

## 腹話術効果の時間特性

小林まおり<sup>\*1</sup> 藤井真治<sup>\*1,2,3</sup> 岩谷幸雄<sup>\*1,2</sup> 坂本修一<sup>\*1,2</sup> 鈴木陽一<sup>\*1,2</sup>

### Temporal Characteristics of the Ventriloquism Effect

Maori Kobayashi<sup>\*1</sup>, Shinji Fujii<sup>\*1,2,3</sup>, Yukio Iwaya<sup>\*1,2</sup> Shuichi Sakamoto<sup>\*1,2</sup> and Yôiti Suzuki<sup>\*1,2</sup>

**Abstract** - The perception of the spatial location of an auditory stimulus can be captured by a spatially disparate visual stimulus, a phenomenon known as the ventriloquism effect. Many studies have shown temporal and spatial dependency of this phenomenon but its temporal dynamics are not fully understood. In this study, we presented participants with a three-minute of audio-visual stimuli, consisting of a pair of light and sound noise with a spatial disparity of either 0, 5 or 10 degrees. Ten participants were instructed to observe the audio-visual stimuli, and to report their position consistency of stimuli by pushing one of two buttons, which indicated "same" or "different" respectively. The time series reported were analyzed as perceptual transitions. The results showed that (1) the mean total duration of the "same location" response increased as the disparities between noise and light decreased, (2) the percept switched during the stimulus presentation in all disparity condition after initial build-up of perception of "same location", and (3) the mean transition times varied depending on the disparity condition. These results show that the ventriloquism effect varies with observation time, suggesting that audiovisual spatial integration have time-varying nature.

**Keywords:** ventriloquism effect, audio-visual interaction, spatial integration, temporal perceptual change

### 1 はじめに

我々は音像の定位方向が視覚刺激に左右される現象を日常よく経験する。テレビや映画をみているときに音声があたかも映像から出ているように感じるのもその一例である。この現象は腹話術効果 (ventriloquism effect) と呼ばれ[1], 視覚と聴覚の統合問題として研究されてきた[1-3]。腹話術効果の生起条件については詳細に調べられており, 視覚刺激と聴覚刺激間の空間距離が小さく[4-6], 刺激を提示するタイミングの同時性が高いほど生じやすい[4,6]ことが明らかにされている。また, 音声と顔といった意味的に一致する刺激の組み合わせは, 雑音と幾何学的図形といった無関係な刺激の組み合わせよりも腹話術効果が生じやすい[7,8]。

しかし, 多くの研究で使用された刺激は時間長が最大でも数 100 ms と非常に短く, その時間的特性に関する検討は十分とはいえない。実際に腹話術効果が成立している期間は, 刺激間の空間距離に依存する[4]。これは腹話術効果が生じるためにある程度の時間を要する, あるいは一度成立した腹話術効果が消失する可能

性を示す。また, 空間的に不一致の視聴覚刺激を数分間観察(順応)すると, 順応した視聴覚刺激間の空間ずれを知覚しにくくなる腹話術残効が生じる[9-11]ことから, 先行経験によって腹話術効果が変動することがわかる。これらのことから, 視聴覚刺激を長時間呈示し観察し続けた場合には腹話術効果は異なる特性を示すことが容易に想像できる。

もし腹話術効果が時間的に変容するならば, 視聴覚間の空間情報の統合が動的に変化する特性を持つことを示す。上述した研究から, その変容は数分のスパンで生じると考えられる。ゆえに比較的短期間で腹話術効果の時間変動を検討することは, 人間の視聴覚の空間統合のメカニズムを解明するうえで有用な基礎知見となりえる。

加えて, バーチャルリアリティ技術の観点から考えると, 見込角が広い映像ディスプレイや, 多チャンネルスピーカによる音場提示システムが普及しつつあり, 視聴覚間の定位ずれに違和感を持ちやすい状況に変化しつつある[12]。そのため腹話術効果の時間特性が明らかになれば, 視聴覚間の定位ずれの補正に応用できる可能性もある。

上述の視座に立ったうえで, 本研究では腹話術効果の時間特性について報告する。

\*1 東北大学電気通信研究所

\*1 Research Institute of Electrical Communication, Tohoku University

\*2 東北大学大学院情報科学研究科

\*2 Graduate school of Information Science, Tohoku University

\*3 現所属:三菱電機株式会社

\*3 Currently working at: Mitsubishi Electric Corporation

## 2 実験 1

実験では視覚刺激と聴覚刺激を3分間呈示し、時間の経過による腹話術効果の変動を検討した。腹話術効果は運動刺激においても生じる[13-15]ことから、視聴覚刺激として運動刺激を呈示した。その際、視覚刺激と聴覚刺激間の空間的距離を操作した。

### 2.1 方法

#### 2.1.1 被験者

視聴覚ともに健常な成人7名(うち女性1名, 平均年齢24.1歳)が, 同意書に署名をした後, 実験に参加した。全ての実験手続きは東北大学電気通信研究所倫理委員会の承認を受けた。

#### 2.1.2 装置

実験は吸音を施した防音暗室で行った。コンピュータ(Dell Precision T5400, MATLAB 7.1)で刺激の生成, 制御および被験者の反応の取得を行った。

視覚刺激の提示には DLP プロジェクタ(Panasonic, PT-D6000LS, 1280×1224 ピクセル解像度)を用い, 音響透過型スクリーン(2650 mm×1988 mm)に投影した。聴覚刺激はサウンドボード(Synthax, HDSP9652), DAコンバータ(BEHRINGER, ADA800), アンプ(MP3016)を通じ, 円形フレーム(直径 26.8 deg.)上に 22.5 deg. (視角 5.2 deg.)間隔で配置した16台のスピーカ(ホシデン 30φソフトエッジスピーカ)から提示した(図1)。円形フレームの下辺を床から約 50 cm になるよう設置した。観察距離は 150 cm とし, いすに座り両眼で刺激を観察した(図2)。被験者の反応はテンキーを用いて, 160 Hz のサンプリングレートで記録した。

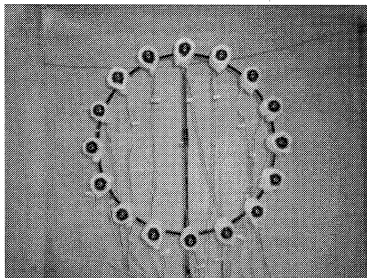


図1 実験で用いた聴覚刺激呈示装置

Fig.1 Presentation system for auditory stimuli in this experiment.

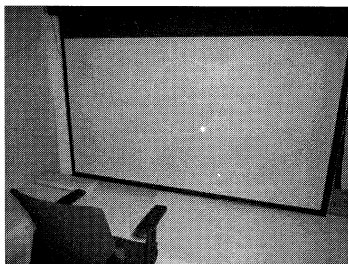


図2 視聴覚刺激提示装置

Fig.2 Audio-visual presentation system

#### 2.1.3 刺激

視覚刺激は, 背景黒の画面に提示した白色ディスク(直径 1.9 deg., 背景とのコントラスト比 95 %)であり, 円(直径 26.8 deg.)を描くように運動した。

聴覚刺激は 1 オクターブ帯域雑音(中心周波数: 1 kHz)とした。各スピーカから提示される聴覚刺激の振幅包絡を連続的に変化させ, 運動音像を作成した(付録A参照)。運動音像自体のA特性音圧レベルは被験者頭部位置で 60 dB に固定した。

視聴覚刺激の時間長は 180 s とし, 運動速度は約 4 deg./s に固定した。この運動速度は予備実験を行った結果, 聴覚による音像定位の精度が高く, かつ腹話術効果が生じることから適切な速度として設定した。

#### 2.1.4 手続き

被験者のキー押しによって1試行を開始した。被験者の課題は, 円中心部の注視点を見つめながら, 聴覚刺激と視覚刺激が同じ位置から提示されているか否かを, 1試行を通じて報告することとした。判断が変化した場合には, できるだけ早くキー押しで報告するよう求めた。実験条件として, 視聴覚刺激の相対位置がそれぞれ 0, 5, 10 deg. 離れた3条件を設定した。各条件につき8試行を行い, 計 32 試行を行った。実験条件の提示順および試行開始時における刺激の提示位置はランダムとし, 先行する刺激のモダリティ(視覚刺激・聴覚刺激)および運動方向(時計回り・反時計回り)については被験者内でカウンターバランスを取った。実験時間は1人につき休憩も含め 120 分程度を要した。

## 2.2 結果

各刺激条件において, 1 試行中に「同位置」と報告した総時間の被験者間平均と標準誤差を図3に示す。視聴覚刺激間の空間距離が近くなるほど, 「同位置」と報告する時間は長くなった( $F(2,12) = 18.07, p < .01$ )。ライアン法による多重比較の結果, 全ての条件間に有意差が認められた。

次に, 1試行を開始してから各反応(同位置・異位置)を最初に判断するまでに要した時間の被験者間平均と

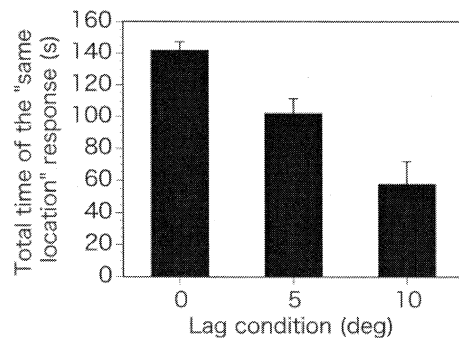


図3 1試行中の「同位置」報告の総時間の平均と標準誤差  
Fig.3 Total duration of "same" location response in a trial

標準誤差を図 4 に示す。刺激間の距離が広がるほど、「同位置」と判断するまでに要する時間が長く、「異位置」と判断する時間は短いことがわかった。反応(同位置, 異位置)と距離条件(0, 5, 10 deg.)を要因とした 2 要因の分散分析を行ったところ, 交互作用が認められた ( $F(2,12) = 10.45, p < .01$ ) が, 各条件の主効果は認められなかった(反応:  $F(1,6) = 0.06, n.s.$ ; 距離条件:  $F(2,12) = 0.6, n.s.$ )。反応についての単純主効果は, 0 deg.条件( $F(1,18) = 9.75, p < .01$ )と 10 deg.条件( $F(1,18) = 7.16, p < .01$ )において認められた。一方, 距離条件の単純主効果も認められた(同位置: $F(2,24) = 5.34, p < .05$ ; 異位置: $F(2,24) = 9.3, p < .01$ )。ライアン法による多重比較の結果, 「異位置」反応時間において 0-10 deg. 条件間 ( $p < .05$ ), および 0-5 deg. 条件間 ( $p < .05$ ) に有意差が認められた。一方, 「同位置」反応時間では, 0-10 deg. 条件間のみ有意差が認められた ( $p < .05$ )。

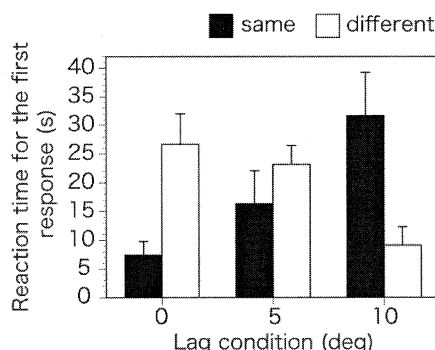


図4 平均初回反応時間と標準誤差

Fig.4 First reaction time for each condition

図 5 に 1 試行中の「同位置」の被験者間平均報告率の推移を示す。0 deg.条件では1試行を通じて 0.8 程度を, 5 deg.条件では 0.5 程度を, 10 deg.条件では 0.3 程度を推移することがわかった。ただし, 10 deg.条件では, 刺激の回転周期に応じた「同位置」報告率の変化が生じることがわかる。

図 6 に, 1 試行中の平均報告数と標準誤差を示す。視覚刺激と聴覚刺激の空間距離が離れるほど, 報告数が増加する傾向が見られた。しかし, 1 要因の分散分析の結果, 空間距離の主効果は認められなかった ( $F(6,34) = 0.83, p = 0.52, n.s.$ )。

次に, 10 s ごとの平均報告数を経過時間に対してプロットしたものを図 7 に示す。この図には, 平均報告-経過時間の近似曲線を 3 次関数で近似した曲線も示してある。近似曲線の決定係数はそれぞれ 0.91, 0.87, 0.81 となった。条件によらず時間経過にしたがって報告数は減少した。特に 0 deg.条件では開始 30 s 後に報告数は急速に減少し, 開始 60 s 前後で 0.2 回程度に安定した。一方, 5, および 10 deg.条件では 20 s 前後で報告数が急速に減少するものの, 120 s 前後まで 0.6 回程度を維持する傾向が見られた。

図 8 に「同位置」と報告されている一回あたりの時間の度数分布を示す。刺激の空間距離によって狭いほど一回あたりに「同位置」を知覚している時間は長くなり, 空間距離が広いほど「同位置」を知覚する時間が短くなることがわかった。

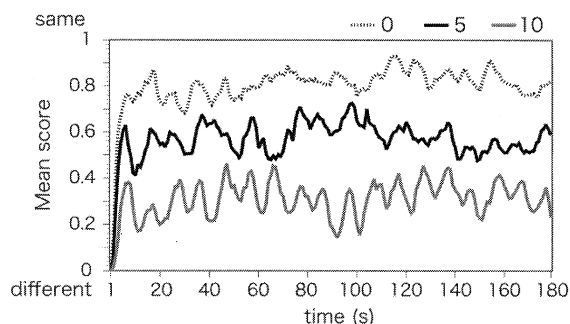


図5 1試行中の「同位置」報告率の推移

Fig.5 Transition of the "same" location rate

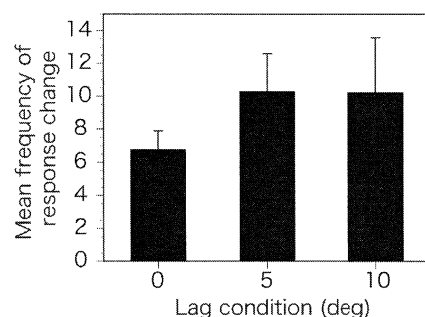


図6 1試行中に報告が反転した平均回数と標準誤差

Fig.6 Mean frequency of response change in a trial

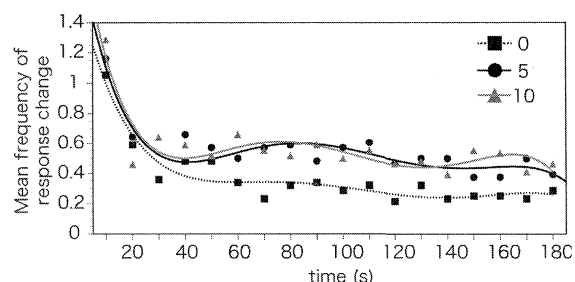


図7 1試行中における 10 秒ごとの報告数の推移

Fig.7 Transition for response frequency of every 10-second in a trial

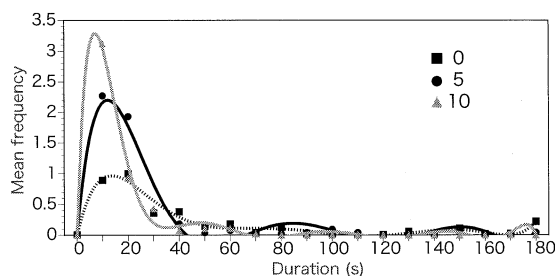


図8 「同位置」知覚時間の度数分布

Fig.8 Frequency distribution of duration of "same" position experience

### 3 考察

先行研究[4]と一致して、視聴覚刺激の空間距離が狭いほど、腹話術効果が成立する期間が長くなることがわかった。加えて刺激間の空間距離に比例して、現象が成立するまでに要する時間は長くなった。通常、視聴覚刺激が8 deg.以上離れてしまうと腹話術効果は成立しにくい[10]とされているが、そのような距離であっても視聴覚刺激を観察し続ければ腹話術効果が生じることが明らかになった。一方、視聴覚刺激を物理的に同位置に呈示した場合でさえも、観察し続けると異位置にあるように知覚することもわかった。ただし、この時間長の効果は、1試行中に複数回の報告があったことから、0および5 deg.条件では「同位置」、10 deg.条件では「異位置」と一度報告した後で知覚が変化した結果と考えられる。さらに全ての条件で1試行中の報告数が5回以上あったことから、腹話術効果は一度成立しても時間経過によって消失、出現を繰り返す動特性を持つことが明らかになった。また、報告数の推移から0 deg.条件では早い段階で「同位置」知覚に収束するのに対して、5, 10 deg.条件では一方の知覚に収束しにくいことがわかった。5 deg.条件では視聴覚刺激が物理的に離れているものの、腹話術効果が生じる距離である。そのため、腹話術効果の生起によって、一致あるいは不一致として知覚しうる多義的な刺激となり、一方の知覚に収束しないと考えられる。

今回の実験結果は、視聴覚間の空間統合が動的に変容する可能性を示す。このような知覚の動的な変化は、単一モダリティの研究において数多く示されてきた[18-20]。典型的な例としては、ネッカーキューブ、ルビンの壺などの多義図形や視覚仮現運動、あるいは両眼視野闘争などがある[17-19]。一方、聴覚事象では音脈分凝や反復単語変形効果などがあげられる[20-21]。最近では、視聴覚の同時性知覚においても知覚の交替(同時・非同時知覚の交替)が報告されている[22]。

こうした知覚反転現象には時間的な動特性の共通点が指摘されている。たとえば両眼視野闘争や図地反転図形ではそれぞれの知覚が経験されている時間の分布はガンマ分布に従う[18,19]。さらに、ネッカーキューブと音脈分凝を用い視聴覚間で動的特性を比較した研究では、2種類の反転現象間で類似した分布が得られることが報告されている[20]。このようにモダリティを超えて時間的動特性が類似することから、知覚の反転現象は、共通した高次のスイッチングメカニズムによって制御されるという説も提唱されている[20,23]。

本研究で得られた腹話術効果が成立している時間の分布(図8)を見てみると度数が小さいために判別は難しいもののガンマ分布に類似しており、単一モダリティにおける多義知覚と共通するメカニズムが関与する可能

性がある。近年、fMRIを用いた研究では、頭頂間溝の活動が多義的知覚経験に関与することが示唆されている[24]。頭頂間溝は視聴覚統合にも関与することが示されており[25]、本実験で観察されたような視聴覚のクロスモダリティにおける知覚現象の動的変化に関与することは十分に予想される。

しかしながら、本研究で示された動的変化が上述したような知覚の動的変化と同じ機序によって生じているかについては不明である。今後、被験者に実際に定位位置を示させるなど測定法を改良し、多角的な検討が必要である。

### 謝辞

本研究は文部科学省特別推進研究(No. 19001004)の助成を受けた。

### 参考文献

- [1] R.B. Welch, D.H. Warren: Immediat perceptual response to intersensory discrepancy; *Psychological Bull*, 88, 638-667 (1980)
- [2] I.P. Howard, W.B. Templeton: *Human Spatial Orientation*. Wiley, New York (1966)
- [3] G.H. Recanzone: Interactions of auditory and visual stimuli in space and time; *Hearing Res*, 258, 89-99 (2009)
- [4] W.R. Thurlow, C.E. Jack: Effect of Degree of visual association and angle of displacement on the "ventriloquism effect": *Percept Mot Skills QT*, 967-979, 1973.
- [5] W.R. Thurlow, C.E. Jack: Effect of Degree of visual association and angle of displacement on the "ventriloquism effect"; *Percept Mot Skills QT*, 967-979, 1973.
- [6] D.A. Slutsky, G.H. Recanzone: Temporal and spatial dependency of the ventriloquism effect; *NeuroReport*, 12, 7-10 (2001)
- [7] R.I. Bermant, R.B. Welch: Effect of degree of separation of visual-auditory stimulus and eye position upon spatial interaction of vision and audition; *Percept Mot Skills*, 42(43), 487-493 (1976)
- [8] M. Radeau: Signal intensity, task context, and auditory-visual interaction; *Perception*, 14(5), 51-57 (1985)
- [9] M. Radeau, P. Bertelson: Adaptation to auditory-visual discordance and ventriloquism in semirealistic situations; *Percept Psychophys*, 23, 343-343 (1977)
- [10] G.H. Recanzone: Rapidly induced auditory plasticity: the ventriloquism aftereffect; *Proc. Natl. Acad. Sci. USA* 95, 869-875 (1998)
- [11] J. LeWald: Rapid adaptation to auditory-visual spatial disparity. *Learn. Mem.* 9, 268-278 (2002)
- [12] 小宮山撰: 大画面テレビ視聴時における音像定位, 日本音響学会学会誌, 43, 9, 664-669 (1987)
- [13] S. Mateef, J. Hohnbein, T. Noack: Dynamic visual capture: apparent auditory motion induced by a moving visual target; *Perception*, 14, 721-727 (1985)
- [14] N. Kitajima, Y. Yamashita: Dynamic capture of sound

- motion by light stimuli moving in three-dimensional space; *Percept Mot Skills*, 89, 1139-1158 (1999)
- [15] S. Soto-Faraco, J. Lyons, M. Gazzaniga, C. Spence, A. Kingstone: The ventriloquism effect in motion: Illusory capture of dynamic information across modalities; *Cogn Brain Res*, 14, 139-146 (2002)
- [16] J.J. Stekelenburg, J. Vroomen: Neural correlates of audiovisual motion capture; *Exp Brain Res*, 198, 383-390 (2009)
- [17] J.M. Hupé, N. Rubin: The dynamics of bi-stable alternation in ambiguous motion displays: a fresh look at plaids. *Vision Res*, 43, 531-548 (2003).
- [18] T. Murata, N. Matsui, S. Miyauchi, Y. Kakita, T. Yanagida: Discrete stochastic process underlying perceptual rivalry; *NeuroReport*, 14, 1347-1352 (2003)
- [19] J.W. Brascamp, R. van Ee, W.R. Wiebe, A.V. van den Berg: Distributions of alternation rate in various forms of bistable perception; *J Vis*, 5, 287-298, (2005)
- [20] D. Pressnitzer, J.M. Hupé: Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization; *Curr Biol*, 16, 1351-1357 (2006)
- [21] R.M. Warren: Auditory perception: An analysis and synthesis; Cambridge University Press (2008)
- [22] O.L. Carter, J.D. Pettigrew: A common oscillator for perceptual rivalries?; *Perception* 32, 295-305 (2003)
- [23] R. Kanai, B.R. Sheth, F.A. Verstraten, S. Shimojo: Dynamic perceptual change in audiovisual simultaneity; *PLoS One*, 12, e1253 (2007)
- [24] R. Cusack: The intraparietal sulcus and perceptual organization; *J Cogn Neurosci*, 17, 641-651 (2005)
- [25] G. A. Calvert: Crossmodal processing in the human brain: insights from functional neuroimaging studies; *Cereb. Cortex* 11, 1110-1123 (2001)

## A 付録

### A.1 音像移動の作成

音像をバーチャルに連続移動させるべく、以下のよう  
に各スピーカから呈示する聴覚刺激の振幅包絡を操作  
した。音像の時間長を  $t$ , 隣り合うスピーカ間を音像が移  
動するまでに要する時間を  $T$  とし, スピーカ  $n$  の位置に  
音像が呈示されるときを  $t=0$  とすると, スピーカ  $n$  の振幅  
 $A_n$ , および隣り合うスピーカ  $n-1, n+1$  の振幅  $A_{n-1}, A_{n+1}$   
を以下のように制御した。

$$A_{n-1}(t) = \cos \left\{ \frac{\pi}{2} \left( \frac{t}{T} - 1 \right) \right\}$$

$$A_n(t) = \cos \left\{ \frac{\pi t}{2T} \right\}$$

$$A_{n+1}(t) = \cos \left\{ \frac{\pi}{2} \left( \frac{t}{T} + 1 \right) \right\}$$

(2010年11月30日)

