論文 / 著書情報
Article / Book Information

論題(和文)	眼電位を用いた音声合成インタフェースの研究
Title(English)	
著者(和文)	尾崎 賢人, 篠崎 隆宏, 武者 利光, 古井 貞熙
Authors(English)	Kento Ozaki, Takahiro Shinozaki, Toshimitsu Musha, SADAOKI FURUI
出典(和文)	日本音響学会2011年春季講演論文集, , No. 3-4-13, pp. 1621-1622
Citation(English)	, , No. 3-4-13, pp. 1621-1622
発行日 / Pub. date	2011, 3

眼電位を用いた音声合成インタフェースの研究*

☆尾崎 賢人，篠崎 隆宏 (東工大)，武者 利光 (脳機能研)，古井 貞熙 (東工大)

1　はじめに

筋萎縮性側索硬化症 (ALS) は，運動ニューロン病の一種であり，重篤な筋萎縮と筋力低下をきたす難病である．ALS の病状末期では，呼吸筋が麻痺するため，延命には人工呼吸器を装着する必要があり，発声が非常に困難になる．このため，運動障害の影響を受けない代替的な意思伝達手段が強く望まれている．これまで，ALS では通常眼球運動障害は現れないことを利用し，閉眼上方視によるスイッチ操作で 50 音の列選択，行選択を順次行う手法 [1] や，画像認識による注視点検出に基づき，ディスプレイに表示されるソフトウェアキーボード上の文字を注視することで入力を行う手法 [2] などが提案されている．しかし，これらの手法は入力毎に一定の待ち時間を要し，即時性の面で問題があった．本研究では，ALS 患者によりインタラクティブ性の高いコミュニケーション手段を提供することを目標としている．入力として即時性に優れる眼電位認識 [3] [4] を用い，音声合成器に繋いで出力を行うインタフェースを提案・実装し，初歩的な評価実験を行った結果について報告する．

2　眼電位測定の原理

眼球の角膜網膜間には，通常角膜側を正，網膜側を負とする電位が存在する．この電位を眼電位，あるいは角膜網膜電位と呼ぶ．顔面皮膚上に眼球を挟み込むように電極を貼付することで，眼球運動に伴う電位の変化を計測することができる．

3　システムの基本設計

本研究で提案するインタフェースを実現するシステムの概要図を Fig.1 に示す．本システムは入力部，認識部，出力部の 3 つの部分から構成されている．ユーザは眼球付近の顔面皮膚上に Fig. 1 のように電極を貼付し，眼球運動を行うことで眼電位の入力を行う．入力された眼電位は前処理を経て認識部に送られる．認識部では，受け取った眼電位を認識して単語列に変換し，その結果を出力部に送る．出力部は，受け取った認識結果に基づいて音声を合成し，出力する．以下で，各部分についての説明を行う．

3.1　入力部

Ag-AgCl 皮膚表面電極を，Fig. 1 に示すように，両眼の上部約 4cm，下部約 2cm，外側約 2cm，及び両眼の中間部，両眼上部の 2 電極の中間部の計 8 箇所に装着する．得られた眼電位を生体アンプを用いて増幅し，時定数 0.3sec のハイパスフィルタを通した上で，AD 変換器を介してコンピュータに記録する．眼電位は周波数 80 Hz 以上で収録すればよいことが報告されている [3] ため，本システムのサンプリング周波数は 100 Hz とした．装着した 8 箇所の電極の内、額中央部に装着したものを接地電極，両眼の中間に装着したものを基準電極とした．次に，記録したデータに対し前処理を行う．訓練データに対する前処理は，PCA(Principal Component Analysis) を用いて行う．テストデータに対しては，訓練データに対して求めた PCA の変換を適用した．

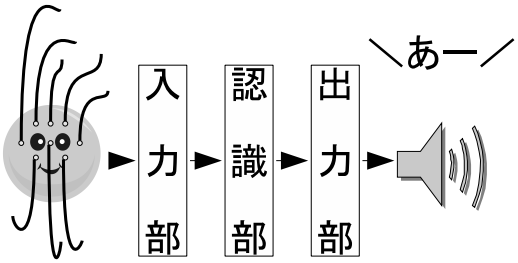

Fig. 1　システムの概念図

3.2　認識部及び出力部

認識部では,「中央」「上」「下」「右」「左」という 5 つの視線方向を，音声認識における 1 つの音素のようなものと捉え，モノフォンによる連続単語音声認識と同様の処理を行うことにより，眼電位を認識する．この際，連続した同一方向の入力における連続回数の区別を精度よく行うことはあまり容易でないと考えられるため，同一方向の連続入力は行わないものとし，各視線方向をそれぞれ一つ一つの状態として持つ認識ネットワークを設計した．認識ネットワークの各状態について，遷移可能な状態の数が常に 4 であることから，それぞれの状態は 2 ビットの情報量を持つ．このことから，連続した 3 状態を 50 音の 1 音 1 音に対応付けることで，提案システムによる自由発話が可能になる．ただし本研究では初歩的な評価を目的として，単純に一つ一つの視線方向をそれぞれ「なか」「うえ」「した」「みぎ」「ひだり」という 5 つの単語に対応付けたシステムとした．

出力部では，認識部から単語列を受け取り，リアルタイムに音声を合成して出力する．

---

*Realtime Speech Synthesis Interface using Electro-oculogram Input by Kento Ozaki, Takahiro Shinozaki(Tokyo Institute of Technology), Toshimitsu Musha(Brain Functions Laboratory), and Sadaoki Furui(Tokyo Institute of Technology)

Table 1　各方向毎の Precision, Recall, F 値

	次元圧縮なし			次元圧縮あり		
	P	R	F	P	R	F
center	0.95	0.26	0.41	0.94	0.76	0.84
down	0.84	0.95	0.89	0.77	1.00	0.87
left	0.96	0.86	0.91	1.00	0.98	0.99
right	0.74	0.92	0.82	0.99	0.72	0.83
up	0.87	0.89	0.88	0.99	0.98	0.99

4　評価実験

4.1　実験条件

提案したシステムを Windows PC 上に実装し，実験を行った．実験に用いたデータは，24 歳の男性被験者 1 名が提案インタフェースの入力部を用いて 2 度に分けて収録を行ったものであり，それぞれ延べ 640 単語，2560 単語の計 3200 単語からなる．オフライン認識実験では，前者をテストデータ，後者を訓練データとした．また，認識部には T^3 デコーダを用いた．事前に行った予備実験の結果から，音響モデルに相当する眼球運動モデルの状態数を 4，混合数を 8，PCA による圧縮次元数を 3 次元とし，T^3 デコーダに与える beam 幅，band 幅をそれぞれ 500，挿入ペナルティを 140 と設定した．また，比較のため，PCA による次元圧縮を行わない場合についても同様の実験を行った．こちらのシステムでの挿入ペナルティは，予備実験の結果から 230 とした．最後に，オフライン実験の際に作成したモデルを使用し，オンラインでシステムの動作確認を行った．

4.2　オフライン評価

PCA による次元圧縮を行わなかった場合の単語認識率が 77.2 %であったのに対し，次元圧縮を行った場合は 86.3 %と向上し，PCA による次元圧縮が有効であることが分かった．また，各視線方向についての F 値を Table 1 に示す．

次元圧縮の効果により，各視線方向毎の F 値はほぼ同等か，あるいは向上した．向上幅の小さかった下方視、及び右方視に着目し，上下，左右方向に分けてそれぞれの原因について考察する．まず上下方向に関しては，眼球上下の構造上の非対称性によると考えられる．角膜と瞼との接地面積の変化による電気抵抗の変化が，観測される眼電位の絶対値に影響するという報告がある [1]．下方視の場合，上方視の場合と比較して角膜と瞼との接地面積が狭いため，電気抵抗が大きく，電位変化の絶対値が小さい．このため中央視状態との弁別が比較的困難となり，認識率低下の一因となったと考えられる．左右方向に関してはこのような構造上の違いはないが，利き目の影響を受けることが考えられる．本実験における被験者の利き目は左目であり，左方視に比べ右方視の電位変化がやや不安定であった．このことが右方視の認識率低下の一因であると考えられる．また，この下方視及び右方視の認識精度の低さは，上方視よりは下方視を，左方視よりは右方視を苦手とする被験者の内観とも一致していた．

4.3　オンライン評価

実装した提案システムのオンラインでの動作確認を行った．実験において，1 視線方向の入力にかかった時間は平均約 0.7 秒であった．この入力速度で 50 音の入力による自然発話を行う場合，1 音の入力に約 2.1 秒かかる計算となる．さらに，本システムは認識に音声認識に用いられる HMM を用いているため，ある程度の入力の伸縮に対応することができる．このため，使用者の訓練により，入力速度の更なる向上が見込まれる．最大で 0.5 秒毎に 1 視線方向の入力を行う程度までは高速化可能であると思われる．

動作確認に際し，入力速度が一定以下になると認識率が大きく低下する現象が確認された．これは認識モデルの元となる訓練データの入力速度がほぼ一定であったためであると考えられる．訓練データ作成時に様々な入力速度のデータを用意することで低速な入力にも対応可能であると考えられる．また，固視微動や瞬目など，不随意の眼球運動によって，使用者の意図とは異なる入力が行われてしまう問題が発生した．この問題に対しては，眼電位によるなんらかのスイッチ操作をシステムの入力受付開始及び終了に割り当てることで対処可能であると考えられる．その他には，現在のシステムでは合成音声の発声速度が固定であるため，合成出力される音声の継続時間よりも高速な入力によって出力待ちの単語が蓄積されてしまい，インタフェースとしての直感性の低下が見られた．この問題に関しては，眼球運動を 50 音に対応付け，自由発話を可能にする場合などには，相対的に入力速度が遅くなることによって緩和されると考えられる．しかし，より積極的な解決法としては，合成音声の再生速度を眼電位の入力速度に同期するように改良することが考えられる．

5　まとめ

本稿では，即時性に優れる眼電位の連続単語認識を入力として用い，音声合成に繋いで出力を行うインタフェースを提案した．初歩的な認識実験を行い，提案インタフェースの性能を評価した．また，オンラインで提案インタフェースを動作させ、問題点を確認した．

参考文献

[1] 大矢哲也, 川澄正史, 生体医工学, 43(1), 172-178, 2005.
[2] 半田聡，海老澤嘉伸，映情学誌，63(5)，685-691，2009.
[3] 久野悦章 他，情処学誌，39(5)，1455-1462，1998.
[4] A. Bulling *et al.*, Journal of AmI and SmE, 1(2)，157-171，2009.