



聴触覚クロスモーダル効果に対する音像定位の影響

The Effect of Sound Source Localization on Auditory-tactile Cross-modality

金 祇¹, 伊東 健一¹, 伴 祐樹¹, 割澤伸一¹

Yi JIN, Kenichi ITO, Yuki BAN, and Shin'ichi WARISAWA

1) 東京大学 新領域創成科学研究科 (〒 277-8563 千葉県柏市柏の葉 5-1-5, jin.yi@s.h.k.u-tokyo.ac.jp, kenichiito@s.h.k.u-tokyo.ac.jp, ban@edu.k.u-tokyo.ac.jp, warisawa@edu.k.u-tokyo.ac.jp)

概要: 聴触覚クロスモーダル効果は今までに多くの検証がなされており, 例えば表面の粗さ知覚が聴覚刺激により変化することが知られている. 音像定位は重要な聴覚知覚だが, テクスチャ知覚に対する聴触覚クロスモーダル効果への音像定位の影響は解明されていない. 音像を頭内または頭外定位する条件で, やすりを棒でなぞる際の粗さ知覚について検証した結果, 頭外定位によって粗さを強く知覚する効果が高まることが明らかとなった.

キーワード: 聴触覚クロスモーダル, 音像定位, 触覚, テクスチャ

1. はじめに

人間は五感やその他の感覚を用いて外界や身体表面の物理情報を取得し, 周囲の世界を認識している. 脳は, 複数の感覚から得られた情報を統合することで, 知覚の信頼性を高めている [1]. 統合された感覚情報間の相互作用により知覚が変化する現象はクロスモーダル効果と呼ばれ, 様々な感覚提示インタフェースに応用されてきた [2].

例えば, 聴覚刺激によって触覚知覚が変化する聴触覚クロスモーダル効果の存在が知られている. Jousmäki と Hari [3] は, 手を擦るときの音の周波数や音圧を操作することで皮膚の粗さ知覚が変化する Parchment skin illusion を発見した. また, 粗い表面である紙やすりに触れた際の粗さ知覚も, 聴覚刺激の周波数によって変化する [4, 5]. 具体的には, 高い周波数の音や不協和音が提示されたときに表面はより粗く感じられる.

しかし, これらの研究の多くは単にヘッドホンで音を提示しており, 音の空間性を考慮した研究は少ない. 人は, 音源の空間位置を知覚できる音像定位が可能なものの, ヘッドホンから単純に音を再生するのみでは方向や距離を感じられない. このように音像定位ができない条件では, 音像が頭の中にもったように感じられるため, 頭内定位が生じる [6]. 再生音像を頭の外側に感じ, 方向感や距離感も知覚できる頭外定位を実現するため, 頭部伝達関数 (Head-Related Transfer Function; HRTF) を用いた立体音響技術はバーチャルリアリティ (VR) やゲームに応用されている [6].

刺激の空間位置がクロスモーダル効果に及ぼす影響に関する既存の研究では, 空間的な応答を求めるタスクに限って影響が生じることが示されてきた [7]. Ro らは, 左右の手に対する電気刺激の方向を回答する実験において, 同一方向の聴覚刺激が識別感度を有意に高めることを示した [8]. 一方, Keetels らは時間的にずれた触覚刺激が視覚刺激の順

序知覚に影響する時間的腹話術効果について, 視覚と触覚の刺激が空間的に不一致であっても効果は不変であると報告した [9].

これらの研究は刺激の空間配置による知覚感度や反応速度への影響に関するものである一方, 粗さのようなテクスチャ知覚 [10] への影響については十分に検証されていない. 刺激位置とクロスモーダル効果に関する既存研究 [7] から考えれば, 空間的な応答ではないテクスチャ知覚への聴触覚クロスモーダル効果は音像位置に影響されないと推定される. 一方で, 一般的には音像の位置の影響を受けないとされているマガー効果の強さが, 刺激位置に対する注意の程度によって変調するとの報告もある [11]. テクスチャ知覚が変化するクロスモーダル効果 [3, 4, 5] において聴覚と触覚の刺激が空間的に一致していれば, 刺激の位置に注意が向くことで音像定位がクロスモーダル効果に影響する可能性があると考えた.

音像定位の聴触覚クロスモーダル効果への影響の有無の解明は, クロスモーダル効果を利用した VR システムにおける立体音響技術の必要性を判断するのに重要である. そこで本研究では, 音像定位の可否によるテクスチャ知覚に対する聴触覚クロスモーダル効果の強さを比較した. 検証の対象とするテクスチャ知覚は, 聴触覚クロスモーダル効果が既に明らかにされている表面の粗さ知覚 [4, 5] とした.

2. 実験方法

2.1 実験参加者

本実験には, 実験の目的を知らない学生 12 名 (男性 9 名, 女性 3 名, 平均年齢 $23.6 \pm 2.23SD$) が参加した. 参加者は正常な視力と聴力を有した. 参加者のうち 1 人は左利き, 残りの全員は右利きだった. 参加者は全員利き手で触覚刺激を体験した.

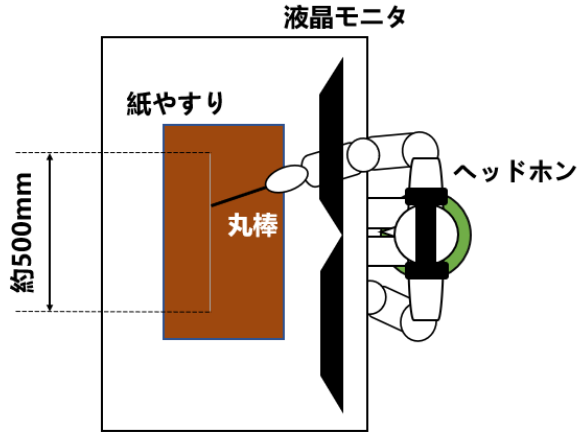


図 1: 実験装置の構成と実験参加者の位置

2.2 実験に用いる刺激の作成

聴覚刺激: 実験では、6種類の聴覚刺激を提示した。粗さ知覚を変化させるため、粒度の異なる3種類の紙やすり(粗目:60番, 中目:120番, 細目:400番)を棒でなぞる音を録音して用いた。中目の紙やすりの粒度は、聴覚クロスモーダル効果による粗さ知覚の変化に関する先行研究[5]で用いられた中間粒度である120番を採用した。明確に粗さの異なる紙やすりから聴覚刺激を作成するため、粗い紙やすりは一般的に粗目に分類される60番を、細かい紙やすりは一般的に細目に分類される400番を使用した。

実験において音像定位を可能とするため、立体的な音場を記録できるアンビソニックマイク(ZOOM H3-VR)を用いて、実験参加者が着座した際の両耳の高さに近い約1150mmで録音した。録音した立体音像および、録音音源をモノラル音に変換したものを、それぞれ頭外定位される音、頭内定位される音として使用した。音声の加工はAdobe Auditionで行われた。実験中、アンビソニック音はValve社によるUnityの立体音響プラグイン「Steam® Audio [12]」で再生した。

触覚刺激: 接触条件を統制するため、棒でやすりをなぞることで、参加者に触覚刺激を体験させた。ステンレス製の丸棒と中目の紙やすりを実験に使用した。紙やすりのサイズは280mm×230mmで、二枚横に並べて長さ560mmの状態を実験した。丸棒の直径は4mm、長さは200mmであった。

2.3 実験装置

実験装置の構成と配置を図1に示す。聴覚刺激は、ヘッドホン(Sony WH-1000XM4)とオーディオインターフェイス(Creative Sound Blaster X4)を用いて提示した。立体音像を再現するために用いるHRTFには、頭部や耳介の形状に起因する個人差があり、他人のHRTFを用いた場合音像定位の誤りが増加する[6]。従って、Sound Blaster X4のSuper X-Fi機能[13]により、参加者個人のHRTFを推定し、音像定位の誤りを軽減した。



図 2: 実験風景

紙やすりの表面をなぞる速度と時間を統制するため、一定速度で動く視覚刺激(黒い円)を図2のように液晶モニターへ表示した。視覚刺激を提示するとともに紙やすりを参加者から隠すため、27インチのモニター2枚を横に並べて使用した。視覚と触覚刺激を提示する実験プログラムはUnityにより作成された。

2.4 実験設計および手順

本実験では、聴覚刺激の提示下で紙やすりをなぞり、粗さを評価する課題を実験参加者に行わせた。

聴覚刺激は、音源の種類と音像定位条件の2要因に基づく6条件(粗目頭内, 粗目頭外, 中目頭内, 中目頭外, 細目頭内, 細目頭外)とした。各条件を3回ずつ繰り返し、1人の実験参加者に対して18回の聴覚刺激を提示した。条件の提示順序を参加者間で均等にするため、提示順序はラテン方格法(Balanced Latin Square) [14]で作成した。

実験を始める前に、Super X-Fiにより、立体音像生成に用いるHRTFを個人に最適化した。個人最適化は参加者の両耳と真正面の写真をスマートフォンで撮影し、専用アプリによってHRTFプロファイルを生成することで行った。次に、頭内定位と頭外定位の概念を実験参加者に理解させるため、棒で紙やすりをなぞる立体音像とモノラル音を実際に聴取させた。

実験中の様子を図2に示す。実験では、参加者は椅子に座り、ディスプレイの下から裏まで腕を伸ばし、棒で中目の紙やすりをなぞった。最初に、ヘッドホンなしで中目の紙やすりをなぞる際に感じた粗さを標準刺激として提示した。

聴覚刺激はヘッドホンを用い、50dB-70dBの音量で提示した。視覚情報の干渉を防ぐため、モニターで視線を遮ることで参加者から紙やすりをなぞっている様子が見えないようにした。参加者はすべての指示を実験プログラムから受けた。聴覚刺激が始まると同時に、ディスプレイ上の黒い円が一定速度で左から右に移動した。参加者は、円の位置とスピードに合わせて紙やすりの表面をなぞり、聴覚刺激と円の移動が停止すると同時に手を止めるよう指示された。円の移動距離は500mmで、速度は100mm/sであった。

紙やすりの実際の粗さを参加者が知ることでクロスモーダル効果が弱まることを防ぐため、実験実施者は1試行ごとに紙やすりを交換する演技を行った。ただし、実際に提

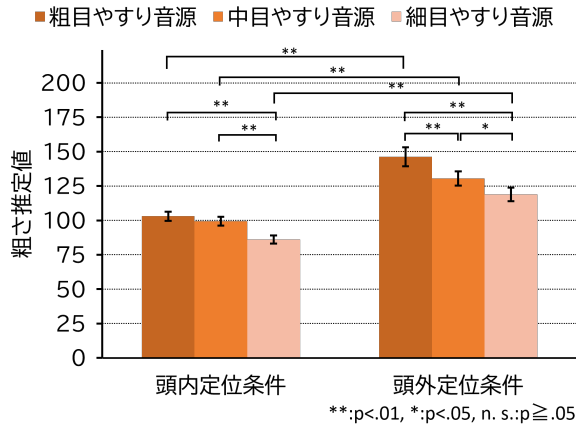


図 3: 粗さ知覚のマグニチュード推定値回答。標準刺激である中目やすりの粗さを 100 とした。誤差棒は標準誤差。

示する触覚刺激はすべて中目のものに固定したため、交換した紙やすりはすべて中目のものであった。

1 試行ごとに、参加者はマグニチュード推定法によってやすり表面の粗さを評価した。実験前に提示した標準刺激の粗さを評価基準値の 100 として、より粗いと感じた場合は 100 より大きい数値で、なめらかと感じた場合は 100 より小さい数値で回答した。また、音像定位について、頭内定位と頭外定位のいずれに感じたか、音像が移動しているように感じたか回答した。

3. 結果・議論

3.1 実験結果

実験における粗さのマグニチュード推定結果の平均値を図 3 に示す。粗さの推定値の条件による違いを確かめるために、2 要因の分散分析を行った。なお、本稿における有意水準は特に断らない限り 5% (0.05) とする。まず、条件ごとに正規性を Shapiro-Wilk 検定にて確認したところ、6 条件中 5 条件で正規分布と有意な差があった。従って、正規性のないデータに分散分析を適用可能な手法である整列ランク変換 (Aligned Rank Transform; ART [15]) を使用した。

ART を用いた二元配置反復測定分散分析の結果、音像定位の主効果は有意であり ($F(1, 199) = 152, p = 2.22 \times 10^{-16}$)、音源種類の主効果は有意でなかった ($F(2, 199) = 2.16, p = 0.118$)。また、2 要因の交互作用は有意だった ($F(2, 199) = 15.2, p = 7.36 \times 10^{-7}$)。従って、2 要因の各水準ごとに Wilcoxon の符号順位検定にて条件間を比較した。多重比較の多重性の補正には Hommel 法を用いた。検定の結果を図 3 に示す。音像定位の条件間では、全ての音源種類で、頭外定位条件での粗さ推定値が頭内定位条件のそれより有意に大きかった。音源種類の条件間では、粗いやすりの音源であるほど粗さ推定値の平均が大きかった。音像が頭内定位の条件では中目音源と細目音源、粗目音源と細目音源の間に、音像が頭外定位の条件では全ての音源間に有意な差があった。

次に、音像定位に関する質問の回答データについて、頭

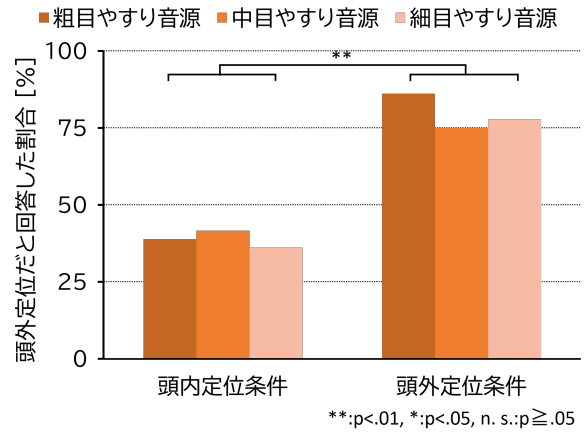


図 4: 音像を頭外定位したとの回答の割合

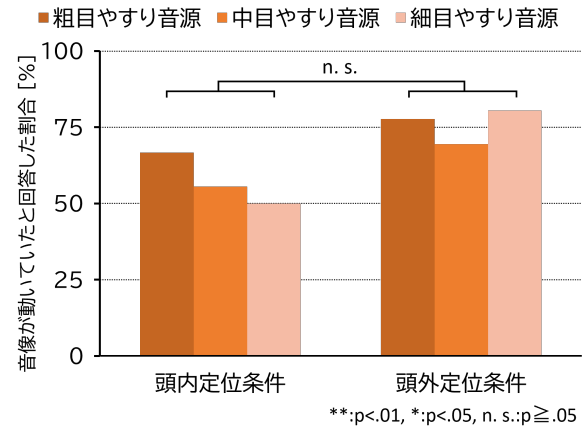


図 5: 音像が動いていたとの回答の割合

内定位条件と頭外定位条件を統計検定で比較した。これらの回答は二値データであるため、適合度に関するカイ二乗検定を実施した。図 4 に示すように、頭外定位の音像条件で頭外定位と回答された回数は、頭内定位の音像条件でのそれより有意に多かった ($\chi^2(1) = 15.1, p = 1.00 \times 10^{-4}$)。一方で、音像の移動感に関する質問では、図 5 に示すように、条件間に有意差はみられなかった ($\chi^2(1) = 2.78, p = 0.0956$)。

3.2 議論

図 3 から、音像定位の条件によって知覚される粗さの変化の大きさが異なることが分かる。検定結果から、頭内定位条件では粗目と中目のやすり音源に有意な差がない一方、頭外定位条件では粗目やすりの音源を聴いたとき有意に表面が粗く感じられることが示された。従って、頭外に定位される音像がテクスチャ表面を粗く感じさせる方向のクロスモーダル効果を強めることが示唆された。一方で、中目と細目のやすり音源の間にはいずれの音像条件でも有意な差が見られ、また粗さ知覚の平均値の比も同程度であった。そのため、表面の粗さを弱く感じる方向のクロスモーダル効果の強度に差があるとは言えなかった。

頭外定位条件の音像では、全体として頭内定位条件の音像より表面が粗く感じられた。その理由として、標準刺激

を提示するときにはヘッドホンなしでやすりをなぞらせたため、ヘッドホンでの音像提示によりなぞる音が強調された可能性が考えられる。また、触覚刺激と一致する方向の聴覚刺激が触覚の感度を高めるという報告 [8] から、指先に音像が定位される頭外定位条件で粗さに対する感度が高まった可能性もある。

図 4 に示した通り、頭外定位条件で「頭外定位」とした回答は頭内定位条件より有意に多かった。これは、本実験で意図したとおりに音像が頭外定位されており、音像定位の効果を検証する実験が成立していたことを示す。しかし、音像の移動感に対しては、図 5 の通り頭内定位条件と頭外定位条件で有意な差がみられなかった。これは、腕の動きと映像の動きに起因する視覚・聴覚・深部感覚の間のクロスモーダル効果によって、頭内定位であっても音像の位置の知覚が変化したことが原因として考えられる。

実験参加者の一部から実験後の感想について、「紙やすりをなぞる最中に、頭内定位か頭外定位かを考える余裕がなかった」「もし頭内定位と頭外定位を回答することを事前に知らなければ、そもそも音像の定位は気にしない(ほど、音像定位を意識することが難しかった)」との回答があった。それらの感想から、体験者が音像定位を意識しなくとも、音像定位の有無による聴覚クロスモーダル効果への影響は生じることが示唆された。

本実験は、視覚刺激を利用して指先を誘導していたため、音像条件に関わらず、視覚刺激が指先への注意を促した可能性がある。今後の実験では、視覚刺激ではなく、アクチュエータを用いて手の動きの速度を統制することを検討する。

4. おわりに

本稿では、音像定位の可否による聴覚クロスモーダル効果への影響の有無を検証するために、音像定位が可能または不可能な条件で、紙やすりの表面の粗さ知覚の変化を調査した。検証の結果、聴覚クロスモーダル効果に対する音像定位の影響が存在していることが示唆された。加えて、頭外定位の音像が、粗さ知覚を増強する方向のクロスモーダル効果を高める現象が見られた。これにより、テクスチャ知覚を変化させるクロスモーダル効果を活用した VR システムでは、音像定位を可能にする立体音響技術を活用することが有効であると考えられる。今後、テクスチャ知覚のみならず、音像定位が他のクロスモーダル効果についても影響を及ぼすかについて検証を進める必要がある。

謝辞 本研究の一部は、科研費 (22K19805, 21H03478) の助成を受けたものである。

参考文献

[1] Shinsuke Shimojo and Ladan Shams. Sensory modalities are not separate modalities: plasticity and interactions. *Current Opinion in Neurobiology*, Vol. 11, No. 4, pp. 505–509, 2001.

[2] 鳴海拓志. クロスモーダル知覚のインタフェース応用. 映像情報メディア学会誌, Vol. 72, No. 1, pp. 1–7, 2018.

[3] Veikko Jousmäki and Riitta Hari. Parchment-skin illusion: sound-biased touch. *Current biology*, Vol. 8, No. 6, pp. R190–R191, 1998.

[4] Steve Guest, Caroline Catmur, Donna Lloyd, and Charles Spence. Audiotactile interactions in roughness perception. *Exp Brain Res.*, Vol. 146, No. 2, pp. 161–171, July 2002.

[5] Didem Katircilar and Funda Yildirim. Harmonicity of sound alters roughness perception. *PsyArXiv*, Jan 2022.

[6] Kazuhiro Iida. *Head-Related Transfer Function and Acoustic Virtual Reality*. Springer, 2019.

[7] Charles Spence. Just how important is spatial coincidence to multisensory integration? evaluating the spatial rule. *Annals of the New York Academy of Sciences*, Vol. 1296, No. 1, pp. 31–49, 2013.

[8] Tony Ro, Johanan Hsu, Nafi E. Yasar, L. Caitlin Elmore, and Michael S. Beauchamp. Sound enhances touch perception. *Experimental Brain Research*, Vol. 195, No. 1, pp. 135–143, May 2009.

[9] Mirjam Keetels and Jean Vroomen. Tactile—visual temporal ventriloquism: No effect of spatial disparity. *Perception & Psychophysics*, Vol. 70, No. 5, pp. 765–771, Jul 2008.

[10] Shogo Okamoto, Hikaru Nagano, and Yoji Yamada. Psychophysical dimensions of tactile perception of textures. *IEEE Transactions on Haptics*, Vol. 6, No. 1, pp. 81–93, 2013.

[11] Riikka Möttönen, Kaisa Tiippana, Mikko Sams, and Hanna Puharinen. Sound location can influence audiovisual speech perception when spatial attention is manipulated. *Seeing and Perceiving*, Vol. 24, No. 1, pp. 67–90, 2011.

[12] Steam[®] audio. <https://valvesoftware.github.io/steam-audio>, (参照 2022 年 7 月 18 日) .

[13] Super X-Fi. <https://jp.creative.com/sxfi/>, (参照 2022 年 7 月 19 日) .

[14] James V. Bradley. Complete counterbalancing of immediate sequential effects in a latin square design. *Journal of the American Statistical Association*, Vol. 53, No. 282, pp. 525–528, 1958.

[15] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. *CHI '11*, pp. 143–146. ACM, 2011.