なお、音声を用いた健康状態の推定アルゴリズムには、参考文献 [9] の手法を採用することとした。

### 2.2.3 結果の蓄積

音声を用いた健康状態の推定処理が行われる度に結果をスマートフォン内部のストレージに保存・蓄積する。これにより、1 回限りの結果だけでなく長期的な結果を蓄積し、それらを見比べることによって平常時との差異や長期的な傾向を把握することが可能となる。

## 2.3 構成要素

### 2.3.1 スマートフォン OS

現在使用されているスマートフォンは OS の種類により分類でき、主なものには Android, iOS, Windows Phone がある。要求仕様から使用するスマートフォンについては、通話音声を取得（録音）できる機能が OS に備わっている必要がある。各スマートフォン OS におけるキャリア通話録音の対応状況は表 1 に示す通りである。

表 1 キャリア通話録音の対応状況

スマートフォン OS	キャリア通話の録音
Android	可
iOS	不可
Windows Phone	一部機種のみ可

ここで iOS はキャリアを介した通話音声の録音・取得は不可となっており、対象から外れる。この機能制限は、アメリカの多くの州や諸外国において、通話録音には相手の同意が必要であることに配慮された結果であると思われる。また、Windows Phone では一部の機種において録音を実現しているアプリがあるものの、ほとんどの機種では不可能である。それらに対して Android では通話時の音声取得が可能となっている。従って、アルゴリズムの実装対象とするスマートフォンの OS については、Android を選択する。なお、Android はスマートフォンにおける世界シェアは 2014 年時点で 81.5 % [10] を占めており、利用者数の点でも有利である。

### 2.3.2 Android のバージョン

Android は速いペースでアップデートを繰り返しており、多くのバージョンが存在している。また、端末に搭載するバージョンの選択やアップデートの提供は、端末メーカーやキャリアに委ねられているため、実際には多くのバージョンが使用されている。提案システムでは、音声の分析処理のためにある程度の CPU パワーが必要となり、比較的新しい機種を対象とするため、バージョンについては Android 4.1(Jelly Bean) 以降に対応することとする。

### 2.3.3 録音開始のタイミング

録音開始のタイミングには、ユーザが呼び出す任意のタイミングでの録音（以降、手動録音と呼ぶ）と通話の開始を検知し

て行う録音（以降、自動録音と呼ぶ）の二つの録音方式が考えられる。日常的にモニタリングを行うためには、録音する事を意識する必要のない自動録音が適していると予想されるが、電話の使用頻度には個人差があるため、電話を掛けない場合の代替手段として、手動録音機能も併せて実装する。ここで、手動録音と自動録音のそれぞれの有効性の比較検討については、別の機会に譲ることにしたい。なお、自動録音における録音の開始は通話の開始を検知して行い、通話は電話を受ける場合と架ける場合があるが、その両方を録音対象とし、録音開始のタイミングは、電話を架けるために受話器を取った時と電話を受けて受話器を取った時のイベントとする。

### 2.3.4 録音する音源

音声の録音には Android の標準クラスである MediaRecorder の AudioRecord を利用し、使用者本人の音声を取得する。AudioRecord では録音する音源を選択することができ、マイクからの音声を直接取得する以外にも、電話の uplink（送信）や downlink（受信）等が選択できる。しかしながら、uplink や downlink は取得できない機種があったり、音質の問題があるため、本システムではマイク音源を選択する。将来的には、downlink を取得して通話相手の音声を取得・解析する等の応用も考えられる。

### 2.3.5 アルゴリズム実装の方式

音声を用いた健康状態の推定アルゴリズムを実装するにあたっては計算実行の速度が重要である。Android の開発では Java の使用が一般的であるが、Java で実装した場合は処理速度は比較的遅くなってしまふ。そのため、音声を用いた健康状態の推定エンジンのアルゴリズム部分を C++ により実装し、Native Development Kit(NDK) を用いてネイティブコードを生成し、Java Native Interface(JNI) の仕組みで Java から呼び出す方式とする。また、この方式では java クラスファイルの解析によるアルゴリズム漏洩のリスクも軽減できる。

## 3. システムの設計と実装

本章では、スマートフォンを用いた音声による健康状態のモニタリングシステムの設計と実装について以下に述べる。

### 3.1 システムの概要

実装する処理の大まかな流れは、次の通りである。

1. 音声の録音
2. 音声による健康状態の分析
3. 分析結果の蓄積
4. 分析済音声の削除
5. 蓄積された分析結果を利用者に提示

以下にそれぞれの処理について述べる。

#### 3.1.1 音声の録音

通話開始や、手動による録音開始を検知して音声の録音処理を開始する。録音処理ではマイクロフォンを音源として音声を取得することで本人の声のみを録音する。この録音から次の分析までの処理はバックグラウンドで動作するよう、Android 上のサービスの仕組みを利用して実装した。

### 3.1.2 音声による健康状態の分析

音声録音の終了後すぐに録音した音声健康状態の推定エンジンに渡し、健康状態の分析を実行する。音声を用いた健康状態の推定エンジンでは、結果は csv ファイルに追記する。また、過去の出力結果を入力としても使用する。csv ファイルは結果やデータが時系列に並んだ形式で保存しており、保存期間に制限などは設けていない。

### 3.1.3 分析結果の蓄積

音声病態分析エンジンによって出力された分析結果の csv ファイルの内容をサーバに設置したデータベースアクセス php スクリプトへ送信し、データベースに登録する。データベースによる結果の蓄積は、担当医や産業医らへの分析結果の提供など他のアプリケーションでの利用を想定しているが、今回は具体的な実装は行っていない。

### 3.1.4 分析済音声の削除

録音した音声には個人情報が含まれている可能性が高いため、分析が終了した音声ファイルはすぐに削除する。

### 3.1.5 蓄積された分析結果を利用者に提示

本研究において実装した健康状態の推定エンジンによる解析結果は、1 回の通話や録音音声からその時点の心の元気さを「元気圧」として、長期にわたって蓄積された複数の通話音声や録音音声から長期的な心の元気さを「心の活量値」として出力される。それらの各結果表示画面（元気圧、元気圧履歴、心の活量値、心の活量値履歴）において、csv ファイルから読み出したデータに従い画面表示を行う。具体的には、元気圧と心の活量値の結果表示に加えて、元気圧履歴および心の活量値履歴では、それぞれの過去の変化を履歴として表示する。

## 3.2 その他

### 3.2.1 データ格納場所

分析結果等の格納場所については、他のアプリや PC からアクセスが可能なパブリックな場所およびアプリのプライベートな場所が考えられる。生データを見せたくない場合等にはプライベートな場所にする実装とすべきである。しかしながら、本研究では使用する際の利便性を優先し、PC やファイル管理アプリからアクセスが可能なパブリックな場所を採用する。録音中の音声ファイルや分析結果ファイルはすべてその配下に保存する。ただし、設定項目などは SharedPreferences に保存する。

### 3.2.2 データベースの構成

データベースサーバは MySQL を使用し、Web サーバからのアクセスのみ受け付けるようにする。アプリは Web サーバの php スクリプトにリクエストを送り、php スクリプトはデータベースサーバにアクセスして登録・取得を行う。この際、php スクリプトでは SHA-1 を使用したチェックサムによる整合性チェックを行い不正な登録を防いでいる。サーバは AWS を利用して構築し、負分散などを行っているが詳細は割愛する。

## 4. 検 討

本章では、2.1 節で示した要求仕様に対して、提案システムとして実現した成果について説明する。なお、本システムを応用したアプリケーションを MIMOSYS (Mind Monitoring

Systems) として提供しており、このシステムを実際に利用し、研究協力者に一定期間使用していただき、音声から健康状態を計測し自動的にモニタリングするアプリケーションの医学的妥当性を検証することを目的とした社会実装研究（研究実施期間：2015/7/20～2016/8/31）が東京大学において実施されている [11]。

### 4.1 音声の録音

音声の録音については、スマートフォンに内蔵されているマイクロフォンから入力される音声を手動録音および自動録音として実現した。ここで、録音ファイルフォーマットは、WAV 形式（量子化ビット数：16bit、サンプリング周波数：11,025Hz、モノラル）とした。

手動録音では、Android 端末のディスプレイのタップをトリガーとして、録音の開始と停止を行う。実装した手動録音機能における、録音開始の待機画面および録音停止の待機画面（録音中）の画面キャプチャをそれぞれ図 1、図 2 に示す。



図 1 録音開始の待機画面



図 2 録音停止の待機画面

自動録音では、電話を架けるために受話器を取った時と電話を受けて受話器を取った時のイベントをトリガーとして録音を開始し、電話を切るタイミングで録音を終了する。なお、自動録音機能においては、ユーザが意識して行う操作が無いため、自動録音に係る画面は不要である。

### 4.2 音声を用いた健康状態の推定

録音された WAV 形式の音声ファイルを入力として、音声を用いた健康状態の推定を行う。この処理を開始するタイミングは、手動録音または自動録音が終了した直後とし、解析開始時および解析終了時に Toast 機能を用いて Android 端末のディスプレイ上にメッセージを表示することにより通知することとした。通知時の画面キャプチャを図 3、図 4 に示す。

また、実装したアルゴリズムの妥当性を評価するため、パソコン上での解析結果とそのアルゴリズムを JNI の仕組みを用いて Android 上で呼び出して解析を行った結果が一致することを確認した。

### 4.3 結果の蓄積

音声を用いた健康状態の推定エンジンによる解析結果は csv ファイルに追記される。本研究において採用した健康状態の推定エンジンによる解析結果は、1 回の通話や録音音声からその時点の心の元気さを「元気圧」として、長期にわたって蓄積された複数の通話音声や録音音声から長期的な心の元気さを「心



図3 解析開始時の Toast 表示



図4 解析終了時の Toast 表示

の活量値」として出力される。それら各解析結果の表示例を図5, 図6に, またそれらの履歴表示例を図7, 図8に示す。



図5 元気圧の表示



図6 活量値の表示



図7 元気圧の履歴表示



図8 活量値の履歴表示

この様に履歴として表示することにより, メンタルヘルス状態として音声による健康状態の長期的な傾向や, 解析頻度などをモニタリングすることが可能となった。

## 5. まとめ

本研究では, スマートフォンを用いた音声による健康状態のモニタリングシステムを開発した。音声では, 非侵襲である上, 特殊な専用装置を必要とせず, 手軽にかつ遠隔的に行えるという利点があり, 日常的に声を発する電話の通話音声に着目し, スマートフォンに実装した。さらに, スマートフォン上で音声入力から解析, 結果の出力までの首尾一貫したシステムとして実現しており, これにより, 使用者のメンタルヘルス状態を日常的に, 簡便に計測できるようになった。

今後の課題としては, Android 以外のスマートフォン OS への対応, 音声を用いた健康状態の推定アルゴリズム自体の性能向上, 産業界など然るべき管理の必要な第三者による健康状態の把握などが挙げられる。

## 文 献

- [1] R. C. Kessler, H. S. Akiskal, M. Ames, H. Birnbaum, P. Greenberg, R. M. A. Hirschfeld, R. Jin, K. R. Merikangas, G. E. Simon, P. S. Wang, Prevalence and effects of mood disorders on work performance in a nationally representative sample of U.S. workers, *Am. J. Psychiatry* 163 (2006) 1561-1568.
- [2] D. P. Goldberg, *Manual of the General Health Questionnaire*, NFER Publishing, Windsor, England, 1978.
- [3] A. T. Beck, C. H. Ward, M. Mendelson, J. Mock, J. Erbaugh, *An Inventory for Measuring Depression*, *Arch. Gen. Psychiatry*, 4 (1961) 561-571.
- [4] Shuhei Izawa, Nagisa Sugayab, Kentaro Shirotsubi, Kosuke Chris Yamadac, Namiko Ogawab, Yuko Ouchib, Yuichiro Naganod, Katsuhiko Suzukie, Shinobu Nomura, Salivary dehydroepiandrosterone secretion in response to acute psychosocial stress and its correlations with biological and psychological changes, *Biological Psychology* Volume79, Issue 3, (2008) 294-298.
- [5] Atsuo Sekiyama, Interleukin-18 is involved in Alteration of Hypothalamic-pituitary-adrenal axis activity by Stress, *Society of Biological Psychiatry Annual Meeting.*, San Diego, USA, 2007.
- [6] "The World Mental Health Survey Initiative". <http://www.hcp.med.harvard.edu/wmh/>, (参照 2015-10-01)
- [7] Kawakami N, Takeshima T, Ono Y, Uda H, Hata Y, Nakane Y, Nakane H, Iwata N, Furukawa TA, Kikkawa T. Twelve-month prevalence, severity, and treatment of common mental disorders in communities in Japan: preliminary finding from the World Mental Health Japan Survey 2002-2003. *Psychiatry Clin Neurosci.* 2005;59(4):441-52.
- [8] S. Tokuno, S. Mitsuyoshi, G. Suzuki, G. Tsumatori, *STRESS EVALUATION BY VOICE: a novel stress evaluation technology*, 9th International Conference on Early Psychosis (Tokyo), 2014.11.17-19
- [9] Shuji Shinohara, Shunji Mitsuyoshi, Mitsuteru Nakamura, Yasuhiro Omiya, Gentaro Tsumatori, and Shinichi Tokuno, *Validity of a voice-based evaluation method for effectiveness of behavioural therapy*, *MINDCARE 2015*(Milan), 2015.9.23-25
- [10] IDC Worldwide Quarterly Mobile Phone Tracker, February 24, 2015
- [11] "社会実装研究 — 音声病態分析学講座". 東京大学大学院医学系研究科 音声病態分析学. <http://univ.tokyo/mimosys/trial/>, (参照 2015-10-01)