様式 C-19

科学研究費補助金研究成果報告書

平成 22 年　5 月 24 日現在

研究種目：若手研究(B)
研究期間：2008～2009
課題番号：20700193
研究課題名（和文）
話者映像及び話声が含む感性情報を考慮した高精度高感性視聴覚音声提示システムの構築
研究課題名（英文）
Development of advanced audio-visual speech communication system for transfer Kansei information
研究代表者
坂本　修一（SAKAMOTO SHUICHI）
東北大学・電気通信研究所・助教
研究者番号：60332524

研究成果の概要（和文）：本研究では，視聴覚情報の持つ様々なパラメータが音韻知覚及び視聴者の受ける感性情報に及ぼす影響を定量化し，その知見を生かした視聴覚音声提示ユニバーサルコミュニケーションシステムの構築手法を提案する。研究の結果，高齢者は健聴者に比べ，音声と映像のズレに寛容であり，かつ，そのような場合でも映像による音声の聴き取り向上が見込めることが明らかとなった。この知見は，単に音声だけ話速を遅くして映像と組み合わせた場合でも，映像の寄与が見込めるということを示唆しており，新しいコミュニケーションシステムの可能性を示すものである。

研究成果の概要（英文）：The aim of this study is to develop next generation audio-visual communication system. For this purpose, I investigated how people integrate speech sound and moving image of talker's face not only for understanding speech signal but also for perceiving a kind of sensational and emotional (Kansei) information. I focused on the effect of speech-rate for audio-visual speech understanding, because "speaking slowly" is very good way to talk to older adults, especially hearing impaired listeners. I used speech-rate conversion technique to slow down speech-rate and synthesized signal was combined with original movie. The results of experiment suggested that older adults are tolerant of audio-visual asynchrony. This fact implies the possibility of new audio-visual communication system, which enhance speech understanding by slowing down auditory information.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2008 年度	1,500,000	450,000	1,950,000
2009 年度	1,800,000	540,000	2,340,000
年度			
年度			
年度			
総　計	3,300,000	990,000	4,290,000

研究分野：総合領域
科研費の分科・細目：情報学・感性情報学・ソフトコンピューティング
キーワード：感性インタフェース，マルチモーダル情報処理，視聴覚統合，ユニバーサルデザイン

１．研究開始当初の背景

マルチメディア技術の進歩により，高品位

な映像・音声情報がネットワークをとおして容易に通信可能となってきている。これに伴い，単に映像，音声情報の伝送だけでなく，話者の細かい表情の変化や，感情の表出による話声の変化といった，感性情報の一端をも視聴者へ伝送する可能性が広がりつつある。

一方で，高齢化社会の急速な進展に伴い，高齢者に優しい音声通信システムの要求も高まりを見せている．高齢者，特に，老人性難聴者は，音声情報の取得が困難な場合が多く，話者の唇の動きを始めとする視覚情報の寄与も大きいと考えられる。これまでの研究により，単に口形情報だけでなく，話者映像と音声からうける感性情報も，単に視聴者の感情に作用するだけでなく，音韻知覚に影響を及ぼすことが示されている。したがって，本来あるべき感性情報を正しく伝送することが，音声情報通信という観点からも重要である。

以上のように，話者映像を含めた音声提示システムを構築する際には感性情報の伝送が重要となると考えられる。しかし現状では，どのように視聴覚情報を処理すれば感性情報が正しく伝送出来るかといった知見は非常に少ない。更に，話者の映像，声質，話速，視聴覚情報の同期といった個々のパラメータ，もしくは，これら複数のパラメータの相互作用が，視聴者が知覚する音韻情報や，視聴者が受ける感性情報にどのように作用し，また，その両情報がどのように結びつくのかということは，ほとんど解明されていないといえる。

２．研究の目的

本研究の最終的な目標は，話者の感性情報をも視聴者へ提示可能な高品位高感性音声提示システムを構築することにある。

そのために本研究では，話者映像の有無や，話速といった話声の性質，話者映像と話声の同期，話者の口の動きといったパラメータが，音韻知覚及び視聴者の感性情報にどのような影響を及ぼすか定量的に明らかにする。そこで得られた知見に基づいて，実際の話者映像や音声情報から音韻知覚及び視聴者の受ける感性情報に影響を与える要因を主観評価実験に基づき抽出する。

以上の結果から，感性情報を効率的に伝送するユニバーサルデザインを指向した高感性音声提示システムの構築手法を提案する。

３．研究の方法

本研究では，話者映像と音声を同時に提示し，そこから視聴者が得る様々な感性情報，音韻知覚情報に基づき，視聴覚音声提示システムが具備すべき要件を洗い出す。したがって，収録した映像・音声の加工，それを用いた主観実験，結果の分析のサイクルを繰り返すことで，最終的に必要となる知見を得ることとなる。

さらに，ユニバーサルコミュニケーションシステムの構築という観点を考えると，単に若齢者に対する知見だけでなく，高齢者を対象とした知覚心理実験を様々な角度から実施し，両者の結果の比較を詳細に行う必要がある。

そこで，以下に示すような方法で，研究を効率的に行った。

(1) 研究に必要な刺激を収録する。申請者はこれまでにも話者映像と音声を用いた聴取実験を多数行っており，その際に収録した多数の話者映像，音声素材がある。音声の内容も，文章から単語まで多岐にわたるものである。そこで，本研究を行う上で必要となる刺激素材の検討を進め，不足した素材の収録を行う。なお，刺激収録時には，収録刺激の質を保つため，これまでと同様に発話訓練を行った放送部所属の話者に依頼する。

(2) 収録した音声と映像刺激とを組み合わせ，特に，映像と音声の時間的同期のズレが，映像と音声から受ける感性情報にどのように影響するかを分析する。時間的なズレを生じさせる手法として，高齢者とのコミュニケーションにおいて重要なポイントとなる話速に着目し，音声の話速のみを伸長し通常速度の映像と組み合わせることでズレを生じさせ，影響を定量化する。合わせて騒音下での明瞭度試験も行い，感性情報と明瞭度との関連についても検討を行う。これらの実験は，若齢者だけでなく高齢者も調査対象とし，加齢に伴う感性情報の知覚の変化や明瞭度の変化，両者の関係についても分析する。

(3) 上記実験で得られた知見が，刺激音の長さに依存して変化するかを調べるため，モーラ数の多い単語と少ない単語を刺激音とし，映像と音声の時間的同期のズレの検知限，許容限が加齢に伴いどのように変化するかを実験的に検証する。

(4) 以上得られた結果を総合的に考察し，特にズレの許容限と単語了解度との関連を分析することで，映像と音声のズレにより引き起こされる可能性の高い違和感などの感性情報の影響が少ない音声通信・提示システムの設計指針を提案する。合わせて，違和感以外の感性情報の影響についても検討を行い，高感性音声提示システムとして備えるべきパラメータも明らかにする。

４．研究成果

得られた結果をまとめると，(1)話速を伸長することによる視聴覚音声の了解度への

影響と，(2)話速を伸長することにより発生する映像と音声のズレの検知限・許容限といった感性指標による評価，の 2 つに大別することができる。以下では，それぞれの項目について，若齢者，高齢者を対象にした実際の実験結果を用いて説明する。

(1) 話速伸長音声と話者映像の提示が音声了解度に及ぼす影響（学会発表 2，3）

高齢者を対象に，7，8 モーラ単語を用いて，話者映像と時間伸長音声による単語了解度を測定し，話者映像と時間伸長音声の統合メカニズムの若齢者と高齢者での差異を考察した。

実験に際しては，ほぼ正常な視力（矯正視力も含め，両眼で 0.7 以上），聴力（4 分法での平均聴力レベル 9.5±3.0 dB）を有する 65 歳以上の成人 16 名（平均 71.3±3.4 歳）を実験参加者とし，話速伸長音声と通常速度話者映像を，両者の開始点が同期するような形で組み合わせて提示し，音声のみを提示した際の了解度と比較した。また，映像の効果の指標として(AV-A)/(100-A)（AV benefit）を採用した。

実験結果を図 1 に示す。全伸長量における単語了解度は A 条件より AV 条件の方が向上していることがわかる。これは，若齢者，高齢者とも同様の結果であった。しかし，音声伸長量が大きい際の単語了解度に関しては，若齢者と高齢者で傾向が異なっていた。「映像＋音声」条件では，若齢者では音声伸長量 0 ms に比べ音声伸長量 100 ms の場合のみ有意に単語了解度が高くなったのに対し，高齢者の実験結果では，それ以外に音声伸長量 200 ，300 ms でも単語了解度が高くなった（$p < .05$）。これについては，「音声のみ」条件で，音声伸長量 0 ms 時の単語了解度に対する音声伸長量 400 ms 時の単語了解度が，若齢者では低下するのに対し，高齢者では低下しないというように，そもそも単語了解度に対する音声伸長の効果が，高齢者の方が大きいことが原因としてあげられる。その結果，映像と音声とのずれによる了解度の低下が発生する可能性が出てくるような音声伸長量が大きい条件でも，その低下分を補う形で単語了解度が高くなったと推察される。この知見は，ユニバーサルデザインを指向した視聴覚音声コミュニケーションシステムを構築する上で，重要な知見となる。

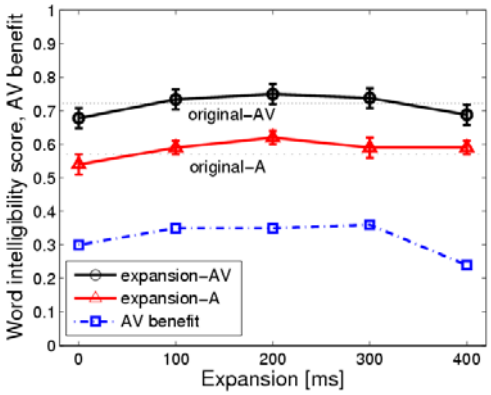

図 1 高齢者の単語了解度試験結果

(2) 話速伸長音声と話者映像のズレの感性評価（学会発表 1）

先に示したように，高齢者は若齢者に比べ，広い範囲で音声伸長による了解度の向上が認められた。しかし，実際に使った刺激は音声のみ伸長しており映像とのズレが生じていることから，若齢者，高齢者ともそのズレによる違和感などの感性情報が了解度に何らかの影響を及ぼしていることが考えられた。そこで，若齢者，高齢者を対象に，話速伸長音声と話者映像とのズレの検知限，許容限を分析した。

日本語を母語とし，ほぼ正常な聴力（4 分法による平均聴力レベル 18.4±3.8 dB）と，正常な視力（矯正も含む）を持つ高齢者 10 名（平均 70.3±2.7 歳）と，10 名の若齢健聴者を実験参加者とし，4 モーラと 7，8 モーラの単語を刺激単語として，先と同様に話速伸長音声と通常速度話者映像を両者の開始点が同期するような形で組み合わせて提示（EXP 条件）し，両者のズレが検出できたか否か（検知限），許容できる範囲であったか否か（許容限）を測定した。なお，実験の際には，通常速度の音声を単純にずらして話者映像と組み合わせた条件（ASYN 条件）でも，検知限，許容限の測定を行った。

4 モーラ単語により得られた検知限，許容限を図 2 に示す。図から明らかなように，検知限に比べ許容限が大きくなっている。

若齢者と高齢者のズレの検知限，許容限の比較を図 3 に示す。図 3 を見ると，4 モーラ単語，7，8 モーラ単語といった単語長や，検知限，許容限といった感性指標の違いにかかわらず，高齢者の方が若齢者に比べ，視聴覚情報のズレに鈍感・寛容であることが見てとれる。このことが，先の実験結果で高齢者の方が若齢者に比べ，広い範囲で話速伸長の効果が観察された理由の一つであると推察され，音声知覚と感性情報知覚の両者の対応関係が得られたと考えている。

(1)，(2)の結果を総合すると，ユニバーサルデザイン指向の新しい視聴覚音声提示システムの姿が見えてくると考えている。特にこれまで漠然としか知られていなかった高齢者に対して「ゆっくり話す」ことの効果を，了解度，感性指標といった様々な角度から明らかにしたことは，今後のシステム構築に向けて重要な知見が得られたと確信している。

今後は，今回得られた結果に基づいて視聴覚情報提示システムを構築し，他の感性情報

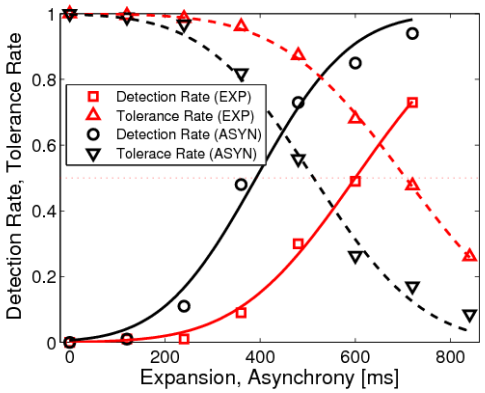

図 2 高齢者における 4 モーラ単語の検知限と許容限

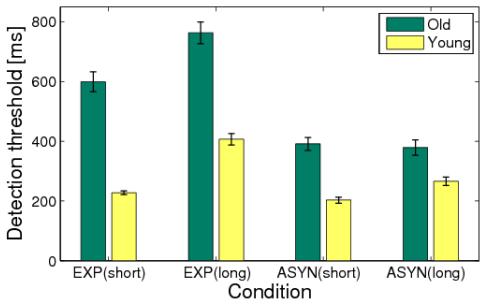

図 3 高齢者と若齢者でのズレの検知限と許容限の比較

も含めた様々な情報を精度高く提示すべく，主観的客観的な実験に基づき知見を集めていく必要があると考えている。

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔学会発表〕（計 3 件）

1. S. Sakamoto, “Aging effect on audio-visual speech asynchrony perception: comparison of time-expanded speech and a moving image of a talker’s face,” Proc. International Conference on Auditory-Visual Speech Processing 2009 (AVSP2009), pp. 9-12, 2009. 9. 10, Norwith, UK
2. 坂本修一, “高齢者を対象とした 7, 8 モーラ単語了解度における音声伸長量と話者映像の影響,” 日本バーチャルリアリティ学会VR心理学研究委員会第 13 回研究会, VRP-2009-4, pp. 9-10, 2009. 4. 24, 仙台
3. S. Sakamoto, “Effect of seech-rate and a moving image of a speaker’s face on 7 and 8-mora recognition in older adults,” Proc. Mini RIEC Workshop on Multimodal Perception, p. 55, 2009. 4. 25, Sendai, Japan

６．研究組織

(1)研究代表者

坂本　修一（SAKAMOTO SHUICHI）

東北大学・電気通信研究所・助教

研究者番号：60332524