

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 24 年 5 月 7 日現在

機関番号：32643

研究種目：基盤研究（B）

研究期間：2008～2011

課題番号：20300080

研究課題名（和文） 対人コミュニケーションにおける視聴覚統合の役割

研究課題名（英文） Integration of auditory and visual information in human communication

研究代表者

望月 要 (MOCHIZUKI KANAME)

帝京大学・文学部・教授

研究者番号：80280543

研究成果の概要（和文）：対人コミュニケーションにおける話し手の表情と音声という視聴覚情報の相互作用について、音声映像非同期の知覚とその影響を指標に、実験心理学的手法による実証的研究を行ない、非同期知覚への文化的要因の関与、発話文の意味の関与、主観的音量知覚の変化などの知見を得た。また音声コミュニケーションにおける視覚情報（話者の表情）の貢献を検討した実験からは、視覚情報利用に個人差があること、表情の誇張が必ずしも視聴成績向上につながらないことなどを明らかにした。

研究成果の概要（英文）：We examined the interactions between audio and visual information in human oral communications using the asynchrony detection tasks. We found some cultural factors in asynchrony detection such as native language of speakers and listeners, facial features, and so on. The understandability of the meaning of sentences are also seemed to be important for asynchrony detection. We also found the phenomenon in which the perceived volume of sound decrease when the audio and visual information is no synchronized. Our experiments on the contributions of the visual information to speech understandings shows there is large individual differences in utilization of visual information to speech cognition.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2008年度	3,900,000	1,170,000	5,070,000
2009年度	1,800,000	540,000	2,340,000
2010年度	2,000,000	600,000	2,600,000
2011年度	1,700,000	510,000	2,210,000
総計	9,400,000	2,820,000	12,220,000

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：感性認知科学

1. 研究開始当初の背景

動画像における映像と音声の時間同期の問題は、いわゆるリップシンク (lipsync) と呼ばれ、専らテレビ放送の品質評価、品質保証の観点から研究されてきた。赤井田ら (1997) はアナウンス映像と、打楽器演奏の2種類について評価実験を行ない、ア

ナウンスの場合、映像・音声の非同期（ズレ）の知覚閾（検出閾）は、音が先行した場合 46 ms、映像先行で 122 msであることを示し、これを基に ITU-R の勧告がまとめられた。近年では、映像・音声の非同期を自動的に補正するシステムも開発され、放送技術の問題としてのリップシンクは一応

の解決をみたと言え。しかし、認知科学の視点からみると、この現象には未だ解明されていない問題が残されており、それを研究することは、ヒューマン・コミュニケーションにおける視覚情報の役割を理解する上で役立つばかりでなく、ヒトの視聴覚情報統合処理一般を理解する上で重要な知見を提供するものと考えられる。

2. 研究の目的

(1) 非同期知覚閾

光点のような単純刺激の場合、点灯時間差の知覚閾は 13 ms から 50 ms 程度であることが知られており、音先行時の非同期知覚閾 46 ms という値は、この範囲に収まっている。僅か 46 ms の非同期に気付くということは、ヒトは日常的に話し手の唇の動きと音声の関係に注意を向けているからであろうと考えられるが、我々は日常生活のなかで、それを意識することはない。リップシンクにおける非同期知覚の感受性が、どの程度のもので、それがどのように形成されるのかを明らかにするのが第 1 の目的である。

(2) 非同期と音量知覚の関係

非同期知覚閾に関する実験的研究の途上で、唇の動きと音声同期していないと感じられるときは、知覚される音声の音量が小さく感じられることが明らかになった。従来、微細なリップシンクのズレは、聞き手に“違和感”や“聞き難さ”、発話内容の“分かりにくさ”といった印象を与えることが知られていた。それに対して主観的音量の変化は、評定の信頼性も高く、聴覚と視覚の相互作用を、より直接的に関係した指標であると思われる。この新たに見出された現象について、詳細な知見を得るのが第 2 の目的である。

(3) 視覚情報の音声知覚への貢献

46 - 122 ms 程度の僅かの非同期を検出できるということは、我々が他人の発話を聞いているとき、その表情に非常な注意を向け

ていることを強く示唆している。日常的に言葉を交わす時、常に注意を払い続けているのではないかも知れないし、その過程を意識しているわけでもないだろう。あるいは、発達段階の初期に、注意を集中した時期があったのかも知れない。可能性は幾つも考えられるが、いずれにせよ、微細な非同期を検出するためには、音声と唇の動きが同期した状態を十分に“知っている”必要があることは間違いない。すると次に、微細な非同期を容易に検出できる程度詳細に、唇の動き或は表情に注目する理由が問題になる。少なくとも晴眼者は、他人の発話を理解するのに、話し手の唇の動きを手掛かりをしている自覚はない。しかし、以上の考察は、話し手の唇の動きないし表情が、音声コミュニケーションに何らかの重要な役割を果たしている可能性を示している。本研究の第 3 の目的は、音声コミュニケーション場面における視覚情報、特に唇の動きを中心とした話し手の表情変化の役割を明らかにすることである。

3. 研究の方法

(1) 恒常法による非同期認知閾の測定

音声と映像の間に数段階の時間差を設けた非同期映像と同期映像を、話し手の顔面がアップになるように無作為な順で反復呈示し、同期の程度や違和感などについて、視聴者の印象を 2 件法ないし 3 件法で測定し、回答の生起率に基づき非同期知覚閾を測定した。呈示する映像の内容を、話し手の国籍、言語、相貌、発話内容（普通文、逆転、無意味文）などについて変化させ、これら要因の知覚閾と主観的音量知覚への影響を評価した。

(2) 一対比較法による非同期の影響の測定

音声と映像の間に数段階の時間差を設けた非同期映像と同期映像 2 つを、話し手の顔面がアップになるように無作為に組合せ、継時的に呈示して、印象を評価する一対比較法によって、話し手の要因、文の要因などが非同期知覚閾と主観的音量知覚に及ぼす影響を測定した。

(3) 画像処理による表情と音声知覚の

4. 研究の成果

(1) 研究用標準動画像

発声訓練を受けた日本人男女に、音素数や音素の繋がり、発話速度などを可能な限り統制した日本語文章、日本語音素を無作為に並べた無意味な音素列、日本語母音、日本語子音を発声させ、ハイビジョン映像に記録した。これは、今後の研究で必要になる各種視聴覚素材作成に広く利用可能である。また、コーカソイド的容貌を持ちながら日本語を母国語とする話し手に、同様の標準的日本語文を発声させた映像も記録できた。これは、相貌が視聴覚コミュニケーションに与える影響を検討する上で貴重な実験素材として活用できると考えている。

(2) 文化的要因

外国人話者が外国語（米語、ロシア語、韓国語、中国語）を話す映像を、日本人視聴者に呈示した非同期閾測定実験の結果から、視聴者にとって外国語の場合、非同期閾が上昇することが確認された。これは常識的に予測されることであったが、韓国語について特異な結果が見出された。即ち、韓国語を全く知らない日本人視聴者であっても、韓国語の非同期知覚閾は、日本語よりは高いものの、米語よりも日本語に近い傾向が明らかになった。実験に参加者した視聴者達は、程度の差はあれ英語教育を受けているので、米語での閾値が全く未知のロシア語より低くなることは常識的に理解できるが、全く未知の韓国語で、ロシア語より非同期検出が容易であることは、非同期検出に、発話言語の音声学的特徴や、話し手の相貌などが関与していることを示唆するものである。コーカソイド容貌の話し手が話す日本語の知覚実験では、相貌の効果は有意ではなかった。一方、視聴者自身が話す映像と、他者が話す映像を比較した実験では、女性においてのみ、自分自身映像を視聴する方が非同期知覚閾が低下すること

が見出されたことから、話者の相貌が非同期知覚に何らかの影響を及ぼしていることは否定できないが、その詳細は明らかにできなかった。

(3) 主観的音量知覚への影響

これは、本研究の途上で新たに発見された現象である。日本語では非同期時に知覚音量が低下し、特に映像が遅延した場合、その影響が顕著であった。一方、韓国語、米語、ロシア語では、この現象は生じなかった。また、通常の日本語と、それを逆転再生したもの、日本語音素を無作為に並べた発話（無意味発話）を比較した実験では、逆転再生だけが非同期の影響を受けなかった。逆転再生では、音素と唇の動きの関係は、通常の日本語と同一であることから、この結果は、非同期が主観的音量知覚に与える影響が、単に音声学的要因だけによるものでないことを示している。

(4) 音声知覚に貢献する顔面部位

素材映像撮影時に、話者の顔面にマーカーを貼付せずに顔面特徴点の動きを検出し、その移動量を加工できる実験システムを開発することができた。これにより、制約の少ない状況で、以前より容易に各種発話素材の製作が可能になった。

唇周辺の特徴点の動きを何段階かに変え、音声にノイズを重畳して、聞き取りの正確さと実験参加者の視点位置を測定した実験からは、顔面の動きが聞き取りに貢献する程度には著しい個人差が存在することが明らかになった。即ち音声情報が劣化した時の聞き取りの正確さが、唇の動きの量によって影響を受ける実験参加者と、必ずしもそうではない参加者が存在した。音声情報劣化時の視線の動きも同様で、唇周辺を注視する傾向にも個人差がみられた。唇の動きを誇張すると、聞き取り正確度が増加すると予測したが、実験結果は、不自然な動きは視聴に貢献しないことを示唆するものだった。動きの加工を、よ

り自然にして追試を実施することと、唇から離れた顔面位置の関与について、今後も研究を進める予定である。

5. 主な発表論文等

[雑誌論文] (計 0 件)

[学会発表] (計 26 件)

- ① 大西 仁・望月 要 (2012). 視聴覚情報の非同期が音量知覚に与える影響 (2), 電子情報通信学会技術報告, HIP2011-84. (2012 年 3 月 29 日, 神戸)
- ② 遠藤浩幸・林 大作・望月 要・大西 仁・中村直人 (2012). 口動作を編集した発話映像による音声聞き取りへの影響, 電子情報通信学会技術報告, HIP2011-85. (2012 年 3 月 29 日, 神戸)
- ③ 大西 仁・望月 要 (2012). 音声・映像の非同期が与える心理的效果: 使用言語の影響 (6), 電子情報通信学会技術報告, CQ2011-67. (2012 年 1 月 20 日, 大阪)
- ④ 大西 仁・望月 要 (2011). 視聴覚情報の非同期が音量知覚に与える影響, 電子情報通信学会技術報告, HIP2011-45. (2011 年 11 月 10 日, 仙台)
- ⑤ 大西 仁・望月 要 (2010). 使用言語が視聴覚相互作用に与える影響: 同時性判断および音量知覚への影響, 日本認知科学会第 27 回大会, P3-20. (2010 年 9 月 19 日, 神戸)
- ⑥ 大西 仁・望月 要 (2010). 音声・映像の非同期が与える心理的效果: 使用言語の影響 (5), 電子情報通信学会技術報告, CQ2010-46. (2010 年 9 月 10 日, 福岡)
- ⑦ 大西 仁・望月 要 (2010). 音声・映像の非同期が与える心理的效果: 使用言語の影響 (4.1), 電子情報通信学会技術報告, CQ2010-23. (2010 年 7 月 8 日, 北海道)
- ⑧ 大西 仁・望月 要 (2010). 音声・映像の非同期が与える心理的效果: 使用言語の影響 (4), 電子情報通信学会技術報告, CQ2010-21. (2010 年 7 月 8 日, 北海道)
- ⑨ 大西 仁・望月 要 (2011). 伝送遅延が知覚・認知・行動に与える影響, 電子情報通信学会総合大会. (2011 年 3 月 14 日, 東京)
- ⑩ 望月 要・大西 仁 (2011). 視聴覚情報の非同期が視聴覚統合に与える影響 (5): 使用言語が音源定位に与える影響, 電子情報通信学会総合大会. (2011 年 3 月 14 日, 東京)
- ⑪ 伊東和廣・望月 要・大西 仁・中村 直人 (2009). 発話理解における視覚情報の認識, 電子情報通信学会第 7 回 QoS ワークショップ, QW7-20. (2009 年 11 月 20 日, 東京)
- ⑫ 山本由紀子・仁科エミ・大西 仁 (2001). 一対比較による感性情報の評価, 電子情報通信学会第 7 回 QoS ワークショップ, QW7-14. (2009 年 11 月 20 日, 東京)
- ⑬ 大西 仁・望月 要 (2009). 音声・映像の非同期が与える心理的效果: 使用言語の影響 (2), 電子情報通信学会技術報告, CQ2009-33. (2009 年 9 月 11 日, 高山)
- ⑭ 伊東和廣・望月 要・大西 仁・中村 直人 (2009). 発話理解における視覚情報の役割, 電子情報通信学会技術報告, CQ2009-34. (2009 年 09 月 11 日, 高山)
- ⑮ 大西 仁・望月 要 (2009). 音声映像の非同期と視聴覚統合 (3), 電子情報通信学会技術報告, CQ2009-12. (2009 年 4 月 17 日, 沖縄)
- ⑯ 大西 仁・望月 要 (2009). 視聴覚情報の非同期が視聴覚統合に与える影響 (1): 使用言語が音量知覚に与える影

- 響, 電子情報通信学会総合大会, B-11-24. (2009年03月19日, 仙台)
- ⑮ 望月要・大西仁 (2009). 視聴覚情報の非同期が視聴覚統合に与える影響 (4): 用言語の意味理解、話速が音量知覚に与える影響, 電子情報通信学会総合大会, B-11-28. (2009年3月19日, 仙台)
- ⑯ 伊東和廣・望月要・大西仁・中村直人 (2009). 発話者の顔映像による音声聞き取りへの効果, 電子情報通信学会技術報告, PRMU2009-278, HIP2009-163. (2009年3月16日, 鹿児島)
- ⑰ 大西仁・山本由紀子・望月要 (2009). 音声・映像の非同期が与える心理的効果: 使用言語の影響 (3), 電子情報通信学会技術報告, CQ2009-83. (2010年01月21日, 京都)
- ⑱ 望月要・大西仁・岡本悠花 (2009). 視聴覚情報の非同期が視聴覚統合に与える影響 (2): 非同期知覚に対する唇の動きの貢献 電子情報通信学会総合大会講演論文集, B-11-28. (2009年3月20日, 松山)
- ⑲ 大西仁・望月要・岡本悠花 (2009). 視聴覚情報の非同期が視聴覚統合に与える影響 (1): 視聴対象話者の相貌の影響 電子情報通信学会総合大会講演論文集, B-11-27. (2009年3月20日, 松山)
- ⑳ 大西仁・望月要 (2008). 音声映像の非同期が与える心理的効果: 使用言語の影響 電子情報通信学会技術報告, HIP2008-147. (2008年12月18日, 仙台)
- ㉑ 大西仁・望月要 (2008). 遅延の変動がユーザーの作業成績と遅延の知覚に与える影響 電子情報通信学会技術報告, CQ2008-55. (2008年11月14日, 郡山)
- ㉒ 大西仁・望月要 (2008). 音声映像の非同期と視聴覚統合 (2) 電子情報通

信学会技術報告, CQ2008-30. (2008年9月25日, 高松)

- ㉓ 望月要・大西仁 (2008). 視聴覚情報の非同期と視聴覚統合: 洋画の吹替えにあまり違和感を覚えないのはなぜか? 日本認知科学会第25回大会, O5-1. (2008年9月6日, 京都)
- ㉔ 大西仁・望月要 (2008). 音声映像の非同期と視聴覚統合 電子情報通信学会技術報告, CQ2008-13. (2008年7月17日, 釧路)

[図書] (計0件)

[産業財産権] (計0件)

[その他] (なし)

6. 研究組織

(1) 研究代表者

望月 要 (MOCHIZUKI KANAME)
帝京大学・文学部・教授
研究者番号: 80280543

(2) 研究分担者

中村 直人 (NAKAMURA NAOTO)
千葉工業大学・情報科学部・教授
研究者番号: 20201676
大西 仁 (OHNISHI HITOSHI)
放送大学・ICT活用・遠隔教育センター・
准教授
研究者番号: 40280549