

バーチャル音空間の評価法*

坂 本 修 —*1

1. はじめに

我々は周囲の環境を把握する際には同時並列的に入力される様々な感覚情報を活用している。その中でも音情報は視覚情報と並んで極めて重要な感覚情報の一つである。これまでの様々な研究をとおして、人間がどのように外界音空間を知覚、認識しているのが明らかとなり、その知見に基づいた音空間収音・再生技術の開発も進められてきた¹⁾。

この時間と空間を超えてある音環境を別の場所にバーチャルに再現する技術は、主にエンターテインメント、アミューズメントをターゲットとして開発されてきたが、システムを構築する際の個々の構成要素の小型化、低廉化等をうけて容易にシステム構築が可能となったこともあり、一般的な場面にも活用されるようになりつつある。ただし、再現される音空間のリアリティは未だ十分ではなく、さらなる技術の高度化が必要である。例えば、昨今の社会情勢を受けてオンラインでの会議実施が増えているが、実際の face-to-face でのコミュニケーションに到底及ばないことは実感していることであろう。離れた環境にいる複数人の協働作業を高い次元で支援するコミュニケーションシステムや、近年研究が盛んに行われている音環境の可聴化 (auralization) 技術を活用した高精度音環境可聴化システムの実現に寄与するような技術の登場が待たれる。

さて、ここで「リアリティ」という言葉を使用した。我々がバーチャル音空間（ここでは様々なシステム、技術を用いてバーチャルに再現される音空間をバーチャル音空間と記す）の「リアリティ」とはどのようなものだろうか？ 我々は多くの場面で「リアリティ」「臨場感」といった言葉でバーチャル音空間（に限らず様々なバーチャル環境）を評価している。しかし、いずれの語も多義的な感性表現であり、定量的に性能を評価する際に使用する語とし

ては適切とは言いがたい。

本稿では、様々なバーチャル音空間再生技術によって創出される音空間の善し悪しをどのように評価するのかを考えるにあたり、まず、人間がどのようにバーチャル音空間を表現し、知覚しているのかについて述べる。それらの知見を踏まえた上で、これまでどのようにバーチャル音空間を評価してきたかについて概観する。

2. 高感性バーチャル音空間の印象表現

まず始めに、バーチャル音空間から受ける印象の表現について述べていきたい。我々がバーチャル環境を体験した際に抱く印象を表出する際に用いる感性表現は多岐にわたる。最も一般的なものの臨場感 (sense of presence) であろう。臨場感は「現在いる場所とは異なる場所にいるような感覚」や「その対象・環境を提供するメディアに対する気づきを失っている状態」²⁾といった意味で用いられてきたが、単にバーチャルな環境を表現するだけに止まらず、「その場で実際に体験しているような」感覚を超え、「好感」、「迫力感」、「動感」、「メカニック感」といった印象を持ち合わせた実際の多感覚事象でも用いられることが明らかとなっている^{2),3)}。

一方で、臨場感が「場」や「空間」といった背景情報に対して想起される感性表現であるのに対し、前景となる「もの」や「対象」に対しての感性表現も存在する。これは「本物らしさ」を示す感性表現であり、迫真性 (sense of verisimilitude)⁴⁾として定義される。また、臨場感という表現を、「場」や「空間」を対象とした「フィールド臨場感」と、「もの」を対象とした「オブジェクト臨場感」に分けて表現するといった場合もある。また、「リアリティ」や「超臨場感」⁵⁾といった語でも表現される。

バーチャル音空間からの印象も同様であり、「(フィールド・オブジェクト) 臨場感」「迫真性」などといった感性表現を用いて表現することが多い。ただし、先にも触れたとおり「臨場感」は心を揺さ

* How to Evaluate the Quality of Virtual Auditory Space?

*1 Shuichi Sakamoto : Tohoku University (東北大学)

ぶる事象に対して用いられる表現であることから、提示されるコンテンツの違いによって「臨場感」が異なることが考えられる。例えば、ライブやコンサートといった非常に心を揺さぶるコンテンツでは、バーチャル音空間の評価にコンテンツから想起される心を揺さぶる情動が影響し、本来求めたい評価が得られない可能性がある。これら考慮して「システム臨場感」と「コンテンツ臨場感」に区別することがある⁶⁾。このような評価を安定的に行うためには評価用標準コンテンツが不可欠であるが、フリーで利用可能な立体音響コンテンツがいくつか提供されている⁷⁾ものの、まだ一般的ではない。

3. バーチャル音空間の知覚

バーチャル音空間の評価をするにあたり、我々がどのように音環境を把握、認識しているのか、その知覚過程を踏まえることは重要である。ここでは、方向、距離、場のそれぞれについて、音空間の知覚手がかりについて簡単にまとめる。

3.1 方向知覚

本特集でも触れられているとおり、方向知覚の手がかり(両耳間レベル差、両耳間時間差・位相差、スペクトラルキュー)を包含している頭部伝達関数(head-related transfer function: HRTF)が音の方向知覚に極めて大きな役割を果たしている。特に正中面の音像知覚については、頭部伝達関数のスペクトル形状の寄与が大きく、スペクトル形状のどの部分に手がかりが含まれているのかといった研究が現在行われている⁸⁾⁻¹⁰⁾。

3.2 距離知覚

距離知覚においても頭部伝達関数は一定の寄与を果たす。ただし、頭部伝達関数の距離依存性は概ね1mまでであり、それよりも遠方では、逆二乗則に基づく音圧レベルの変化や反射音といった他の手がかりが支配的となる。

なお、距離の知覚には、聴取者から対象物までの距離の他に、対象物が背景からどれくらい離れているかという奥行きに類する距離の知覚も含まれ¹¹⁾、それぞれについての知覚様相が調べられている。

3.3 場の知覚

方向、距離に比べ、場自身の知覚を直接明らかにした研究はそれほど多くなく、未解明な部分が多い。しかし、建築音響などの分野ではよいホールの設計といった観点で研究がなされており、音の広がり感の知覚に関しては「みかけの音源の幅(apparent source width, auditory source width: ASW)」や「音

に包まれた感じ(listener environment: LEV)」を中心に盛んに研究が行われており、様々な物理的パラメータとの対応関係も報告されている。

3.4 多感覚情報の手がかり

先にも述べたように我々は同時並列的に入力される様々な感覚情報を統合処理して外界を知覚、認識している。音空間知覚に関しても例外ではなく、音を聴く際には音源が見えている場合も多く、また、日常生活では静止した状態で音を聴くことはほとんどない。ここでは、視覚情報と自己運動情報に着目して音空間知覚に与える影響について触れる。

視覚情報と音空間知覚の統合処理として有名な現象は腹話術効果(ventriloquism effect)¹²⁾である。音源と映像の種類にも依存するが、映像と音源のずれが10°前後までであれば、音像の位置は映像の位置に引き寄せられる方向に偏位する。

音空間知覚中に聴取者が動くことにより、両耳に入力される信号が時間的に変化する。これは大局的に考えると音空間知覚手がかりが増えることになるため、頭部運動は音空間知覚に寄与すると考えるのが妥当である。ただし、前述のように頭部運動を含む自己運動が音空間の正確な知覚に寄与するといった報告^{13),14)}だけでなく、音空間知覚を変容させるといった報告¹⁵⁾⁻¹⁷⁾もあり、より詳細な研究が待たれる。

これらの知見は、バーチャル音空間を評価する際に、単に聴覚情報の精度を求めるだけでなく、他の感覚情報を含めた形で評価を行う必要があることを示唆している。

4. バーチャル音空間の精度評価

ここまでで、我々がバーチャル音空間(環境)から受ける印象をどのように感性表現をしているか、また、我々がどのように音空間を知覚しているのかについてまとめた。バーチャル音空間の評価をするにあたり、これらの知識を踏まえた上で評価法を考えることは、実際に音空間を知覚するのが人間である以上極めて重要である。

ただし、システムを構築、改良して高精度にバーチャル音空間を提示することを考えると、操作可能な物理パラメータと人間の「場」としての音空間の知覚過程との対応関係を知り、その知見に基づいて要求された精度を満たすように物理パラメータを操作することになる。しかし、現状では人間の「場」としての音空間の知覚自体に不明な点が多く、操作可能な物理パラメータとの対応関係はほとんどわ

かっていない。したがって、物理指標に基づいて物理的に忠実に所望の音空間が合成できているかの評価と、実際に我々が体験した際の印象が適切であるかどうかの評価がそれぞれ行われることになる。

ここでは、比較的よく使われる評価指標について、物理指標と主観指標に分けていくつか示していく。

4.1 バーチャル音空間の物理評価指標

提示したい所望の音空間情報があれば、単純に再現された音空間との物理的な特性を比較することになる。例えばスペクトル歪み (spectral distortion: SD) であれば、様々な方向に配置された音源から聴取点に音が到来した際に、その方向から到来した音と、それを模擬した音との周波数スペクトルを音源が配置された全周囲にわたって比較する。例えば以下のような式で差分を算出する。

$$SD = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(20 \log_{10} \frac{|H_{\text{synthesize}}(i, f)|}{|H_{\text{target}}(i, f)|} \right)^2}$$

ここで N は評価する音源の数、 H_{target} 、 $H_{\text{synthesize}}$ はそれぞれ、目標とする音空間の周波数スペクトル、合成された音空間の周波数スペクトルである。通常、 N 個の音源は全周囲に配置し、全ての方向から到来する音が正しく表現されているかを計算することになる。一方時間波形の比較では、相互相関関数や、二乗平均誤差を求めるといった方法が取られる。例えば清水らは、ある点 r における時間波形について、目標とする音空間の音圧 p_{target} と合成された音空間の音圧 $p_{\text{synthesize}}$ を以下の式で比較して精度を分析している¹⁸⁾。

$$E = \max_t \left[10 \log_{10} \left\{ |p_{\text{synthesize}}(r, t) - p_{\text{target}}(r, t)|^2 \right\} \right]$$

ある領域全体にわたって再現される音空間の精度を検討する際には、領域内に評価したい点を稠密に配置し、上記のような方法で、それぞれで所望の音空間と合成された音空間との物理パラメータを比較する。時間波形、周波数スペクトルで比較する場合の他に、高次アンビソニックス (higher order Ambisonics: HOA) では球面調和関数の展開係数自体を比較するといった場合もある。

ただし、物理パラメータがどれだけ合っていれば聴感上差がないといえるかはわかっていないことが多く、体系的な知見も存在しない。したがって、実際に人間が体験するようなシステムの評価では、最終的に人間が聴いて精度を評価することになる。

4.2 バーチャル音空間の主観評価指標

人間がバーチャル音空間を体験して最終的な評価

を行う必要があるのだが、その際に使用する評価指標や評価方法も統一されたものではなく、様々な研究者が目的に合わせて評価法を検討し実施している状態である。ただし多くの場合、目的に合わせて選ばれた形容詞群を用いて実際に聴いた印象を判断して分析していることが多い。

例えば、バーチャル音空間から受ける印象のうち最も一般的な臨場感については、臨場感や実在感、音像定位や音響空間の性質などの評価に用いられる多数の形容詞を用いて音の臨場感・空間印象を評定尺度させて因子分析した結果、「感動」「移動」「評価」「奥行」「躍動」「空間」「臨場」の7因子に分類できることが明らかとなっている¹⁹⁾。また、「臨場感」が心を揺さぶる事象に対して用いられるということもあり、「感動」という尺度を用いてバーチャル音空間の評価を行うといった試みもなされている²⁰⁾。この研究プロジェクトでは、評価に使用するために標準的に使用する形容詞の選定が行われ、実際に22.2 マルチチャンネル音響方式の評価に使用されている。さらに、バーチャル音空間の印象の多次元構造を明らかにしてシステム評価を行うことを念頭に、コンテンツに依存しない評価が可能な評価用形容詞の選定も進められている²¹⁾。この際には、様々な方式で提示されたバーチャル音空間から受ける印象を選定された形容詞対を用いて評定尺度させ、システムの差異を表現可能な因子を抽出している。評定尺度法だけでなく、評価対象全ての組み合わせに対してある尺度で回答させる一対比較法や、複数の両極の形容詞対の尺度に対して回答させる SD (semantic differential) 法を用いた印象評定に基づいて評価するといったことも比較的多く行われるようである。

一方、バーチャル音空間から受ける多次元的な印象を統合し、得たい指標に対する評価を直接行うといった場合もある。例えば「臨場感」を「あたかもその場にいる感じがする」と説明を付記した上で、その高低を判断させるというものである。ただしこの場合、回答時に想起されるその指標の解釈が統一されていないと得られた結果のばらつきが大きくなるため、全員が同じような印象を想起するような説明を付記するか、より一般的で紛れの少ない評価語を用いて判断させる必要がある。さらに、全体的な結果のみが得られるだけなことから、システム改良へのフィードバックを得るには工夫が必要となる。

主観評価結果から物理パラメータへの対応関係が比較的考えやすいという点では、音像定位や距離定

位などの基本的な聴覚特性の測定といったことも考えられる。実際頭部伝達関数を用いたバーチャル音空間合成技術の評価では、水平面や正中面の音像定位性能を実音源とバーチャル音源の両者を用いて測定し、その差異を分析することで評価することになる。しかし、音像定位は実験で聴取者が音像位置を意識して聴いている状態であり、我々が日常音空間を知覚する際の様相とは必ずしも一致しない。

近年では、視線、瞳孔径変化、脳波などを指標として生理指標を用いてバーチャル環境から受ける印象を客観評価するといった試みもなされている。

5. さ い ご に

バーチャル音空間の評価に関しては、人間の音空間知覚過程の分析も含め、物理指標との対応が良い主観評価指標、もしくは主観評価に対応する物理指標が未だ存在せず、現在でも多くの研究者が取り組む研究課題である。今後、より定量的な評価指標が提案されることが期待される。

参 考 文 献

- 1) 鈴木陽一, トレビーニョ ホルヘ, 坂本修一: 高臨場感音響技術の最新動向と将来展望, 電子情報通信学会誌, vol. 101, no. 8, pp. 786-792 (2018).
- 2) 行場次朗: 臨場感と迫真性の感性心理学的特性, 電子情報通信学会誌, vol. 101, no. 8, pp. 793-797 (2018).
- 3) 寺本渉, 吉田和博, 浅井暢子, 日高聡太, 行場次朗, 鈴木陽一: 臨場感の素朴な理解, 日本バーチャルリアリティ学会論文誌, vol. 15, no. 1, pp. 7-16 (2010).
- 4) 寺本渉, 吉田和博, 日高聡太, 浅井暢子, 行場次朗, 坂本修一, 岩谷幸雄, 鈴木陽一: 「迫真性」を規定する時空間情報, 日本バーチャルリアリティ学会論文誌, vol. 15, no. 3, pp. 483-486 (2010).
- 5) 超臨場感コミュニケーション産学官フォーラム (URCF), <https://www.urcf.jp/>
- 6) 小澤賢司: 聴覚臨場感の基礎特性—コンテンツ臨場感とシステム臨場感—, 信学技報, vol. 108, no. 333, EA2008-115, pp. 83-88 (2008), <https://ci.nii.ac.jp/naid/110007115046>
- 7) Lenna (22.2ch) free distribution, <https://miyuhosoi.com/lenna/>
- 8) K. Iida, M. Itoh, A. Itagaki, M. Morimoto: Median plane localization using a parametric model of the head-related transfer function based on spectral cues, *Appl. Acoust.*, vol. 68, pp. 835-850 (2007).
- 9) F. Asano, Y. Suzuki, T. Sone: Role of spectral cues in median plane localization, *J.A.S.A.*, vol. 88, no. 1, pp. 159-168 (1990).
- 10) Y. Iwaya, T. Magariyachi, M. Otani, Y. Suzuki: Role of 4-8kHz band component for wideband noise localization in median plane, *J.A.S.A.*, vol. 131, p. 3216 (2012).
- 11) F. Rumsey: Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm, *J. Audio Engineering Society*, vol. 50, pp. 651-666 (2002).
- 12) R.B. Welch, D.H. Warren: Immediate perceptual response to intersensory discrepancy, *Psychological Bull.*, vol. 88, pp. 638-667 (1980).
- 13) W.R. Thrlow, P.S. Runge: Effect of induced head movement in localization of direction of sound, *J.A.S.A.*, vol. 42, pp. 480-488 (1967).
- 14) Y. Iwaya, Y. Suzuki, D. Kimura: Effects of head movement on front-back error in sound localization, *Acoust. Sci. & Tech.*, vol. 24, no. 5, pp. 322-324 (2003).
- 15) J. Leung, D. Alais, S. Carlile: Compression of auditory space during rapid head turns, *Proc. Natl. Acad. Sci. USA*, vol. 105, pp. 6492-6497 (2008).
- 16) W. Teramoto, S. Sakamoto, F. Furune, J. Gyoba, Y. Suzuki: Compression of auditory space during forward self-motion, *PLoS ONE*, vol. 7, no. 6, e39402 (2012).
- 17) A. Honda, Y. Masui, Y. Suzuki, S. Sakamoto: Effect of passive whole-body rotation on sound localization accuracy of listener subjective straight ahead, *Acoust. Soc. & Tech.*, vol. 41, no. 1, pp. 249-252 (2020).
- 18) T. Shimizu, J. Trevino, S. Sakamoto, Y. Suzuki, T. Ise: Evaluation of the extension and coloration of multiple listening zones synthesized by the shared field reproduction system, *Acoust. Sci. & Tech.*, vol. 40, no. 4, pp. 241-249 (2019).
- 19) 谷口高士: 高臨場感コミュニケーションにおける聴覚的臨場感の階層的印象推定モデル, 電子情報通信学会誌, vol. 101, no. 8, pp. 804-811 (2018).
- 20) 大出訓史, 安藤彰男, 谷口高士: 音楽によって喚起される感動の種類と音楽や音響の印象の関係, *信学論 (A)*, vol. J97-A, no. 4, pp. 323-331 (2014).
- 21) 岡部敏貴, トレビーニョ ホルヘ, 山高正烈, 坂本修一, 鈴木陽一: 音空間再生方式の差異が音空間の印象に与える影響, *音講論 (秋)*, 2-1-18, pp. 1271-1272 (2019).