

高い音は上？ — 乳児の視線計測からのアプローチ — Eye-tracking Study on Infants' Audiovisual Congruence in “Height”

丸山 慎^{1,2,3}・渡辺はま¹・多賀巖太郎¹

Shin Maruyama, Hama Watanabe, Gentaro Taga

¹ 東京大学, ² 国立情報学研究所, ³ (財) ヤマハ音楽振興会

The University of Tokyo, National Institute of Informatics, Yamaha Music Foundation

shindiana2@yahoo.co.jp

Abstract

Studies on the inter-sensory influence have reported that young infants as well as adults exhibit the perceptual congruence of “height” between sound pitch (hearing) and vertical space (vision). In the present study, we attempted to replicate the findings in previous infant studies but in different (and more precise) ways by introducing an eye-tracking system. In the experiment, we used computer generated movies as stimuli: an original character, “the smiling sun” encircled by flashing lights, moved horizontally accompanying the electric piano sound in an identical pitch (C3) and, after a few seconds, it suddenly disappeared. At that moment, the sound changed to the C major scale played in either direction (ascending: C3 to C4 or descending: C3 to C2) or did not change (identical: C3), and the flashing lights circle appeared at three spatial areas (top, middle and bottom) on the screen. Our logic was that, if infants exhibited the perceptual congruence between the spatial and auditory stimuli, their gaze direction would be biased due to the direction of the sound scale for detecting the missing character. Seventy three 10-month old infants participated to the study. The movies for all sounding patterns were randomly presented. We analyzed the infants' total looking time to the target areas in each direction of the music scales and partially replicated a tendency that the ascending scale induced the infants to watch the highest areas on the screen. This result demonstrates that the audio-visual congruence is a robust phenomenon that occurs *even* when the visual target is *absent*, suggesting that the direction of the change of the sound pitch intrinsically has an influence on the infants' prospective perception of the outcome of the visual event.

Keywords — infant, audio-visual congruence, pitch, spatial height, eye-tracking

1. 目的

数多くの感覚間知覚に関する研究が明らかにしているように、私たちの知覚の対象となる刺激は、単一のモダリティにおいて処理されるだけでなく、複数の異なる感覚モダリティ間に“冗長な”

仕方で影響を与える。本研究は、このような多感覚間知覚のなかでもとりわけ空間の高低の知覚（＝視覚）と音声のピッチの知覚（＝聴覚）との対応関係に焦点を当てる。空間的に「高い」ということと音声のピッチが「高い」ということの対応は、単なる言語的な表記やメタファーとしての一一致ではなく、私たちの知覚的経験として実証されている [1]。さらに近年では、3–4 か月齢の乳児が視聴覚間で「高さ」を対応させていることが確認されている [2]。ピッチの高低が変化する音声に合わせて上下方向に移動するボールの動画刺激を乳児に見せ、ボールが上から下に移動する際に、音声も高いピッチから低いピッチへと移行する場合（＝視聴覚間の一致）と、ボールは上から下に移動しているのにピッチは低から高へと移行する場合（＝視聴覚間の不一致）のどちらが選好されるかを検討し、乳児が前者（ボールの動きと音声のピッチの上下行が一致している刺激）の方により注意を向けていたことを示したのである。

このように視聴覚間での一致という異なる感覚モダリティ間の結びつきは、人間の発達初期から認められるものである。しかし、果たしてそれはどの程度“頑強な”関係といえるのだろうか。例えばそれは、一方のモダリティからの入力突然消失してしまったような場合、刺激の入力が消失したモダリティの知覚内容が、他方のモダリティからの入力によって影響を受けるようなことはあるのだろうか。そして、そのようなある種の“予見的な”知覚を導く視聴覚間の協応は、乳児においても確認できるのだろうか。以上の問題にアプローチするために、本研究では、音声を伴って動くキャラクターが刺激の提示中に消失し、その直

後から音声のピッチの高低が変化し始めるという動画を使用して、視覚刺激としては不在となったキャラクターの行方を追いかける乳児の視覚的な探索のパターンが、音声のピッチの変化（すなわち聴覚刺激の知覚）からの影響を受けるのかどうかを検討した。また本研究では、キャラクターの探索に関わる乳児の「見る」反応をより厳密に捉えるために、視線計測装置を用いた測定を行い、人間の発達初期における視聴覚間の協応について眼球運動の水準から検討した。

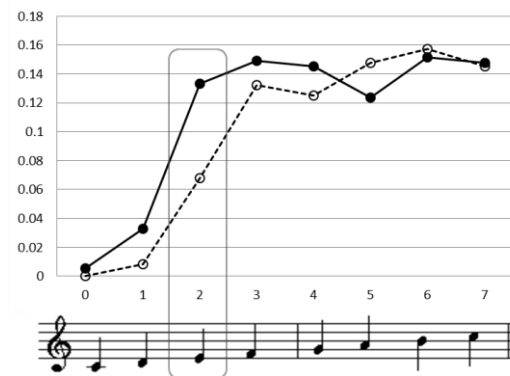
2. 方法

研究協力者は10か月齢の乳児73名とその保護者であった。計測は東京大学大学院教育学研究科発達脳科学研究所の防音ブース内にて行われた。研究協力者には研究内容および計測装置について説明し、参加同意書への署名を確認の上、計測を開始した。本研究では視線計測装置 Tobii T-60（トビー・テクノロジー・ジャパン：サンプリング・レートは60Hz）のモニタ上に独自に作成した動画刺激および注意を喚起するための誘因刺激を投影した。乳児は保護者の膝の上に座った状態でそれらを視聴した。動画刺激は以下の3つの構造を持つ：(1) 架空のキャラクター（点滅する日輪で囲まれた笑顔の“お日さま”）が登場し、1秒間に1音のペースで再生されるピアノの「ド（C3：ハ長調）」の音に合わせて、モニタの左右いずれかの端から中央に向かって水平方向に移動する(2) 中央に差し掛かった時点で、そのキャラクターが突然消失し、それと同時にピアノの音でハ長調の上行音階（ドレミファソラシド）、下行音階（ドシラソファミレド）あるいは変化しない一定音（「ド」のまま）のいずれかのパターンが再生される(3) このときモニタの上部、中部、下部の3か所にキャラクター本体を囲んでいた点滅する日輪だけが目印として現れる。各乳児に3パターンの音階の動画をランダムに提示した。もし乳児が音高と空間の高低とを対応づけた知覚をしているのだとすれば、消失したキャラクターを探索する彼らの視線の方向が、再生される音階の方向（上行、下行）

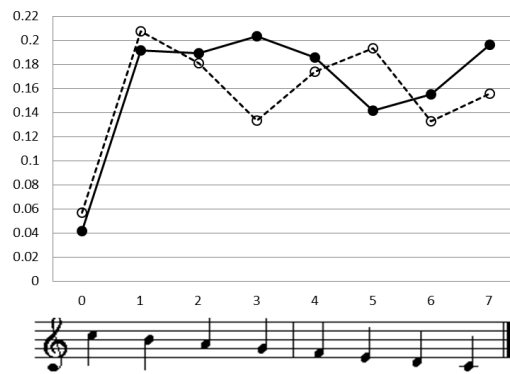
に影響を受けるはずである。つまり上行する音階が再生された場合には、乳児の視線はモニタの上部に向けられやすいのではないかと考えている。この動画1回の再生時間は約15秒であり、誘因刺激を含めた全ての刺激を視聴した場合に要する時間は約7分であった。

3. 結果

キャラクターが消失する直前の1秒間のうちの80%以上（0.8秒以上）キャラクターを追視していたことが確認でき、かつこの基準をクリアした試行が音階の各方向につき2試行以上あった乳児31名分のデータを分析対象とした。



a. 「上部」への注視時間の割合



b. 「下部」への注視時間の割合

図1 動画刺激の指定領域への注視時間の割合
(音響刺激の各音ごとに算出)

キャラクターが消失して音階のパターンが再生され始めた直後に動画刺激の上部、中部、下部に現れる日輪の周囲を重複する領域がないように区切り、その3つの領域のいずれかを注視していた時間の割合を有効試行数に応じて算出した。この

注視時間の割合を、音階を構成する1音ごとにまとめたものが図1である。同a.は動画刺激の「上部」の領域、同b.は「下部」の領域への注視時間の割合であり、それぞれに上行音階を提示した試行（実線）と下行音階を提示した試行（破線）の結果を示した。

各音ごとに音階の方向（上行か下行か）を要因として対応のあるt検定を行ったところ、上行音階が再生され始めて間もない時点（ド、レ、ミ、ファ、ソ、ラ、シ、ドのうち「ミ」の音が再生された時点：図1a.の四角で囲った2の部分）において、動画刺激の「上部」領域への注視が、下行音階が再生された場合よりも有意に長かった（ $t(30)=2.563$, $p=.016$ ）。しかし「ミ」が再生された直後、「ファ」が再生された以降は、音階の方向による上部への注視時間に差は見られなくなった。また下部への注視に関しては（図1b.）、音階が上行あるいは下行のどちらの方向に変化しても、各音ごとの注視時間の割合には差が見られなかった。

4. 考察

動画刺激の中部から下部にかけての領域への注視は、上下行いずれの音階の方向にも影響されることなく全般的に大きな割合を占めていた（下部の領域に関しては図1b.に示した結果から明らかである）。各領域における1音ごとの注視パターンの変化を比較してみると、中部および下部についてはピッチが変化し始めた瞬間から（ピッチの変化の方向に関係なく）即座に注視が向けられる傾向が表れており、中部から下の領域は10か月齢の乳児にとって注視することが比較的容易な領域であったことが推測される。

これに対して上部は、注視時間の割合が他の領域に比して小さかったことから、最も注視しにくい領域であったと考えられる。実際、乳児の注視能力の発達に関しては、水平方向に比して、垂直方向への注視の発達がやや遅れるという報告もあり[3]、本研究の結果はそれと一致する傾向を示している。それにも関わらず、音階の再生過程において、上部への探索的な注視時間が、音階の上

行下行によって変化していたという結果は重要である。すなわち上部への注視は、上行音階が再生された場合には2番目の音「ミ」への移行時から生じていたのに対し、下行音階が再生された場合には、それより後の3番目の音「ファ」への移行時から生じていた（図1）。上行音階を提示した方が、下行音階を提示した場合に比べて“より早い”段階で上部への探索の割合が増加していたのである。これは、視覚的に不在となったキャラクターの移動先に関する乳児の空間的な探索行動が、上行音階、すなわちピッチが次第に高く上がっていくという聴覚刺激により強く影響を受け、視線が上部の空間に一瞬早く誘因されたということを反映しているのではないだろうか。ところが、この上行音階による上部への注視の影響は、たった1音の間でしか確認されなかった。つまり2番目の「ミ」の音が再生された時点での上部への注視の先行性が、続く「ファ」の音では消滅し、上部と下部への注視時間の間に差が見られなくなったのである（図1a.）。このような反応から示唆されることは、乳児たちが本実験の課題において、「音が高くなっているから上方の空間ではないか？」といった、いわば音高と空間との“意味的な”対応による処理を行っていた可能性は低いということである。なぜなら、仮に乳児が音高と空間とを意味的な水準で関連付けていたならば、音が上行し続けている間は、少なくとも上部への注視がもっと持続的であってもよいはずだと考えられるからである。したがって上部への反応の生起と消滅が、きわめて短い時間のなかで起こっていたということは、むしろ視聴覚間での刺激の処理が、上行音階における音高の変化に内在する運動の方向性によって、乳児の視線が上部の空間へと“期せずして”導かれたからであると考えた方が適切ではないだろうか。

乳児の視聴覚間協応は、視覚的に不在となったキャラクターの行方を探索するという課題においても生起した。このような本研究の結果は、音高と空間知覚における視聴覚間の協応が、「見えるものと聴こえる音」との一致だけではなく、消失し

たキャラクターの行方を知覚的に誘導し、その行方を予見させる、すなわち「“見えないもの”と聴こえる音」の間にも生じる、頑強な事実であるということを示しているのである。

参考文献

- [1] Evans, K., & Treisman, A. (2010) Natural cross-modal mappings between visual and auditory features. *Journal of Vision*. 10 (1): 6, 1-12.
- [2] Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010) Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science*, 21 (1), 21-25.
- [3] Gredebäck, G., von Hofsten, C., Karlsson, J., & Aus, K. (2005) The development of two-dimensional tracking: a longitudinal study of circular pursuit. *Experimental Brain Research*, 163: 204-213.