

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 22 日現在

機関番号：51303

研究種目：基盤研究(C)

研究期間：2011～2014

課題番号：23500637

研究課題名(和文) 視聴覚非同期刺激を用いた次世代両耳補聴システムの実用性能評価

研究課題名(英文) Practical evaluation for binaural hearing-aid system by using asynchronous stimuli

研究代表者

本郷 哲 (Hongo, Satoshi)

仙台高等専門学校・専攻科・教授

研究者番号：80271881

交付決定額(研究期間全体)：(直接経費) 3,800,000円

研究成果の概要(和文)：雑音除去等の処理を行うことのできる次世代の両耳補聴器において、両耳が塞がれるために、処理による時間遅れが問題となる。特に視聴覚間で時間差が生じる場合に、音声認知の不具合、定位感の不具合、などがあるため、この時間差条件を明らかにし、処理の遅れ時間の許容範囲を求めることが必要である。本研究の結果、この許容時間については、約100ms程度までの遅れ時間に収めることが必要であることがわかった。また、特に子音における影響が大きくなることも明らかとなった。さらに、移動音源に対して定位感を評価したところ、音源が人の歩行速度程度場合には、約100ms程度まで、定位に違和感を感じないことが明らかとなった。

研究成果の概要(英文)：In next-generation binaural ear-aid system, the asynchronous between audio signal and visual signal can be significant problem for speech recognition and sense of sound source localization. The aim of this study is to clarify the limit of time delay for processing time of binaural ear-aid system from the point of view of speech recognition rate and sense of sound localization.

We performed the speech recognition tests and movement sound source tests by using human examinee. As a result of this study, we found that it is necessary to reduce the delay time of processing in ear-aid system to 100ms for good Japanese speech recognition. In addition, it was revealed that the influence for recognition rate in the consonant sound is larger than the vowel sound. Furthermore, from the movement sound source test, it became clear that the limit delay time not to feel sense of incongruity is less 100ms when the speed of movement sound source is human walking speed.

研究分野：音響信号処理

キーワード：バイノーラル 両耳補聴器 マガーク効果 視聴覚刺激 時間遅れ

1. 研究開始当初の背景

軽度の聴覚障がい者が音情報を得るためには、古くから補聴器を活用し、音信号を増幅する手段を用いている。近年のデジタル技術の発展により、補聴器はデジタル化され、音声信号を単にそのまま増幅するのではなく、ハウリング抑圧、雑音除去、必要な信号のみの増幅など、高度な処理を行うのが一般的になっている。さらに、このような補聴システムは雑音除去や信号分離の機能を活かして、工場や発電所等の雑音環境下における、一般健聴者のコミュニケーションツールとしても活用が高まっている。

これまで、補聴器はモノラルの片耳装用が主であったが、近年は両耳聴の補聴システムが数々出現している。両耳聴にすることで、補聴システムユーザが元々持っている聴覚機能を活かすことができ、片耳装用に比べて倍以上の補聴効果が得られること、音源の定位という重要な情報を伝えることができること、2ch を基本とするデザインになるため、ブラインド信号処理などの雑音除去と相性が良いことなどの長所があるためである。

このような社会的要求がある中、本研究代表者は、これまで両耳補聴システムのための様々なタイプのバイノーラル音声強調手法 TS-BASE/WT, TS-BASE/SDMF, 等を開発してきており、科学研究費補助金の交付を受けてその実用研究も行ってきた。これらの研究成果によって、有効な両耳補聴性能が得られており、かつハードウェアに実時間実装も可能になっている。その性能は概ね 15dB 程度(全周波数帯における Segmental SNR)のノイズ除去(目的音強調)が可能となっており、かつ音声の品質劣化を表す LSD(log spectrum distance)が小さく高品質でノイズ抑圧ができています。さらに、本研究代表者らの提案する TS-BASE では、目的音はその定位感を損ねることがなく、ユーザに対して両耳による音源定位情報の提供も可能とする。また TS-BASE を中心とする特許を取得しており、その実用化研究が期待されている。

これまで提案してきた両耳聴補聴システムでは、音声情報のみを単独で評価した場合には、Segmental SNR や LSD 等の定量的な性能評価は良い結果が得られている。一方、フレーム間の音声に混在するパルスノイズ、ミュージカルノイズ、補聴システムのバッファリングによる時間遅れなどの影響があり、これらに起因する主観的な評価実験が実用化には不可欠となる。また、主観評価により、ユーザが実用する上での補聴システムの性能が評価されるので、実質的な性能向上と実際に使うユーザに適したチューニングが期待される。

このような音声評価は一般に音声単独で評価されることが殆どであるが、補聴システムを用いるユーザは音情報と同時に視覚情報も得ている。補聴システムでは、フレームごとに信号処理を行うため、どんなにプロセッサの性能を向上させてもフレーム時間窓長に対応する時間遅れが発生し、先行する視覚情報と遅延を有する聴覚情報が非同期による不整合状態の視聴覚情報をユーザが得ることになる。このような状況が起きた場合、例えば、道路上で両耳補聴システムを装用した歩行者に対し定位感のある音を提供できたとしても、定位音のわずかな時間遅れにより重大な事故をもたらしたりすることも考えられる。

視聴覚不整合が生じた場合の補聴システムの評価は、放送大学の西原が先行して行っている研究があるが、補聴システムの具備すべき要件を与えるようなものではない。

さらに、音源が移動していくときに、視聴覚の情報の非同期/不整合が生じた場合に、ヒトがどのように違和を感じるか、如何に定位するか、マガーク効果などの異聴現象の有無などは調査されていない。

2. 研究の目的

本研究では、以上の研究背景と本研究代表者のこれまでの研究成果を踏まえて、TS-BASE 補聴システム及びより一般的な両耳補聴システムについて、視覚情報まで含めた主観的観点の性能評価を行い、両耳補聴システムの実現上具備すべき要件を明らかにしていくことを目的とする。

具体的には、本研究代表者がこれまで提案してきた TS-BASE システムおよび理想的な補聴システムを想定した場合について、主観的な性能評価、主観的な定位感評価、視聴覚相互作用による補聴システムの遅れ時間の影響、の観点にたち、音源が移動する状況下での視聴覚同時に情報が入ったときの性能評価実験 (PESQ, MOS, 了解度など) を行う。音源が移動する状況は、実際に音源移動装置を作成して実現し、そのときの現実の視覚刺激とバッファにより遅れが発生する聴覚刺激を混在させた視聴覚情報を用いて、評価を行う。

また、これらの評価実験を通して、より一般化された両耳補聴システムが具備すべき要件を聴覚、人間工学的観点から明らかにすることを目的とする。

3. 研究の方法

1) 音声のみの両耳聴音声了解度評価

上記のように確立された主観評価実験に従って、様々な条件下 (例えば環境の SNR, 残響量, 雑音の種類など) で補聴システムの評価を行う。なお、原音声のみの場合、原音声中に雑音を加え処理を行わない場合等についてもバイアス試験のために評価する。なお、この音声のみの主観的評価実験は、次の視聴覚相互作用を考慮した評価実験と比較することを前提に実験を行う。すなわち、視覚情報を付加した場合の条件と同様な音条件において主観的評価実験を行う。さらに、様々な条件における、音声のみの音声了解度を調べ、以降の実験と比較するための知見を得る

2) 音声と映像の時間ずれが生じるときの音声了解度評価

購入した高速度カメラを用い、音声の遅れ時間を唇の動き等から精密に測定する。また、それぞれの音声遅れ時間に対して、音声了解度を調査し、母音、子音別の調査、マガー効果等の異聴現象の影響等を精査する。

3) 移動する音源に対する定位感の評価

移動する音源に対する定位感の評価を行う。具体的には、スピーカーとディスプレイをレールに沿って移動させ、ディスプレイに話者映像を写し、様々な音声遅延を与えながら、了解度を調べる。

4. 研究成果

本章では、1) 音声のみの両耳聴音声了解度の評価及び、2) 映像と音声遅延に対する両耳聴音声了解度の評価を合わせて示す。

音声データセット (FW03) に収録されている単語を発声した映像を、図 1 に示すようなタイムラインで音声に遅延を与え、ピンクノイズを加えて流した。本実験で用いる FW03 (Familiarity - controlled Word Lists 2003) とは、親密度別単語了解度試験用音声データセットであり、難聴者のための単語了解度試験用単語リストに含まれる全単語が発声・収録されている。FW03 のリストのうち、使用するのは親密度が低いものであり単音節の実験に近いものとなる。また、視覚による情報補完が起りやすいように、ピンクノイズを付加して実験を行った。FW03 のリスト名と音声遅延は表 1 のように対応している。このような秒数の音声遅延を与えたのは、坂本らの論文 [3] において、非同期条件の検知限がおよそ 200ms であったため、その前後が了解度の臨界と考えたからである。

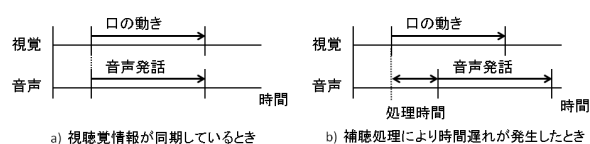


図 1 補聴処理による音声時間遅れ

表 1 FW03 と音声遅延の対応

list a	list b	list c	list d	list e	list f	list g	list h
0ms	10ms	20ms	50ms	100ms	200ms	300ms	400ms

本実験では、1 単語につきおよそ 2 秒の映像を提示し、その映像の話者が何と言っているように聞こえたかを解答用紙に記入させた。実験は情報デザイン棟 2 階防音室で、20 歳前後の聴覚正常な人を対象として行った。被験者数は男性 3 人、女性 3 人の計 6 人とした。聴取実験の結果は以下の通りである。作成したグラフの、縦軸は正答率、横軸は用いたリスト、エラーバーは標準偏差を表している。

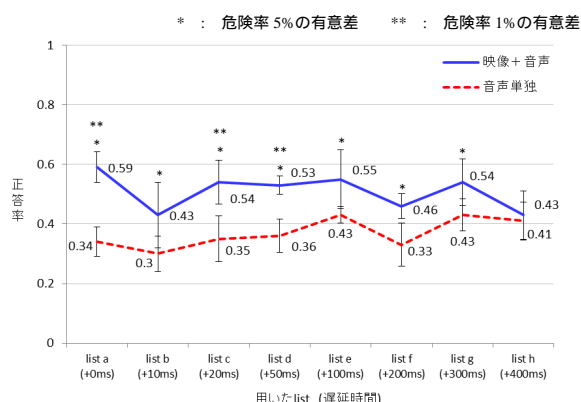


図 2 聴取実験正答率の平均と標準偏差

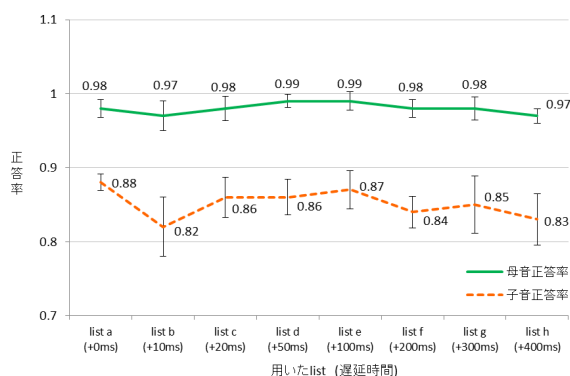


図 3 聴取実験母音・子音別正答率の平均

図 2 の結果より映像付加による音声理解度の向上は、100ms 程度の時間差で収束しており、100ms 程度以上の時間差がある場合、視覚と聴覚の情報補完がほとんど行われなくなることを示している。すなわち、両耳補聴システムにおいても 100ms 秒の処理時間（音声 が 100ms 視覚よりも遅れる）以下であれば、視覚による情報補完が行われることを示し

ている。

また、図 3 の結果より子音の正答率変化が遅れ時間とともに大きく変化していくことがわかる。このことは、視聴覚相互作用により主に子音の発話認識に対する視覚の情報補完が大きく、かつ処理時間の遅れは、主に子音認知に影響を与えることを示唆している。マガーク効果が、子音の異聴であることから従来からの知見からも妥当な結果であると考えている。

さらに、このような映像と音声を移動する状況下で、映像を動かしながら、音声提示スピーカーを移動させると、速度 1m/s 程度（室内歩行速度程度）の場合、100ms の遅れで定位上の違和感が大きくなることが明らかとなった。

以上のことから、両耳聴補聴器においては、了解度の観点および定位感の観点からも 100ms 以内の処理時間に抑えることが重要な要件であることが示される。また、この遅れ時間は、サンプリング周波数 16kHz の場合には、1600 点に相当するものである。このことは、例えば、サンプリング周波数 16kHz のシステムにおいては、FFT に基づく雑音抑制処理を行う場合、512 点、1024 点程度までの処理点数に抑える必要があることを示している。

5. 主な発表論文等

〔雑誌論文〕(計 2 件)

Shuichi Sakamoto, Satoshi Hongo, Yoiti Suzuki, "3D Sound-Space Sensing Method Based on Numerous Symmetrically Arranged Microphones," IEICE Trans. Fundamentals Vol. E97-A 2014 pp. 1893-1901 (査読有)

Satoshi Hongo, Shuichi Sakamoto, Yoiti Suzuki, "Binaural Speech Enhancement Method by wavelet transform based on interaural level and argument differences," Proceedings of International Conference on Wavelet Analysis and Pattern Recognition. CD-ROM (2012) (査読有)

〔学会発表〕(計 6 件)

高橋琴音、本郷哲, " 両耳補聴時における聞き取り性能評価 ", 高専シンポジウム 2015 in 函館予稿集, 2015 年 1 月 10 日 函館工業高等専門学校

板垣李夏、本郷 哲, " 音声プライバシー保護のためのマスキング音生成法の検討 " 平成 26 年高専シンポジウム in 久留米予稿集, 2014 年 1 月 25 日久留米工業高等専門学校

A Nakai, M Luimula, S. Hongo H. Vuoda, " Ecaluating an Inter-Cognitive Game Equips Motion-Based Control by Using Kansei Engineering, " CogInfoCom2013, 2013 年 12 月 2 日~5 日 Budapest, Hungary.

板垣李夏、本郷 哲, " 音声プライバシー保護のためのマスキング音に関する考察 " 平成 24 年度電気関係学会東北支部連合大会, 2012 年 8 月 30~31 日, 秋田県立大学

平間有紀、本郷哲, " 音楽のリズム鮮鋭度による音楽ジャンル判別の検討 ", 平成 24 年度電気関係学会東北支部連合大会, 2012 年 8 月 30~31 日, 秋田県立大学

中井彬人、矢入聡, " Setting a System to Control Multiple-Channel Loudspeakers Outputing Acoustic Signals " 平成 24 年度電気関係学会東北支部連合大会, 2012 年 8 月 30~31 日, 秋田県立大学

6. 研究組織

(1) 研究代表者

本郷 哲 (HONGO, Satoshi)

仙台高等専門学校・専攻科・教授

研究者番号 : 80271881

(2) 研究分担者

矢入 聡 (YAIRI, Satoshi)

仙台高等専門学校・専攻科・准教授

研究者番号 : 00447187