

Headphone Taps : 通常のヘッドホンへのタップ入力

真鍋 宏幸^{1,a)} 福本 雅朗^{1,†1,b)}

受付日 2013年6月21日, 採録日 2014年1月8日

概要: 簡便な回路でヘッドホンへのタップを検出する Headphone Taps を提案する. 提案手法を用いることで, マイクやスイッチなどの入力手段を備えていない通常のヘッドホンを, 入出力デバイスとして利用することが可能となる. 様々なヘッドホンでタップ入力を検出できること, 歩行中でも入力が可能であることを確認した. 一方のスピーカをマイクとして利用した音声入力を組み合わせることで, 複雑なコマンドを素早く入力することもできる.

キーワード: ヘッドホン, タップ, コマンド, 音声, 入力

Headphone Taps: Tap Control for Regular Headphones

HIROYUKI MANABE^{1,a)} MASAOKI FUKUMOTO^{1,†1,b)}

Received: June 21, 2013, Accepted: January 8, 2014

Abstract: We propose Headphone Taps, which detects a tap on the headphone by a simple circuit. This technique allows regular headphones that have no input functionality to be used as input-and-output devices. Tests confirm that taps on various headphones are detected and the user can input commands while walking. It is easily combined with voice commands, possible if the speaker is used as a microphone, and various commands including complicated ones can be supported.

Keywords: headphones, tap, command, voice, input

1. はじめに

携帯型音楽プレーヤの普及により, いつでもどこでも音楽を楽しめるようになった. 通勤・通学中, また家事を行いながら, ジョギングをしながら, ヘッドホンを装着して音楽を楽しむことは, もはや日常的な光景となっている. 音楽プレーヤの種類以上に, 多種多様なヘッドホンが市販されており, ユーザは好みのヘッドホンを選ぶことができる. たとえば, 大型のオーバヘッド型, イントラコンカ型, 外耳道に挿入するインイヤー型など, 大きさや形状による違いの他, ダイナミック型, バランスドアーマチュア型などアクチュエータによる違い, さらにマイクの有無や, 有

線なのか無線なのかという接続方法の違いなどにより, 音質や装着感の異なる様々なヘッドホンがある. また, ヘッドホンは日常生活と密接に結び付いているため, 機能的要素だけでなく, 見た目などのファッション的な要素も重要になってきている. ユーザは市販されている多くのヘッドホンの中から, 総合的な比較を行い, 好みのヘッドホンを選定している.

ヘッドホンはスピーカを内蔵した単純な出力デバイスであることが多く, 音楽プレーヤを操作する場合には, スイッチなどの別な入力デバイスが必要であった. また, 高機能な音楽プレーヤや, 携帯電話を音楽プレーヤとして利用する場合などでは, 音声コマンドや通話などを行うためにマイクが必要となることもある. それらのために, スイッチやマイクなどの入力機能付きのヘッドホンの利用が考えられるが, ユーザによる自由な選定に制限がかかってしまう.

ここで, 通常のヘッドホンに入力機能を加えることができれば, ユーザはヘッドホンを自由に選定することがで

¹ 株式会社 NTT ドコモ先進技術研究所
Research Laboratories, NTT DOCOMO, INC., Yokosuka,
Kanagawa 239-8536, Japan

^{†1} 現在, Microsoft Research
Presently with Microsoft Research

a) manabehiroyuki@acm.org

b) fukumoto@acm.org

き、好みのヘッドホンを使い続けることができる。本稿では簡単な回路を付与することで、通常のヘッドホン筐体へのタップを検出し、入力として利用する Headphone Taps を提案する。

2. 従来研究

音楽プレーヤ操作における利便性向上という観点に加え、日常的に利用されているヘッドホンを、入出力を備えたウェアラブルデバイスとして活用しようとする観点からも、ヘッドホンに入力機能を付与する試みが数多く行われてきた。ヘッドホン筐体や接続ケーブルにセンサ類を組み込むことは、数多く行われてきており、すでに多数市販されている。有線接続方式のヘッドホンでは、接続ケーブルの途中にスイッチとマイクが組み込まれることが多いのに対し、無線接続方式ではヘッドホン筐体に組み込まれる。それらは、イヤホンマイク、ヘッドセットと呼ばれることも多い。イヤホンマイクがユーザの音声を胸元に配置されるマイクで検出するのに対し、外耳道を経由して気導音として伝わる音声を、イヤホン内部に設置したマイクで検出する製品もある [1]。また、ヘッドホンにタッチセンサを組み込むことも試みられている。タッチセンサでヘッドホンの脱着を検出する製品 [2]、タッチパッドを搭載して指先のジェスチャを検出する製品 [3] のほか、操作時の不快感低減を目的としてタッチセンサを用いた研究 [4], [5] などがある。また、フォトリフレクタなどの光学センサを組み込んだ例もある。たとえば、表情の変化にともなう外耳道の変形検出 [6]、左右どちらの耳に装着されているのかの検出 [7]、光電脈波の計測 [8] など、様々な応用がなされている。ヘッドホンを用いて、脈波以外の生体信号を取得する研究もある。たとえば、電極を設置して EMG の計測 [9] や EOG の計測 [10] が行われている。さらに、加速度センサやジャイロセンサにより検出した頭部の動きを用いたインタラクションの研究も行われている [11]。このように、スイッチやセンサを新たに設置することで、多様な入力手段を実装することができる。しかし、それらのヘッドホンを利用する場合、ユーザはヘッドホンを取り替える必要があり、好みのヘッドホンを使い続けることができない。

入力機能のない通常のヘッドホンを使い続けようとした場合、(1) ヘッドホンを用いた入力をあきらめ、音楽プレーヤ本体を直接操作する、(2) 入力機能付きのデバイスを新たに導入する、のいずれかを選択しなければならない。しかし、音楽プレーヤは鞆やポケットの中に収納されていることも多く、素早い操作が困難である。腕や胸元など即座にアクセス可能な位置に配置されていたとしても、デバイスの位置を確認したうえで、正しく入力が行えるよう注意深く操作する必要がある。新たな入力デバイスを導入する場合であっても、同様である。たとえば、有線接続のヘッドホンと音楽プレーヤの間に、スイッチとマイクを備えた

アダプタを介在させた場合、操作を行うためにはまずそのアダプタを探さなければならない。特にジョギング中などでは、アダプタが固定化されずに動き回ってしまうため、素早い操作は難しい。また、入力デバイスは小型であることが望ましいが、小型であればあるほど、素早いアクセスが難しくなり、さらに慎重な入力操作が求められてしまうという課題もある。

新たにスイッチやセンサを追加することなく、通常のヘッドホンに入力機能を付与することができれば、上記の課題を解決することができる。その実現方法として、スピーカをマイクとして利用することがあげられる。スピーカをマイクとして利用できることは広く知られており、一部のインターフォンは、内部に設置されたスピーカのみで、半二重の通話を実現させている。通話方向の切替えは、ユーザの能動的なスイッチ押下によって行われる。また、ヘッドホンには左右に2つのスピーカがあることを利用し、片側のスピーカをマイク、他方のスピーカはそのままスピーカとして用いることで、音声通話が行える製品もある [12]。さらに、エコーキャンセラを用いることで、1つのスピーカのみで全二重通話ができる製品もある [13]。それらの主たる目的は音声通話であるが、音声認識を用いたコマンド入力に適用することが可能である。しかし、実用的な実装を考えると、音声入力を開始するためには、ユーザが音声以外の手段を用いて操作を行う必要があり、先に述べた課題が再度浮上してきてしまう。また、エコーキャンセラを用いた全二重通話を行う場合、複雑な処理が必要のために、コストが上昇し、消費電力も増加してしまう。

スピーカをマイクとして利用する従来技術では、音声の検出が試みられてきたのに対し、本研究ではヘッドホンをタップしたときに生じる衝撃を検出対象とし、それを用いた入力を試みる。タップ入力は、タッチパネルを搭載したスマートフォンにおける代表的な入力手法であり、すでに広く用いられている。マイクを用いたタップ入力も多く研究されてきており、たとえば、窓に設置した4つのマイクによるタップ検出 [14]、腕に巻いたセンサアレイによるタップ検出 [15] などがある。これらでは、タップ位置を検出することで、位置に応じた多様な入力を可能としている。また、マイクを用いることでタップの種類を識別し、入力の種類を増やすことも行われている [16]。ヘッドホンへのタップを考えた場合、ユーザが左右どちらのヘッドホンにタップするのかを使い分けるのは容易であるのに対し、目視が困難なためタップする位置や種類を正確に制御するのは難しい。しかし、左右2種類のタップしか検出できない場合でも、複数のタップを組み合わせることで、多様な入力が可能である。たとえば、手首に設置した加速度センサでタップを検出し、タップの時間間隔を組み合わせることで様々な入力が行う [17]、タップのリズムをパスワードに用いる [18]、などが行われてきている。なお、スイッチは

ON と OFF を任意の時間継続させることができるのに対し、タップの場合には生じた時刻情報のみしか得られないため、イヤホンマイクに搭載されているスイッチを用いた入力とは異なる。

3. 提案手法

提案手法である Headphone Taps は、新たなセンサを追加することなく、通常のヘッドホンに入力機能を付与する。スピーカをマイクとして利用可能なことはよく知られており、スピーカを用いた音声検出は様々に行われてきた。それに対し提案手法では、ヘッドホン筐体へのタップを、内部に設置されたスピーカで検出する。音声の場合と比較して、ヘッドホンへのタップはスピーカに大きな加速度が加わるため検出が容易であり、安価で低消費電力な回路で実現できる。音楽プレーヤを操作することを考えた場合には、ただか数種類の入力が行えれば十分であるため、タップのリズムを用いた入力に対応できる。ヘッドホンへのタップは、目視による確認を行う必要がなく、歩行中やジョギング中など身体が揺れている場合であっても、正確にタップすることが可能である。そのため、ヘッドホンの接続ケーブルにあらかじめ備えられたスイッチや、新たに設置するアダプタのスイッチを用いる場合よりも、素早く正確な操作ができるようになる。さらに、一方のスピーカをマイクとして用いる機能を追加することで、音声での入力も可能となる。この拡張を行うことで、より多様で汎用的な入力を行うこともできる。スピーカを、スピーカとして使用するのか、マイクとして使用するのかを、別に設置されているスイッチで切り替える機能を持つデバイスはすでに存在するが [12]、提案手法では機能の切替えをヘッドホンへのタップによって行うことができるため、素早い音声入力が可能となる。

3.1 基本回路

図 1 に提案手法の基本回路を示す。ヘッドホン内部に設置されているスピーカは、ヘッドホンアンプと抵抗 R1 を通して、音楽プレーヤと接続されている。通常の音楽プ

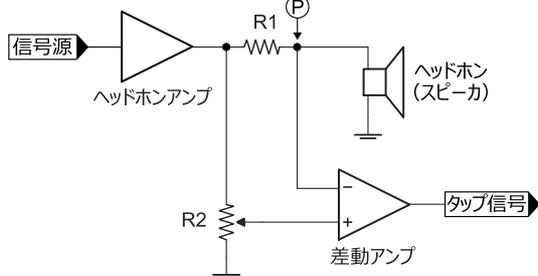


図 1 基本回路. タップにより生じた電流を、R1 の電圧降下量の変化として検出する。R2 はタップ信号が小さくなるように調整される。R1 は電流制限、発信防止のためにも有効である

Fig. 1 Schematic to detect taps.

レーヤでも、ヘッドホンアンプはあり、電流制限や発信防止を目的として R1 が設置されていることも多いため、タップを検出するために、新たに可変抵抗 R2 と差動アンプが追加された構成となっている。なお、この回路は片側のヘッドホンのみを示しており、両側のヘッドホンで独立にタップを検出する場合には、同様の回路をそれぞれ用意する必要がある。

スピーカが音楽を出力しているとき、R1 には音楽信号に応じた電圧降下が生じる。スピーカが純粋な抵抗（リアクタンス成分が 0、抵抗値が一定）である場合には、R1 の電圧降下量は音楽信号に比例するため、差動アンプの出力（以降、タップ信号）がつかねに 0 になるよう、R2 を調整することが可能である。しかし、スピーカには誘導リアクタンスが含まれるため、R1 の電圧降下量は音楽信号には比例せず、位相差が生じる。そのため、R2 をどのように調整しても、タップ信号には信号源由来の成分が表れる（以降では、この成分を残渣成分と表記）。ここで、スピーカがタップされると、スピーカに加えられた衝撃により、パルス的な電流が生じる。R1 の電圧降下量は、この電流の影響を受けて変化するため、タップ信号にその様子が表れる。このタップによる信号変化が、残渣成分よりも十分に大きければ、容易にタップ検出を行うことができる。

市販のイヤホン（図 3 の A）を接続したときの計測例を図 2 に示す。R1 に 100Ω の抵抗を用い、P 点での振幅が 100 mVpp となるように 500 Hz の正弦波を加えた。R2 は残渣成分の振幅が最も小さくなるよう調整してある。タップは、ヘッドホンに装着した状態で行った。タップにより、P 点の波形は乱れる（図 2(i)）。一方、タップ信号は、残渣成分にタップによるピークが加わった波形となっている（図 2(ii)）。タップによるピークに対し、残渣成分は十分小さく抑えられている。

提案手法を広く適用できるかどうかは、ヘッドホンによる違いと、ユーザによる違いが調整可能な範囲にあるかど

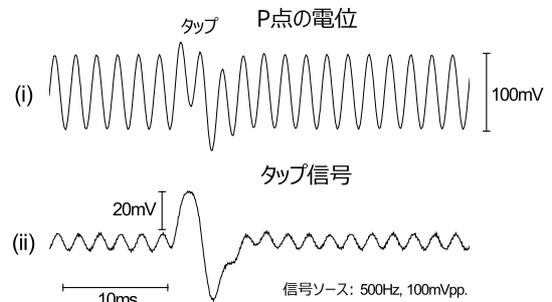


図 2 ヘッドホンにタップしたときのタップ信号。(i) P 点の電位、(ii) タップ信号。タップによって P 点の電位が乱れており、タップ信号では対応するピークが明確に確認できる。P 点とタップ信号に見られる正弦波には位相差が生じており、残渣成分は 0 とはならない

Fig. 2 Tap signal when a headphone is tapped.



	製品名	形状	駆動方式	インピーダンス
A	アップル iPod touch ヘッドホン	イントラコンカ	ダイナミック	32Ω
B	JVC HA-FXC71	カナル	ダイナミック	16Ω
C	シュア SE110	カナル	BA	27Ω
D	オーディオテクニカ ATH-EM700	耳掛け	ダイナミック	32Ω
E	ゼンハイザー HD25-1	オーバーヘッド	ダイナミック	70Ω
F	ゼンハイザー HMD25-1	オーバーヘッド	ダイナミック	600Ω

図 3 使用したヘッドホン. 形状, 駆動方式, インピーダンスが異なる 6 種類を用いた. E と F はインピーダンスが異なる

Fig. 3 Examined headphones.

うかにある. 以下では, それぞれの違いによる影響の調査を行った.

3.2 ヘッドホンによる違い

ヘッドホンには, 駆動方式や形状などの違いにより, 様々な種類がある. 駆動方式には, ダイナミック型, バランスド・アーマチュア (BA) 型, コンデンサ型や圧電型などが知られているが, 多くのヘッドホンはダイナミック型もしくは BA 型である. 形状には, オーバヘッド型, 耳掛け型のほか, 通常はイヤホンと呼ばれる, イントラコンカ型やカナル型などがある. さらに, 構造による開放型, 密閉型の区別があるほか, インピーダンスをはじめとする電気特性もそれぞれのヘッドホンで異なる. それらの違いが, 音質や装着感, 遮音性, 可搬性, ファッション性などに影響し, ヘッドホンに多様性をもたらしている. これら様々なヘッドホンに提案手法が適用できることが望ましい.

そこで, 形状, 駆動方式, インピーダンスの異なる市販ヘッドホンを用意し, 提案手法の適用性について実験を行った. 用意した 6 種類のヘッドホンを図 3 に示す. A は, アップル社の iPod touch に付属していたヘッドホンであり, E と F はインピーダンスが異なる同形状, 同駆動方式のヘッドホンである (マイクが付属しているが, 今回は使用しない). それぞれのヘッドホンを製作した回路に接続し, 先ほどと同様に, 500 Hz の正弦波をヘッドホン両端の電圧が 100 mVpp となるように加え, R2 を調整した. それぞれのヘッドホンをタップしたときのタップ信号を図 4 に示す (図中 A は, 図 2 の (ii) と同一). ピークの振幅や時間幅, 負のピークの有無などに違いは見られるが, すべてのヘッドホンでタップにともなうピークが観測できる. また, 残渣成分はタップにともなうピークと比較して, 十分に小さく抑えられていることが分かる. なお, 図には示していないが, ヘッドホン B ではピークの約 60 ms 後に負の小さなピークが見られた.

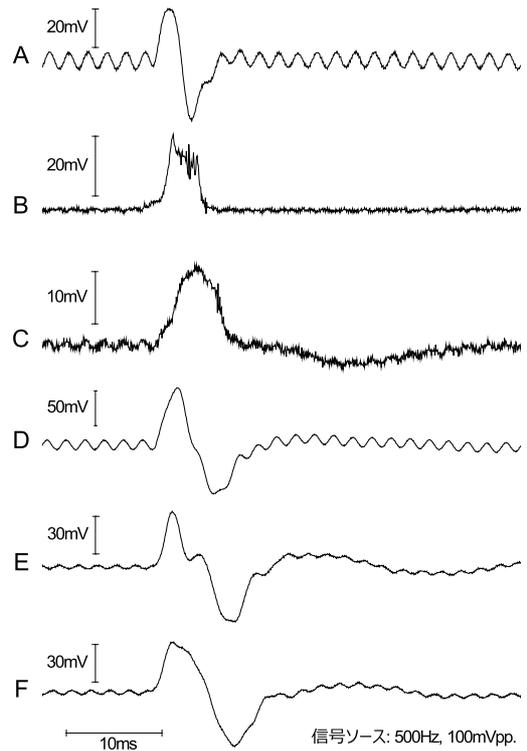


図 4 それぞれのヘッドホンを接続したときのタップ信号. 振幅や形状はそれぞれ異なるが, すべてのヘッドホンで, タップにともなうピークが観測できる

Fig. 4 Tap signals of examined headphones.

3.3 ユーザによる違い

タップの強さや位置によって, タップ信号で観測されるピークの大きさは変わる. ユーザごとに, どの程度ピークが変わるのか実験を行った. 12 名の被験者 (男性 9, 女性 3, 26~46 歳, 平均 33 歳, 全員右利き) に, 図 3 に示した 6 種類のヘッドホンを装着してもらい, 右側のヘッドホンを 10 回タップしてもらった. 被験者には, インジケータに従って, 1 秒ごとに 1 回タップを行うという指示のみを与えた. 観測されたタップ信号から, 正のピークの大きさ (ヘッドホン A や D のように正負 2 つのピークが観測される場合でも, Peak-to-peak の値ではない) を求め, その平均と標準偏差を図 5 に示す. 被験者によって, タップによるピークの大きさが異なっている. ある被験者 1 名はヘッドホン B を装着した場合にタップによるピークが観測できなかったが, 他の被験者ではすべてのヘッドホンでタップによるピークが観測できた. ピークを正しく検出するためには, 残渣成分のうち, ピークと同程度の周波数成分の大きさが特に重要となる. そこで, 一例として 62.5 Hz の正弦波を与えたときの残渣成分の大きさ (peak-to-peak の値) を横棒で示す. 残渣成分よりも, タップによるピークの方が大きければ, 正確なタップ検出が期待できるが, 逆の場合, 残渣成分をタップであると誤検出 (false positive) する可能性がある. ヘッドホン B や E, F では, 300 mVpp の正弦波を加えた場合でも, 多くの被験者でタップ検出が

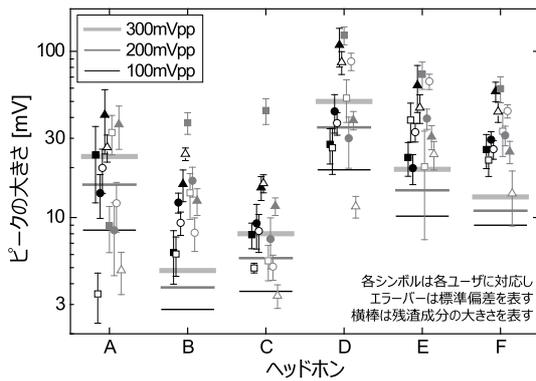


図 5 タップにより生じるピークの高さ。被験者によりピークの高さが変わる。ピークが残渣成分よりも大きいことが多いが、被験者とヘッドホンの組合せによっては、ピークが残渣成分よりも小さい場合がある

Fig. 5 Peak levels of tap signals.

行えるが、それ以外のヘッドホンでは一部の被験者で false positive が生じると考えられた。

4. 実装

提案手法を実装する場合、複数の方法が考えられる。今回は、提案手法の有用性を検証することが目的であるため、製作が容易な、音楽プレーヤとヘッドホンの間に接続する外付けアダプタとして実装を行った。後述するように、ヘッドホンやユーザによる違いを吸収し、false positive を抑制するためのキャリブレーション手法、一方のスピーカをマイクとして用いる通話機能も組み込んだ。

4.1 音楽プレーヤへの実装

Apple 社の iPod touch は音楽プレーヤとして広く普及しているだけでなく、外部コネクタを利用することで、電源の取り出しやコマンドの送信が行えるという特徴があり、実装のベースとして適している。図 1 で示した差動アンプとデジタルポテンシオメータに加え、タップの認識、ポテンシオメータの調整、コマンドの発行を行う PIC マイコンなどから構成される回路基板を、市販のリモコンアダプタのケース内に収納した (図 6 の左)。タップ信号が設定された閾値を超えると、タップであると認識し、対応する LED が点灯する。音楽プレーヤを操作するための主要な 5 つのコマンドに対して、図 7 に示すタップ入力を対応させた。片手モードと両手モードの 2 つの入力モードがあり、片手モードの場合には、文献 [17] と同様に、タップのリズムにより入力が行われる。両手モードでは、左右のヘッドホンへのタップの順番により、コマンドが選ばれる。タップがコマンドとして認識されると、シリアル通信を介して音楽プレーヤにコマンドが送信され、音楽プレーヤが操作される。



図 6 製作したプロトタイプ。iPod touch に接続するアダプタとしての実装 (左) と、Galaxy S II 用のスマートジャケットとしての実装 (右)

Fig. 6 Developed prototypes.

コマンド	片手モード	両手モード
再生 / 停止	♪	シングルタップ (R or L)
次の曲	♪♪	ダブルタップ (R,R)
前の曲	♪♪	ダブルタップ (L,L)
音量アップ	♪♪♪	ダブルタップ (R,L)
音量ダウン	♪♪♪	ダブルタップ (L,R)
オンフック / オフフック	♪	シングルタップ (R or L)

図 7 実装したコマンド。片手モードのときは、各コマンドにはタップのリズムが割り当てられており、両手モードのときには、タップの順番が割り当てられている

Fig. 7 Implemented commands.

4.2 スマートフォンへの実装

Samsung 社のスマートフォン Galaxy S II は、USB ホストとして動作することができるため、外部に電力を供給することが可能である。スマートフォンと USB 接続し、拡張バッテリーとして機能するジャケットからバッテリーを抜き取り、回路基板を搭載したスマートジャケットとして実装した (図 6 の右)。簡略化のために、音楽信号やコマンドの伝達はスマートフォン本体のイヤホンマイク端子とプロトタイプとを有線でつなぐことで行った。タップとコマンドの対応は音楽プレーヤと同様に図 7 のとおりである。音声入力を可能とした点が、音楽プレーヤへの実装と異なる。これは、ヘッドホンの一方のスピーカをマイクとして機能させることで実現している (この切替え機能は、文献 [12] と同様)。この機能により、音声コマンドも実現可能となるが、今回は通話機能のみを実装した。

4.3 キャリブレーション

ヘッドホンの電気的特性は異なっており、タップによるピークの高さも、ユーザによって異なる。それらの違いを吸収するために、それぞれキャリブレーションが必要である。まず、タップ信号に表れる残留成分を最小化するために、図 1 の R2 の調整を行う。音楽の再生を停止したう

えて、特定周波数の基準信号をスピーカに与え、タップ信号をモニタしながら R2 を調整する。

ユーザによってタップによるピークの大きさが異なる点に関しては、2点補正する必要がある。タップとして認識する閾値と、音量の最大値の2点である。ユーザに複数回タップを行ってもらい、タップとして認識する閾値を設定する。しかし、これだけでは残渣成分がその閾値を超えて、false positive が発生することがある。図 5 から分かるように、残渣成分の大きさは音量に依存するため、最大音量を制限することにより、false positive をなくすることができる。ヘッドホンにより残渣成分の大きさは異なるが、R2 を調整するときはその大きさは判明しているため、タップの閾値から適切な最大音量を設定することが可能となる。これら2つのキャリブレーションは、接続するヘッドホンもしくは使用するユーザを変更する際にそのつど必要となるが、10秒程度の短時間で行うことができる。

5. 評価

製作したプロトタイプを用いて、音楽再生中にタップを行ったときのタップ信号の例を図 8 に示す。キャリブレーションにより残渣成分は小さく抑えられており、タップによるピークが明確に確認できる。以下では、製作した音楽プレーヤとスマートフォン用の2種類のプロトタイプを用いて行った、2つの評価実験について示す。

5.1 コマンド入力評価実験

製作した音楽プレーヤ用プロトタイプ、ヘッドホン A と C を用いて、評価実験を行った。6名の被験者(男性5, 女性1, 26~46歳, 平均32歳)は両手モードで自由に入力ができるよう、約10分間の練習を行った。このとき、頭部を振る、ジャンプするなどにより、身体の動きによる false positive が発生するかどうかの確認を行った。その後、静止状態(5名着座, 1名立位)で、5種類(図7の上から5つ)のコマンドをランダムに指示し、計10回ずつ入力を試みた。次に、歩行および階段の昇降中に、同じ5種類のコマンドを計6回ずつ入力した。コマンドに対して、正しく入力できた割合を成功率と定義し、静止時と歩行時それぞれの平均と標準偏差を図9に示す。多くの被験者が、

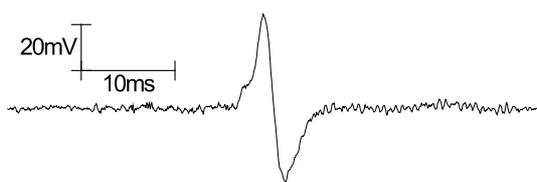


図 8 プロトタイプを適用したときのタップ信号の例。音楽再生中のタップ入力を行っている。残渣成分が小さく抑えられており、タップによるピークが明確に確認できる

Fig. 8 An example of tap signal with the prototype.

静止時だけでなく歩行時でも高い成功率で入力を行うことができている。コマンド別の成功率は、再生 95 (93), 次の曲 91 (79), 前の曲 93 (76), 音量アップ 98 (90), 音量ダウン 93 (94) (全被験者の静止時および歩行時の平均値。括弧内が歩行時)であり、被験者 vi を除いた場合には、コマンド別の成功率に大きな違いは見られなかった。コマンド入力が失敗したのは、被験者 vi を除き、すべて false negative であり、false positive は発生しなかった。たとえば、ダブルタップをシングルタップとして認識してしまったことによる。ヘッドホン A に対して、被験者 i は歩行中にタップするのが難しい、被験者 iii はタップすると外れそうになるとコメントした。ヘッドホン A は、耳からの飛び出しが小さいため、誤った位置をタップしやすく、またユーザの耳形状によってはうまく固定できないことがある。両被験者ともヘッドホン C では高い成功率を達成していることから、それぞれのユーザに適したヘッドホンを選定することが重要であると考えられる。一方、被験者 vi では、false negative も false positive いずれも発生した。被験者 vi はヘッドホンをタップすることにより生じる不快を低減させるために、複数の指で弱くタップしており、ヘッドホンだけでなくその周囲も同時にタップしていた。これにより、ヘッドホンをうまくタップできていない場合には、false negative が発生し、複数の指が時間差をもってタップした場合には false positive が発生した。

キャリブレーションにより最大音量を制限したため、再生している音楽による false positive が発生することはなかった。しかし、一部の被験者は音量が小さいと指摘した。また、ヘッドホンを机などに置くとときや、頭部を強く叩いたときには、false positive が生じることがあった。

5.2 通話実験

スマートフォン用プロトタイプを用いて、通話が可能な

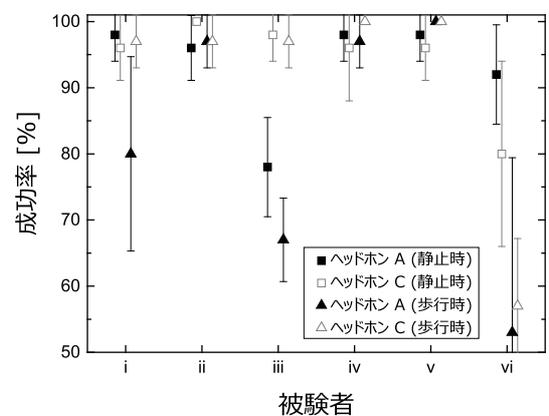


図 9 タップによるコマンドの入力成功率。コマンドごとの成功率の平均と標準偏差を表す。成功率が低下している場合も見受けられるが、多くの場合、歩行時であっても高い成功率が得られている

Fig. 9 Command acceptance rate.

どうか確認を行った。音楽再生中にシングルタップを行った場合には、再生が停止されるが、電話がかかってきたときにシングルタップを行うと、スマートフォンにオフフックコマンドを送信する。同時に、片側のヘッドホンをマイクとして機能するよう切り替える。装着者の発声音は外耳道からも漏れており、ヘッドホンを装着したままで、発声音を検出することができる [1], [12], [13]。本実験においても、ヘッドホンを装着したまま通話を行った。終話する際には、スピーカとして機能している側のヘッドホンをタップすることで、オンフックコマンドの送信と、両方のヘッドホンをスピーカとして機能するよう切り替える。ヘッドホン A を接続し、装着者の声を検出できること、会話が可能であることを確認した。

6. 議論

提案手法により、マイクやスイッチなどの入力手段を有さない通常のヘッドホンに入力機能を加えることができる。そして、複数のヘッドホンをを用いた実験を通じて、提案手法の有効性を確認した。本章では、今後検討すべき点について議論する。

6.1 R1 の抵抗値

ヘッドホンアンプとヘッドホンの間に挿入される抵抗 (図 1 の R1) には、通常 10Ω 程度のものを用いることが多いのに対し、製作したプロトタイプでは 100Ω を用いている。高い抵抗を用いることで、位相差の影響は小さくなる。そのため、残渣成分を小さくすることができ、タップ検出の観点からは有利となる。しかし、ヘッドホンが出力する音量を維持するためには、ヘッドホンアンプはより大きな信号を出力しなければならないため、効率が低下する。また、R1 とヘッドホンの合成インピーダンスが変化するため、出力される音の周波数特性、つまり音質が変化する。音楽を聴くことが第 1 の目的である以上、容易に知覚できるほどに音質が変化することは望ましくない。

ここで、プロトタイプで用いた 100Ω と、通常用いられる 10Ω とで音質の違いを知覚できるかどうか、実験を行った。十分な駆動能力を持つオペアンプに 100Ω を接続した回路と、 10Ω を接続した回路を同一基板上に製作し、ヘッドホン A と iPod touch の間に接続した。それぞれの回路にヘッドホンジャックが設置されており、ヘッドホンを差し替えることで、回路を切り替えることができる。また、 1kHz の正弦波を入力したときのヘッドホンによる電圧降下量が同一となるよう回路上で調整しているため、両回路で音量の違いはない。音楽プレーヤには、MP3 形式で 256kbps のビットレートで記録されたジャズやポップスを含む複数ジャンルの楽曲が収められている。被験者 (男性 9, 女性 1 の合計 10 名, 26~37 歳, 平均 32 歳) に、両回路による音質の違いを知覚できるかどうかを尋ねた。実験

中、被験者は任意の楽曲を、任意の音量で (音楽プレーヤの本体の音量を変えることで制御可能)、任意の時間聴くことができ、任意のタイミングでヘッドホンを差し替えながら、比較を行った。実験の結果、10 名中 7 名は両者の違いが分からないと回答した。残りの 3 名は、わずかに違いを感じると答えたが、その後行ったブラインドテストでの正解率は 7, 7, 5 割 (ランダムに答えた場合の期待値は 5 割) であった。このことから、 100Ω の抵抗を用いた場合であっても、音質への影響は十分に小さく、実用上問題ないこと、また、さらに高い抵抗を利用できる可能性が示唆された。ただし、使用する音楽プレーヤ、アンプ、ヘッドホン、楽曲のジャンルやビットレート、さらに被験者のスキルなどの組合せによって、被験者が感じる音質への影響は異なることが予想されるため、さらなる調査を行い、慎重に抵抗値の設定を行うことが望ましい。

6.2 不快感の低減とタップ検出性能の向上

ヘッドホンを強くタップすると、不快感を生じることが指摘されている [5]。外耳道や周辺部位に衝撃が加わるだけでなく、機密性の高いカナル型のヘッドホンでは、タップによって外耳道内部の気圧が上昇することにより、鼓膜が刺激される。コマンド入力評価実験での被験者 v_i は、特に不快感を感じやすいと考えられる。不快感を抑制するためには、弱いタップでも誤動作なく認識できるようにすることが必要となる。しかし、今回のプロトタイプでは弱いタップを誤動作なく認識するためには、最大音量に大きな制限をかけることになり、実用上問題が生じてしまう。最大音量の制限を緩和しながら、弱いタップを認識できることが望まれる。

現状の実装では、1 つの閾値を用いた単純なピーク検出を用いてタップ認識を行っているために、最大音量を制限する必要があった。より高度な手法を導入すれば、最大音量の制限を緩和しつつ弱い力でのタップにも対応することが可能である。具体的な手法として、残渣成分を抑制する手法と、高度なピーク検出手法の 2 つが考えられる。残渣成分を小さくするためには、R1 をより大きな抵抗とする、簡単な位相補償回路を追加する、タップによるピークの周波数に合わせたバンドパスフィルタをタップ信号に適用することなどがあげられる。また、高度なピーク検出手法として、ピークの形状に合わせたピーク検出アルゴリズムの導入があげられる。図 4 から分かるように、それぞれのヘッドホンで生じるピークの形状は異なる。キャリブレーション時に得られるピークを含むタップ信号と、観測されるタップ信号との類似度を算出することで、より頑強なタップ認識が期待できる。また、多くのヘッドホンでは正のピークの後に負のピークが観測されることから、正負 2 種類のピーク検出を行う簡便な手法も有効であろう。

上記のような技術的手段だけでなく、被験者のトレーニ

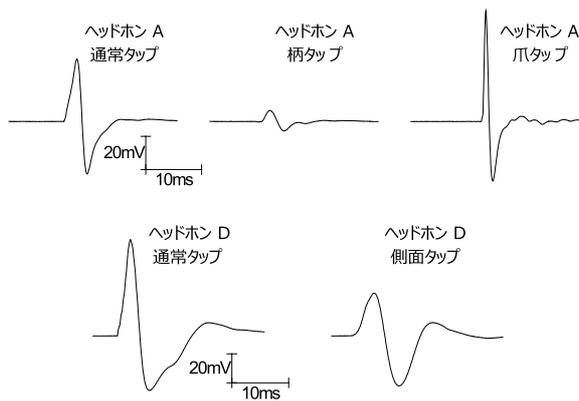


図 10 異なるタップに対するタップ信号. 上段の図はヘッドホン A へ 3 種類のタップを行ったときの, 下段はヘッドホン D へ 2 種類のときのタップ信号の例を表す. タップする位置や, 力のかかる方向によって, タップ信号の大きさが異なる

Fig. 10 Tap signals with different tapping.

ングによって, タップにより生じるピークを大きくすることも有効な手段となりうる. スピーカをタップセンサとして用いる場合, スピーカは 1 軸の加速度センサのように振る舞う. そのため, 同じ力でタップした場合であっても, 力の向きがその軸と一致していない場合や, 指や耳によって衝撃が吸収される場合には, 観測されるタップ信号は小さくなってしまふ. ここで, 異なるタップに対するタップ信号の違いを図 10 に示す. 耳に装着したヘッドホン A に対して, 通常行われる指の腹を用いたヘッドホンのスピーカ収容領域へのタップ, ヘッドホンの柄の部分 (図 3 に示されているヘッドホン A の樹脂製筐体の中で, ケーブルが通っている細い筒状の領域) へのタップ, 爪を用いたヘッドホンのスピーカ収容領域へのタップを行った際のタップ信号を上段に, 下段には装着したヘッドホン D に対する, 通常のタップと, ヘッドホン筐体の側面に, スピーカの軸と直交する向きにタップしたときのタップ信号の例を示す. なお図中, ヘッドホン A への 3 種類のタップ信号, ヘッドホン D への 2 種類のタップ信号は, それぞれ同一スケールで表示している. それぞれのタップを 50 回繰り返したときの, タップ信号の正のピークの大きさの平均値は, ヘッドホン A で 39 (6), 7 (1), 64 (20), ヘッドホン D で 63 (7), 25 (10) であった (単位は mV. 括弧内は標準偏差). タップする位置や方向によって, タップ信号の大きさは異なっており, ユーザがどのようなタップを行うかは, タップ検出性能に大きな影響を与える. 逆にいえば, ユーザが効果的な位置や方向を習得することで, 弱い力でのタップ検出や, 音量の制限を緩和することが可能となる.

6.3 音声コマンドの併用

音楽プレーヤの制御の例など, 入力するコマンドの数が少ない場合, タップを用いた入力是有効に機能する. しかし, 入力可能なコマンド数が多い場合には, 一般にタップ

回数が増大するため, 入力に時間がかかるとともに, 対応するリズムや順番をすべて記憶しておく必要があるため, 実用的ではない.

ヘッドホンを接続する音楽プレーヤやスマートフォンなど携帯端末の多くが, すでに音声コマンド機能を有している. そこで, プロトタイプで実装した音声通話機能を, 音声コマンド入力に転用することで, 任意コマンドの発行が可能となる. タップにより接続機器側の音声コマンド機能を起動させ, ヘッドホンをマイクとして機能させればよい. 複雑な操作, たとえば「特定の曲を再生する」, 「電話帳に登録されている友人に電話をかける」などを, 音声を用いることで短時間に行うことができるようになる. あるコマンドを, タップで入力するのか音声コマンドで入力するのかは, 自由に設定可能であるが, 頻繁に行われるたかだか数種類の操作はタップのみによる入力で行い, 複雑な操作は音声コマンドで入力するのが, 実用的で望ましい設定となるだろう.

6.4 実装形態

今回のプロトタイプは, ヘッドホンと音楽プレーヤなどの機器の間に接続される外部アダプタとして実装を行った. この実装形態は, 製作が容易である反面, 多くの電子部品を必要とし, 利用の前にアダプタを接続しなければならないという煩わしさがある. 他の形態として, 機器本体に内蔵することが考えられる. これにより, 接続の煩わしさが低減され, 利便性が向上する. さらに, 電源やヘッドホンアンプ, CPU など多くの部品を共通化することができるため, 新たに追加する部品が少なくなり, 基板サイズの拡大や, 追加部品によるコスト上昇も小さくなる. また, 外付けアダプタとして実装する場合であっても, 音楽信号やコマンドの伝達を Bluetooth などの無線で行うことで, コマンド実行までの時間短縮, ケーブルの接続の手間の削減などが可能である.

6.5 適用範囲

提案手法が, 様々なヘッドホンに適用できることが確認できた. しかし, ノイズキャンセリングヘッドホンやワイヤレスヘッドホンのように, ヘッドホン側にアンプが搭載されている場合には, 提案手法を適用することができない. また, 本稿で紹介していない様々なヘッドホンを含め, 多くのヘッドホンで試してみたところ, アンプが搭載されていないヘッドホンであっても, 強い力や爪を立てたタップが必要なタップ検出が困難なヘッドホンや, 検出した音声のレベルがきわめて低いヘッドホンが存在することが分かった. いくつかの原因が考えられるが, 製作したプロトタイプに問題がある可能性がある. 製作したプロトタイプは, 図 3 のヘッドホンを参考に設計しており, 図 4 に示した大きさよりも, 極端に小さなピークや残渣成分を持つ

ヘッドホンを想定できていない。単純にタップ信号が小さいだけであれば、プロトタイプ内部の差動アンプのゲインを調整することで、対応することができる。そのためには、より多くのヘッドホンを用いて、それぞれのヘッドホンがどのような反応をし、その範囲はどの程度なのかを明らかにしていくことが必要である。

6.6 ウェアラブルデバイスとしての視点

本稿では、主に音楽プレーヤを操作することを想定してきた。しかし、ヘッドホンがすでに日常的に利用されている状況を考えれば、ヘッドホンを単に音楽を聴くためのデバイスとしてではなく、即時利用可能なウェアラブルデバイスとしてとらえることもできる。ウェアラブルデバイスの開発では、現実的にそのデバイスを日常的に装着することが難しい、という致命的な問題がしばしば生じる（バッテリー容量が原因となることが多い）。提案手法を適用することで、特別なヘッドホンではなく、ユーザの好みのヘッドホンを入出力デバイス化することができるため、その問題を小さくすることができる。これにより、ヘッドホンを用いた所望の情報への素早いアクセス、ユーザのコンテキストに応じた適切なタイミングでの情報提示など、様々なサービスの実現可能性の向上、ウェアラブルコンピューティングのさらなる発展に貢献できると考えられる。

7. 結論

ヘッドホンへのタップを入力とする Headphone Taps を提案した。マイクやスイッチなどの入力手段を有していない通常のヘッドホンであっても、タップを用いた入力機能を付加することができる。ヘッドホンはユーザの耳に固定されているため、目視による確認が不要で素早い入力が可能となる。特性の異なる複数のヘッドホンへのタップを、簡便な回路で検出できることを確認した。プロトタイプとして、市販の音楽プレーヤおよびスマートフォンと、ヘッドホンの間に接続する2種類のアダプタを製作した。評価実験により、静止中だけでなく、歩行中であっても正しくコマンド入力ができることを確認した。また、一方のスピーカをマイクとして利用した音声入力と組み合わせることも容易である。提案手法は、音楽プレーヤやスマートフォンの操作のためだけでなく、ウェアラブルコンピューティングの発展にも寄与すると考えられる。

参考文献

- [1] ナップエンタープライズ：インコア，入手先 (<http://www.incore.jp/products/product.html>) (参照 2013)。
- [2] ソニー・エリクソン：MH907 (released in 2009)。
- [3] Parrot: Zik, available from (<http://www.parrot.com/zik/usa/>) (accessed 2013)。
- [4] Buil, V., Hollemans, G. and van de Wijdeven, S.: Headphones with touch control, *Proc. MobileHCI '05*,

pp.377–378 (2005)。

- [5] Buil, V. and Hollemans, G.: Acceptable Operating Force for Buttons on In-Ear Type Headphones, *Proc. ISWC '05*, pp.186–191 (2005)。
- [6] 谷口和弘, 西川 敦, 小林英津子, 宮崎文夫, 佐久間一郎: みみスイッチ: 外耳の動きを入力情報とする常時着用型入力装置, インタラクシオン 2010 予稿集, pp.243–246 (2010)。
- [7] Matsumura, K., Sakamoto, D., Inami, M. and Igarashi, T.: Universal earphones: Earphones with automatic side and shared use detection, *Proc. IUI '12*, pp.305–306 (2012)。
- [8] Poh, M.-Z., Kim, K., Goessling, A., Swenson, N. and Picard, R.: Heartphones: Sensor Earphones and Mobile Application for Non-obtrusive Health Monitoring, *Proc. ISWC '09*, pp.153–154 (2009)。
- [9] Sano, A., Tomita, T. and Oba, H.: Applications using Earphone with Biosignal Sensors, ヒューマンインタフェース学会研究報告集, Vol.12, No.6, pp.1–6 (2010)。
- [10] 真鍋宏幸, 福本雅朗: ヘッドホンを用いた EOG 法による視線入力インタフェース, 情報処理学会論文誌, Vol.52, No.4, pp.1515–1526 (2011)。
- [11] Brewster, S., Lumsden, J., Bell, M., Hall, M. and Tasker, S.: Multimodal ‘eyes-free’ interaction techniques for wearable devices, *Proc. CHI '03*, pp.473–480 (2003)。
- [12] チェンジャー II, 入手先 (http://www.big-tail.com/product/i_changer.html) (参照 2013)。
- [13] 日鉄エレックス: e 耳くん, 入手先 (<http://www.ns-elex.co.jp/solution/security/factory/factory02.html>) (参照 2013)。
- [14] Paradiso, J.A., Leo, C.K., Checka, N. and Hsiao, K.: Passive acoustic knock tracking for interactive windows, *Extended Abstracts of CHI '02*, pp.732–733 (2002)。
- [15] Harrison, C., Tan, D. and Morris, D.: Skinput: Appropriating the body as an input surface, *Proc. CHI '10*, pp.453–462 (2010)。
- [16] Harrison, C., Schwarz, J. and Hudson, S.E.: TapSense: enhancing finger interaction on touch surfaces, *Proc. UIST '11*, pp.627–636 (2011)。
- [17] 福本雅朗, 外村佳伸: “指釦”: 手首装着型コマンド入力機構, 情報処理学会論文誌, Vol.40, No.2, pp.389–398 (1999)。
- [18] Wobbrock, J.O.: TapSongs: Tapping rhythm-based passwords on a single binary sensor, *Proc. UIST '09*, pp.93–96 (2009)。



真鍋 宏幸 (正会員)

1976 年生。1999 年東京工業大学工学部卒業。2001 年同大学大学院修士課程修了。同年 (株) NTT ドコモ入社。以来、ウェアラブルコンピューティング、生体信号を用いた入力インタフェースの研究に従事。電子情報通信

学会, ACM 各会員。



福本 雅朗 (正会員)

1964年生。1988年電気通信大学応用電子工学科卒業。1990年同大学大学院修士(電子)課程修了。同年日本電信電話(株)入社。以来、各種インタフェースデバイスの研究に従事。

NTT ヒューマンインタフェース研究所，NTT ドコモ先進技術研究所を経て，現在，Microsoft Research (Lead Researcher)。電子情報通信学会，ヒューマンインタフェース学会，ACM 各会員。