芸術科学会論文誌 Vol. 20, No. 5, pp. 219 – 230 (2021)

サラウンド音響環境における視聴覚情報の時間的不一致の影響

矢島春香 [1)](非会員)　越智景子 [2)](非会員)　大淵康成 [2)](正会員)

1) 東京工科大学大学院バイオ・情報メディア研究科メディアサイエンス専攻
2) 東京工科大学メディア学部

Influence of Temporal Inconsistency of Audio-Visual Information in Surround Sound Environment

Haruka Yajima[1)]　**Keiko Ochi**[2)]　**Yasunari Obuchi**[2)]

1) Graduate School of Bionics, Computer and Media Science, Tokyo University of Technology
2) School of Media Science, Tokyo University of Technology

g3118029b9 @ edu.teu.ac.jp, { ochikk , obuchiysnr} @ stf.teu.ac.jp

概要

本研究では，サラウンド音響を視覚情報と組み合わせた環境に着目し，両者に時間的な不一致がある場合の影響について調査を行うことで，高臨場感コンテンツの実用性向上を目指す．左右方向の知覚実験では，映像と音声を時間的にずらした状態で提示し，ずれの程度に対する違和感の調査を行った．その結果，音が少し遅れている方が違和感は少ないという結果が得られた．また，音源までの視覚上の距離を変えると，最適な遅れが変わることがわかった．前後方向の知覚実験では，聞き慣れている音に比べて，聞き慣れていない音の前後定位の精度が低いことが確認された．また，視覚による定位が聴覚による定位を妨げることや，視聴覚の同期が定位精度劣化をもたらすことがわかった．これらの知見をもとに，高臨場感のコンテンツを作成するための指針を示すことができた．

Abstract

This study focuses on the environment of surround audio and visual information. We aim at enhancing the practicality of highly realistic media contents by investigating the influence of temporally mismatched audio-visual interaction. In the left-to-right localization experiments, we presented temporally mismatched audio and visual signals, and investigated the relationship between the delay and feeling of oddity. The results showed that it sounds more naturally if the sound comes behind the image. In addition, the optimal delay depended on the visual distance to the object. In the fore-to-aft localization experiments, it was found that the accuracy degrades if the listener is not familiar to the sound. We also found that the visual localization interferes the auditory localization, and the synchronicity between audio and visual stimuli degrades the localization accuracy. These findings have shown guidelines of highly realistic contents creation.

1 はじめに

人間は，主に視覚と聴覚を使って周囲環境を立体的に認識している．人に環境を疑似体験させるメディアでは，人間が立体感を知覚する仕組みを，様々な形で模擬することを試みてきた．視覚における立体感の再現は，絵画における遠近法に端を発し，両眼視差を用いた立体ディスプレイなどに発展している．一方，聴覚における立体感の再現は，オーディオ機器におけるステレオ再生から，劇場やホームシアターのサラウンド音響などに繋がっている．近年，そうした技術はさらに発展してヴァーチャルリアリティとして結実し，視覚と聴覚の両方の点で，極めて臨場感の高いコンテンツが提供されるようになりつつある．

こうした視聴覚融合コンテンツにおいて，システムの性能上の制限や，演出上の意図などにより，音声と映像とが時間的不整合を起こすことがある．しかし，時間的不整合があっても必ずしも臨場感が低減するとは限らない．テレビの衛星中継では，音声と画像にずれが生じることがあるが，わずかなものであれば許容されている [1]．ステージ音響では，先行音効果 [2] によってステージの方向を意識させるために，意図的にスピーカー音声に遅延を入れる場合もある．

本研究では，サラウンド音響環境における方向定位を主眼に置き，こうした時間的不整合の効果を明らかにすることを試みる．視覚における方向定位が比較的容易なタスクであるのに対し，聴覚における方向定位には不確かさが伴う．そうした音源定位のタスクを更に二つに分け，

1. 不確かさの程度が低い左右定位において，視覚との時間的不整合がどのように許容されるのか
2. 不確かさが顕著である前後定位において，視覚との不整合が定位精度にどのように影響するのか

を明らかにする実験を行う．1 のタスクにおいては，対象となる物体の形や大きさ，移動速度などが人間のメンタルモデルに影響を及ぼす可能性がある [3, 4, 5] という予測に基づき，対象物の見え方をコントロールしながら違和感の変化を測定する．これは，視聴者が左右を間違えることは殆どないため，コンテンツにおける臨場感の主眼は，視聴者にとって自然で快適な情景の再現という点に置かれることが多いことを意識したものである．一方，2 のタスクにおいては，そもそも正しい定位ができるかどうかが状況によって大きく変わるはずだと考え，違和感のような主観評価ではなく，定位精度という客観指標による評価を試みる．ここでは，コンテンツ作成においても，視聴者に対していかに正確に方向感を伝えるか，あるいは意図的に混乱に誘導するかといったことが，製作者の主眼となることを意識している．

誤りが頻繁に生じる前後定位においては，どのような信号を聴かせるのかも重要な要素である [6]．特に，視聴覚融合コンテンツへの応用を考える場合には，正弦波のような単純な信号と，人間の声や様々な日常生活音などとでは，異なる影響が生じる可能性も高い．そこでタスク 2 の実験では，複数の音響信号を用意し，それらの間の違いについても検討する．

本論文は，著者らによる会議論文 [7, 8, 9] の内容を整理し，さらなる解析結果を加えてまとめたものである．本論文の構成は以下の通りである．第 2 節では，視聴覚相互作用に関する先行研究を示し，本研究の課題の位置づけを明確化する．第 3 節は，前述のタスク 1 についての章で，左右方向の知覚に関する実験の内容を紹介する．第 4 節は，タスク 2 に対応する実験の章で，前後方向の知覚についての検討を行う．第 5 節で実験全体に対する考察を述べたのち，第 6 節をまとめとする．

2 先行研究

2.1 視聴覚相互作用に関する研究

本研究では，視覚と聴覚で矛盾した情報が与えられた場合の人間の知覚を扱う．矛盾の形態には様々なものが考えられるが，これまでに報告されている研究例の中で重要なものを紹介する．

視聴覚相互作用の例として良く知られているのが，マガーク効果 [10] である．マガーク効果の実験では，単音節発声を用い，視覚と聴覚で異なる刺激が与えられた場合に，知覚の内容が変化する現象が確認されている．このとき，耳で聞いた音節を知覚するケース，目で見た音節を知覚するケースだけでなく，どちらとも異なる音節を知覚するケースがある．これは，視覚が聴覚に優先するのみならず，両者が混合的に作用して知覚される可能性を示唆している．また，マガーク効果をさらに拡張させた例として，長田ら [11] は，音声提示方向と映像提示

方向の空間的なずれがマガーク効果に与える影響を調査した．この結果からは，視覚情報と聴覚情報の空間的一致度が，相互作用の大きさを決める重要なパラメータであることが示されている．

視覚と聴覚の矛盾については，腹話術効果 (ventriloquism) [12] も有名である．腹話術効果とは，聴覚刺激と視覚刺激が異なる場所で提示されたときに，聴覚的な方向知覚が視覚の影響を受けることを指す．Bertelson らが行った正弦波の音の方向知覚の繰り返し実験では，視覚刺激の存在により，正しい知覚への収束が遅くなったり，最後まで正しい知覚に収束しなかったりする様子が示されている．視聴覚間の不整合に関する実験を行う際には，両者が同じ対象物に起因すると解釈されることが重要だが，腹話術効果の研究例はそのような観点での参考となりうる．

マガーク効果が内容的不一致，腹話術効果が空間的不一致を扱うのに対し，時間的不一致を扱う例としてダブルフラッシュ効果 [13] がある．この研究では，聴覚が視覚に影響を与える例として, 聴覚刺激の存在により, 一度しか光っていないディスクが複数回光ったように知覚されるという実験結果が示された．また，視覚に影響を及ぼしうる聴覚刺激の時間間隔として，およそ 100 ミリ秒という上限値も示されている．類似の時間差閾値は [14] でも示されており，大脳皮質における処理の初期段階に結び付けられている．また，[15] では，聴覚刺激に残響を加えることで距離感を模擬した実験を行い，音源までの距離が長いほど時間差が許容されることや，その場合の時間差が音速で説明できることが示されている．これらの実験はビープ音などのシンプルな音響刺激を用いて行われているが，マガーク効果の実験に提示タイミングのずれを導入することで，音声発話に対する時間的不一致の影響を調べた研究 [16] もある．これらの結果は，時間的に不整合な視聴覚刺激が，同一対象物に起因すると解釈されるための必要条件を考える際に重要である．

これらの研究で得られた視聴覚相互作用についての知見は，テレビコンテンツの制作などにも生かされている [17]．テレビコンテンツの中で，映像に合わせて詳細に音源位置を定位するには大変な労力が伴うが，腹話術効果により補正が行われるのであれば，必ずしも厳密な位置に定位しなくても済む．NTSC 方式のテレビでは，標準的な視距離からの画面の見込み角が 10 度程度と考え，こうしたずれは許容されることが多い．一方，HDTV の場合には見込み角が更に大きくなることから，音響定位性能を上げるための機構が追加されている．本研究の成果を実用化する際には，こうした研究例が参考になるはずである．

2.2 音の方向知覚に関する研究

音源定位は主に，両耳間時間差，両耳間強度差，頭部伝達関数に由来する周波数特性などを使って決められる [18]．特に，前後の知覚には両耳間時間差や両耳間強度差が役立たないことから，周波数特性の役割が大きい．Kumpik らは，片耳をふさいだ状態での音源定位の実験により，フラットな周波数特性を持つ音に対して前後を含む定位の学習効果がある一方，ランダムな周波数特性を持つ音にはこの学習効果が働きにくく，特に前後の定位の誤りが生じやすいことを示した [6]．本稿第 4 節では，これらの性質を考慮に入れて前後知覚の実験を設計する．

3 音の左右知覚における視聴覚不一致

本節では，不確かさの程度が低い左右方向の定位において，視聴覚間の時間的不整合が許容されうること，また場合によっては臨場感に良い影響をもたらしうることを実験により明らかにする．左右方向の定位は，視覚はもちろん聴覚においても比較的容易なタスクであり，定位そのものに失敗することはほとんど無い．そこで，視聴覚間の時間的不整合を持つコンテンツを用意し，被験者が感じる違和感を答えてもらう主観評価実験を行い，その結果を元に，最も違和感を覚えさせない視聴覚間の関係はどのようなものであるのかを明らかにする [7].

3.1 実験準備

静かな会議室において，図 1 に示すような正方形配置となるようスピーカ (Fostex PM0.3) を置いた．スピーカの高さは 80 cm(床面からスピーカ底面までの距離) とした．さらに，上辺の中央部にモニタ (HITACHI 46 インチプラズマテレビ P46-GP08) を置いた．4 つのスピーカは，USB オーディオ (Roland Studio-Capture UA1610) を経由して USB で，モニタは HDMI で，それぞれ iMac に接続されている．

正方形の大きさは，一辺の長さ 130 cm と 260 cm の 2 種類で行ったが，以下に示す結果は両者を合計したものである．前方左側と右側のスピーカをそれぞれ $L1$, $R1$

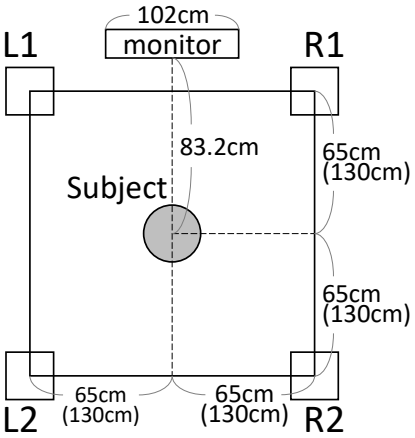

図 1　実験環境 (1)

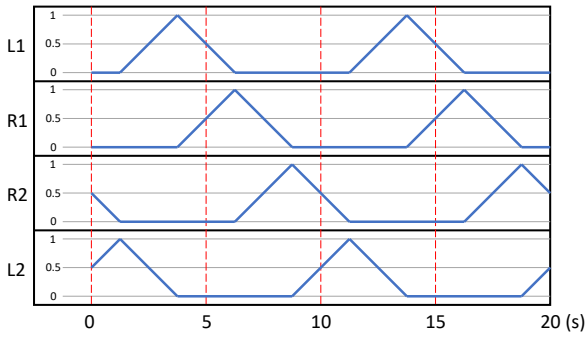

図 2　各チャンネルの音量変化

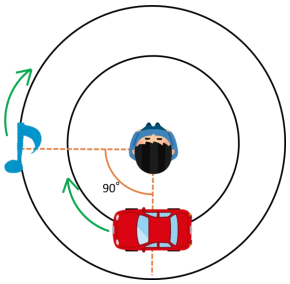

図 3　音を 90° ずらした様子

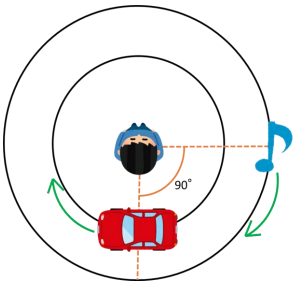

図 4　音を-90° ずらした様子

と呼び，後方左側と右側のスピーカを $L2$，$R2$ と呼ぶことにする．被験者は，4 つのスピーカの中央の位置に，前方を向いて座っている．

この状態で，視覚および聴覚情報の提示を行った．モニタに提示される視覚情報は，車が自分の周りを周回している様子である．車は円を描いてゆっくりと被験者の周りを右回りに 2 周する．ただし，固定されたカメラは前方のみを撮影しているため，前方以外を走っているあいだは車は映らない．撮影は，車が描く円の大きさを変えながら，半径 2.3 m，11.0 m，36.7 m の 3 種類で行った．いずれの場合も一周にかかる時間が 20 秒になるよう速度を調整している．

スピーカから提示される聴覚情報は，視覚情報の撮影と同時に録音した車の走行音である．走行音は，サンプリング周波数 44100 Hz，量子化ビット数 16 ビットでモノラル録音を行った後，4 チャンネル分を複製し，さらに図 2 のような時間的音量変化になるよう音量の調節をした．つまり，各スピーカから聞こえる音は，音量のみが異なっている．これらの音を 4 つのスピーカから再生することにより，車が周回しているように感じさせることができる．例えば，図 2 の 0 秒時点では，$R2$ と $L2$ の音量が等しく，なおかつ $L1$ と $R1$ の音量が 0 なので，真後ろにいるように聞こえる．5 秒時点では，$L1$ と $L2$ の音量が等しく，なおかつ $R1$ と $R2$ の音量が 0 なので，左側にいるように聞こえる．被験者の顔の中心位置で測定した音量は，最大 62.2 dB SPL であった．

なお，図 2 は，実験開始時に音が真後ろから聞こえ始めるような設定となっているが，実際の実験では，このデータの途中から再生を開始することにより，任意の角度から音が聞こえ始めるようにする．このとき，20 秒以降は先頭に戻って再生を続ける．真後ろから時計回りに $r°$ 進んだ角度から再生を始める場合には，音データの $r/36$ 秒のところから再生を開始する．一方，映像は常に真後ろから始まり，20 秒間で 2 周して終了する．以下では，この r を映像に対する音の角度のずれと定義する．

実験は，作成した動画をランダムで流し，1 つの動画を視聴するごとに，ノート PC のキーボードを用い，映像と音の空間的位置に違和感を覚えれば ‘y’ のキーを，感じなければ ‘n’ のキーを押すように指示した．その際，どちらかのキーを押すと次の動画が再生される旨だけを

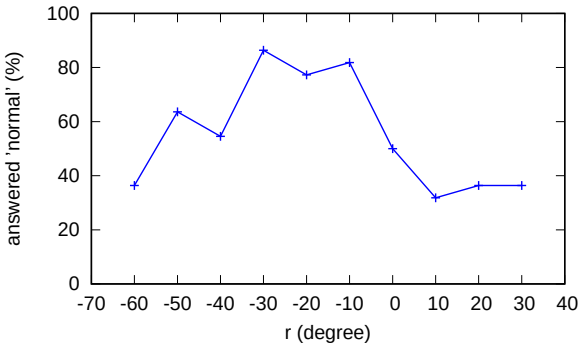

図 5　ベースライン実験の結果

図 6　近い距離を走る車の映像 (near)

図 7　遠い距離を走る車の映像 (far)

伝えており，判定の制限時間は設けていない．

3.2　ベースライン実験

はじめに，角度のずれ r に対して，違和感を覚えない人の割合がどのように変わるかの概略を確認するため，r の値を小刻みに変えながらの実験を行った．このとき，動画撮影用のカメラと自動車の距離は中程度 (11.0 m) とした．予備実験において，$r>30$ では「違和感が無い」と答える人の割合が安定的に少ないという結果が得られていたため，$r=-60$ から 10° 刻みで $r=30$ までずれの値を変えて実験を行った．各回の実験では，各被験者は，10 個の動画を全て 1 回ずつ，ランダムな順番で視聴した．被験者は女性 10 名と男性 12 名の合計 22 名で，10 代が 4 名，20 代が 16 名，30 代が 1 名，50 代が 1 名という年代分布となっており，全員が健聴者である．

実験結果を，図 5 に示す．横軸がずれの角度 r，縦軸が「違和感が無い」と答えた人の割合を示す．映像と音の定位方向が同じである $r=0$ の動画よりも，音がやや遅れている $r\leq -10$ の動画の方が，違和感が無いと答える人数が多いという結果になった．こうした傾向は $r=-30$ 程度まで続き，それより更にずれが大きくなると，違和感を覚える人が多くなる．

被験者ごとの結果を見てみると，$r=-60$ から $r=30$ までの 10 回の試行に対し，違和感が無いと答えた回数が 3 回から 8 回までばらついており，そもそも被験者ごとに違和感を覚える基準が異なっている様子が見て取れる．被験者ごとに，違和感が無いと答えたときのずれ r の平均を取ると，−21 から −5 の範囲に分布しており，全体の平均は −10.4 であった．

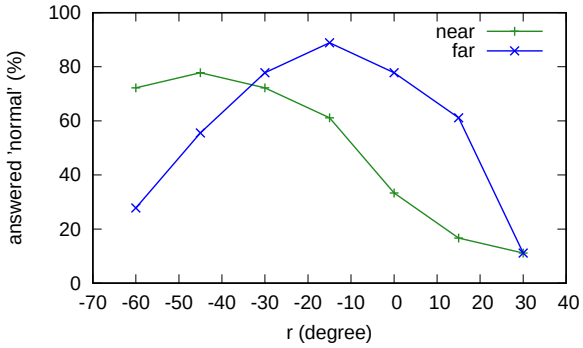

図 8　違和感が無いと判断した人の比率 (距離別)

3.3　対象物までの距離の影響

次に，対象物までの距離によって，違和感を覚えないと答える範囲がどのように変わるかを調べるため，距離 2.3 m と 36.7 m の映像を用いた実験を行った．以下，前者を near，後者を far と呼ぶ．それぞれの実験における画面の例を，図 6 および図 7 に示す．なお，この実験では，同一の被験者に 2 種類の距離の動画を見せるので，トータルでの実験回数削減のため，15° 刻みで 7 種類の動画を用意し，near と far それぞれについて 7 種類，合計 14 種類の実験をランダムな順番で行った．被験者は女性 9 名と男性 9 名の合計 18 名で，10 代が 4 名，20 代が 13 名，50 代が 1 名という年代分布となっており，全員が健聴者である．

3.4　実験結果

距離別に行った実験の結果を，図 8 に示す．ベースライン実験と同様に，映像より音が遅れている場合の方

が，違和感が無いとする被験者の割合が高いという結果となった．また，near では，$-60 \leq r \leq -15$ の範囲で違和感を覚えないと答えた人が 60 % を超え，ピークは $r = -45$ のときであった．far では，$-30 \leq r \leq 15$ の範囲で違和感を覚えない人が 60 % を超え，ピークは $r = -15$ のときであった．なお，各被験者の実験において，near と far の回答が r のすべての値で完全に一致したケースは無かった．

図 8 の結果は，1 条件あたり 18 回の試行に対する回答分布を示したものであり，試行回数が十分であるとはいえないが，この実験の範囲内での統計的な分析を行うため，統計分析ソフト R を用いてフィッシャーの正確確率検定を行った．r の値を固定した条件で，near の正解と不正解，far の正解と不正解から成る 2×2 の分割表に対してフィッシャーの正確確率検定を行ったところ，$r = -60$ (near:正解 13/不正解 5, far:正解 5/不正解 13), $r = 0$ (near:正解 6/不正解 12, far:正解 14/不正解 4), $r = 15$ (near:正解 3/不正解 15, far:正解 11/不正解 7) の 3 つの条件でそれぞれ $p = 0.0184, 0.0176, 0.0153$ となり，有意差が見られた (有意水準 0.05).

3.5 考察

ベースライン実験では，車の進行方向に対して，音が遅れて聞こえる場合は，ある程度のずれがあっても，違和感が無いと答える人が多かった．音の定位方向が，映像よりも 10° から 30° 遅れたときに，一番違和感が無く動画を視聴できることがわかった．これは，音は耳に届くまでに少し時間がかかることを，多くの人が認識していて，微量な遅れは日常生活においても許容されているからではないかと考えられる．反対に，音が映像よりも早く聞こえるものについては，日常生活では考えにくく，違和感を覚えやすいのではないかと考えられる．違和感が無いと答えた被験者が最も多かったのは，映像と音の定位方向が一致している 0° の場合ではなかった．

図 8 の $r = -60, 0, 15$ における near と far の違いは，near の反応が far の反応に対して遅れていると解釈すれば説明できる．実際，near の結果を r の＋方向に 30 度ぶんだけ移動させて far の結果と比較すると，あらゆる条件で有意差が無いという結果となる．

音の伝達にかかる時間を考えると，far の方が遅れが大きくなると考えられるが，実験結果はそうなってはおらず，音の伝達時間がずれに対する違和感の有無を決める重要な要因にはなっていないということがわかる．距離により違和感が異なる第一の要因としては，far の方が車の走行速度が速いため，わずかな時間のずれが長い距離として解釈され，違和感を生むということが考えられる．さらに，第二の要因として，車は点音源ではなく，横方向に広がりを持っているため，near では画面上の大きな部分を占めるということがある．仮に車長を 3.4 m とすると，それが 2.3 m 離れた位置にあるときの視野角は約 73° である．そのため，車の中心から前後に 35° 程度ずれた方向から音が聞こえても，車のボディの一部から音が鳴っているように感じられる．一方，36.7 m 離れている場合，視野角は約 5° である．仮に車の中心よりもやや後方の位置を基準として違和感の有無が決まると仮定すれば，そのピークの位置が near においてより後方に来ることはある程度説明できる．ただし，なぜ中心よりも後方になるのかについては，車の認知モデルそのものについての分析が必要であろう．また，有意といえる差ではないものの，図 8 において near の方がやや幅広く分布しているように見えるのは，第一の要因により遅れに対する許容度が高いためと考えることも可能である．

4 音の前後知覚における視聴覚不一致

本節では，不確かさが顕著である前後方向の音源定位において，視覚との時間的不整合が定位精度にどのように影響するのかを実験により明らかにする [9]．前後知覚では定位そのものに失敗することも多いため，定位の正解率を指標とし，それが視聴覚間の時間的不整合の有無によってどのように変化するかを調べる．

前後方向の定位では，Kumpik らの実験 [6] で示されたように，被験者が聴覚刺激に慣れているかどうかによって正解率が影響を受ける．これは，過去に聞いたことがある音であれば，どの周波数にディップが生じているかといったことをヒントに定位を行えるのに対し，知らない音がどのような周波数特性の影響を受けたかを推測することは難しいためである．そこで本節の実験では，多様な種類の聴覚刺激を用意し，定位正解率を比べることにより，音源の種類が視聴覚不整合に与える影響を定量的に評価する．

4.1 実験準備

実験は防音室内で行った．実験時の器材配置は図 9 の通りである．

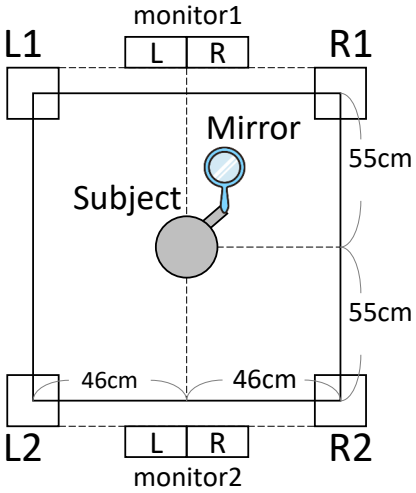

図 9　実験環境 (2)

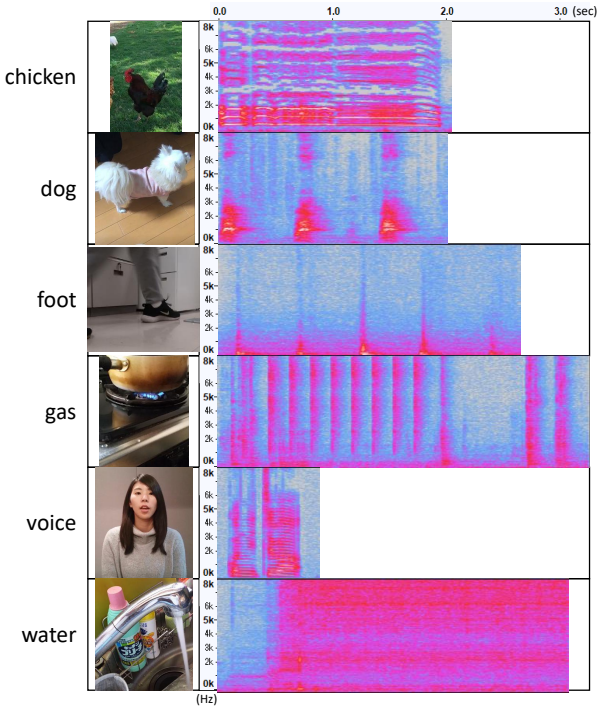

図 10　6 種類の非電子音のスクリーンショット (左) とスペクトログラム (右): 上から順に，鶏の声，犬の声，足音，ガスの音，人間の声，水道の蛇口の音

$L1, R1, L2, R2$ の 4 つのスピーカ (Fostex PM0.3) は，縦 110 cm 横 92 cm の長方形の頂点上，高さ 80 cm の位置に置かれており，実験用 iMac 内の多チャンネル音響処理ソフトウェア (JRiver Media Center 23) で再生し，USB オーディオ (Roland Studio-Capture UA1610) を経由して出力した．なお，再生にあたっては，音量の絶対値が定位のヒントにならないよう，試行ごとに 20 %〜100 % の範囲でランダムに音量を変更した．

前方の映像刺激は，iMac の 21.5 インチモニタに提示される (monitor1)．提示用映像を 2 種類用意し，モニタの左半分 (L) もしくは右半分 (R) に画像が現れるようにする．後方の映像刺激は，単焦点プロジェクタ (Ricoh PJ WX4152) を使用してスクリーンに投影される (monitor2)．こちらも呈示用映像を 2 種類用意し，スクリーンの左半分 (L) もしくは右半分 (R) に投影されるようにする．被験者からは後方スクリーンが直接は見えないため，直径 14.5 cm の手鏡を渡し，後方の確認は手鏡を用いて行うこと，その際にはできる限り頭を動かさないようにすることを指示した[*1]．

実験に用いる聴覚刺激は，以下に挙げる 10 種類を用意した．日常生活で聞く機会が多そうな音として，鶏の声，犬の声，足音，ガスコンロの音，人の声，水道の蛇口の音の 6 種類，それに電子音が 4 種類である．これらに合わせた視覚刺激として，日常音については該当するものの様子を撮影した動画を，電子音に対しては抽象的な図形を音にリズムに合わせて変化させた動画を用意し，音の再生のタイミングに合わせて提示した．6 種類の日常音の視覚刺激の 1 シーンとスペクトログラムを図 10 に示す．これらの音は，どれも被験者の多くが聞き慣れた音であり，スペクトルの変化にも気が付きやすい．これに対し，4 種類の電子音の視覚刺激の 1 シーンとスペクトログラムを図 11 に示す．これらの音は，今回の実験用に新規に作成したものであり，被験者の記憶には無いはずである．また，スペクトル上でも強い部分と弱い部分がランダムに現れ，変化に気付きにくいものになっている．以下では，4 種類の電子音をそれぞれ piko, powa, wow1, wow2 と呼ぶ．

視覚刺激と聴覚刺激を組み合わせる際には，時間同期について 2 種類の組み合わせ方法を採用した．1 つは，映像が先に提示され，そのあと 2〜3 秒後に音が鳴り始めるもので，以下ではこれを視聴覚非同期の実験と呼ぶ．もう 1 つは，映像と音が完全に同じタイミングで表示さ

[*1] 鏡を使った後方の確認に際しては，当初は大型の鏡を screen1 の横に置いていたが，予備実験の被験者から「だんだんと右前にスクリーンがあるように思えてくる」という意見が聞かれたため，このような方法に変更した．

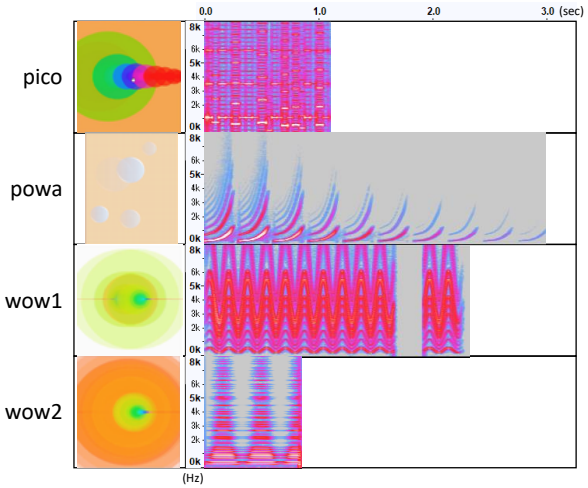

図 11　4 種類の電子音のスクリーンショット (左) とスペクトログラム (右)

れるもので，以下ではこれを視聴覚同期の実験と呼ぶ．

4.2　実験手順

20 代前半の女性 2 名，男性 12 名の被験者に対する定位実験を行った．被験者は全員が健聴者である．各試行においては，視聴覚刺激を提示し，それに対して音の前後方向をキーボード入力させた．その際「回答時間も測定するので，方向がわかったらすぐにキーを押すように」と指示した．これは，聞いた音に対する直接的な反応を見ることにより，刺激の提示タイミングの影響が現れやすくなることを期待しての指示である．

実験の各試行では，映像と音の提示方向を毎回変更した．ここで，映像・音それぞれの提示方向が 4 種類 (左前・右前・左後・右後) あるので，それぞれを独立に選ぶと 16 種類の組み合わせがあるが，その中で，映像と音の左右関係が一致するもの 8 種類のみを使用した．このような組み合わせを用いたのは，定位がわかりやすい左右方向の情報について，視覚と聴覚が一致していることにより，視覚と聴覚の連動性を感じさせるためである．これら 8 種類のうち，視覚の前後方向と聴覚の前後方向が一致しているものを一致条件，そうでないものを不一致条件と呼ぶ．これら 8 種類の提示方向と，10 種類の音源とを組み合わせた 80 種類の中から，ランダムに抽出した 30 個を使って 1 回のセッションを行う．なお，すべての試行を通じて，抽出された 30 個のうち日常音の個数の平均は 17.6，標準偏差 2.2，最大値 22，最小値 12 であった．また，抽出された 30 個のうち映像と音の提示方向が一致しているものの個数の平均は 14.7，標準偏差 2.4，最大値 19，最小値 11 であった．

なお，音の提示に際しては，スピーカを 1 つだけ用いる実験と，2 つ用いる実験とを別々に行った．2 つ用いる場合には，想定される方向のスピーカから音を鳴らすのに加えて，前後が同じで左右が異なるもう 1 つのスピーカから，20 dB ほど音量を下げた音を鳴らした．これは，左右方向の定位が若干わかりにくくなることが，前後方法の定位に影響を及ぼすかどうかを調べるためのものである．これにより，被験者 1 名あたり 2 回のセッション，合計 60 回の試行を行う．

実験は，視聴覚非同期，視聴覚同期の 2 回に分けて行った．ただし，両方の実験に参加したのは男性 8 名のみである．これらの被験者を「対応のある被験者」と呼ぶ．これに加えて，視聴覚非同期の実験には女性 2 名と男性 1 名，視聴覚同期の実験には男性 3 名の被験者が加わった．各実験における被験者数は 11 となり，それぞれ 60 試行を行うと合計で 660 回分のデータが得られるはずだが，視聴覚非同期の実験では，1 名の被験者が 2 スピーカのセッションに参加できなかったため，630 回分となっている．視聴覚同期の実験では，660 回分のデータが得られた．

4.3　実験結果と考察

実験結果を表 1 に示す．ここでは，対応のある被験者 8 名だけに対する誤答率と，全被験者に対する誤答率とを示した．全体的な傾向として，視聴覚同期の実験では，視聴覚非同期の実験に比べて誤答率が増えた．音の方向別の内訳で見ると，音が前で鳴っているときの誤答率が特に増え，他の条件に比べて顕著に高くなっている．視覚情報を確認してから聴覚情報に意識を向けられる場合と異なり，視覚情報に注意力が削がれている状況での音の知覚が困難になっている様子が見てとれる．特に，手鏡を使っている場合を含めて，視覚に対する注意力が前方に集中しているため，音が前で鳴っているにもかかわらず，後ろで鳴っていると感じてしまうケースが一定数存在しているのではないかと考えられる．なお，全被験者のデータを用いて，非同期と同期の 2 群に対するカイ二乗検定を行った結果，全体データでは有意差は見られなかったが，前方の音に限定した場合には，非同期で正解 298/不正解 22，同期で正解 298/不正解 37 となり，$p = 0.084$ で有意傾向が見られた ($p < 0.1$)．

表 1　条件別の誤答率 (%) 上段:対応のある 8 被験者/下段:各実験に参加した全 11 被験者

	全体	音の方向		動画種別		スピーカ数	
		前	後	一致	不一致	1	2
視聴覚非同期	4.7	6.8	2.3	4.0	5.4	2.9	6.3
	6.7	6.9	6.5	4.8	8.4	5.0	8.2
視聴覚同期	6.5	10.7	2.1	4.7	8.2	4.2	8.8
	8.5	11.0	5.9	5.0	11.9	7.0	10.0

動画種別ごとに見ると，一致動画に比べて不一致動画での誤答率が高くなり，映像に騙されて音を知覚することがあることを確認できた．予備実験 [8] において，視覚と聴覚に左右の差を持たせない場合，こうした差が小さくなる結果が見られていたが，左右の差を明確にしてなおかつ両者を常に一致させることにより，視聴覚の相互作用が大きくなり，こうした効果が観測されやすくなったものと思われる．最後に，スピーカを 1 つ使った場合と 2 つ使った場合では，2 つの場合の方が誤答率が高くなっており，左右定位の曖昧さが前後定位にも影響していることが確認された．

次に，音源の種類別の誤答率を図 12 (対応のある被験者) および図 13 (全被験者) に示す．“pico”, “powa”, “wow1”, “wow2” が電子音である．日常音に比べて電子音での誤答率が顕著に高くなっており，従来研究と同様の傾向が本実験においても見られた．全被験者のデータを用いて，日常音と電子音の 2 群に対してカイ二乗検定を行ったところ，非同期の実験では日常音の正解 342/不正解 18，電子音の正解 204/不正解 24 で $p=0.0289$，同期の実験では日常音の正解 345/不正解 17，電子音の正解 203/不正解 39 で $p=3.46\times10^{-5}$ となり，いずれの場合も $p<0.05$ で有意差が見られた．また，視聴覚同期による誤答率の増加は，電子音において特に顕著にみられていた．

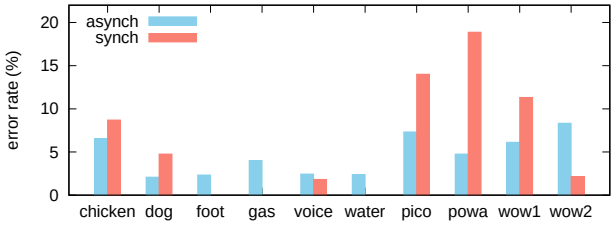

図 12　動画種類別の誤答率: 対応のある被験者 8 名の平均．asynch は映像と音の前後関係が不一致の場合，synch は一致の場合を表す．

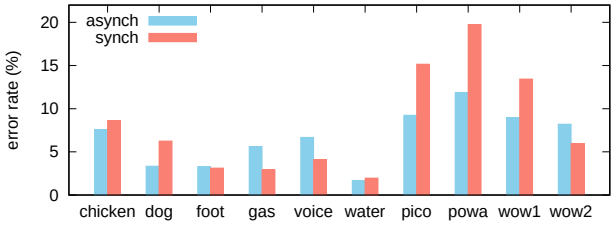

図 13　動画種類別の誤答率: 全被験者 11 名の平均．asynch は映像と音の前後関係が不一致の場合，synch は一致の場合を表す．

5　総合考察

第 3 節および第 4 節では，視聴覚融合コンテンツにおける視覚と聴覚の時間的不整合に着目し，音の定位という観点から，違和感の有無や定位の精度について調査した．

まず，音の左右方向の知覚に着目し，視覚情報と聴覚情報から得られる空間的ずれが，どのように意識されるのかを検証した．音と映像の定位方向が異なる動画を視聴し，違和感が有るか無いかを判断してもらう実験を行った．車の進行方向に対し音が先行して聴こえる場合は，違和感が有ると感じる人が多く，映像が先行して見える場合は，違和感を覚える人が少ないことがわかった．また，映像と音の定位方向が一致しているよりも，少し音が遅れる方が映像と音が合っていると感じる人が多いことがわかった．

聴覚情報の遅れに対し，対象物までの距離がどのように影響するのかについても調査を行った．その結果，近くに見えるものに対しては大きな遅れを許容するのに対し，遠くに見えるものに対しては小さな遅れしか許容しないと解釈することにより，実験結果を説明できることがわかった．ただし，今回の実験では，対象物そのもの

の視野角の影響があると考えられるため，今後さらに手法を変えての実験が必要である．

これらの結果から，サラウンドコンテンツにおける効果音の再生タイミングについての知見が得られる．対象物の位置から機械的に音の定位位置を決めるのではなく，対象物の動きを考慮した遅れを入れること，対象物までの距離 (もしくは対象物の見込み角) に応じてこの遅れの量を調整することなどが，高臨場感コンテンツ作りに必要となる．

なお，今回の実験設定では，near と far の距離の比が約 16 倍であり，音源や聴取環境が全く同じであれば，約 24dB の音量差が生じるはずである．実際の実験では，距離差の知覚そのものを調べているわけではないという点で，対象となる音の聴き取りやすさを優先して，そうした音量の差は再現しなかった．しかし，音量差が遅延の違和感に与える影響についても，今後検討する必要がある．

次に行った音の前後知覚の実験では，聞き慣れている音に比べて，聞き慣れていない音の前後定位が難しいという，従来からの知見が確認された．また，視聴覚相互作用が感じられるコンテンツにおいて，視覚による定位が聴覚による定位を妨げる様子も見られた．さらに，視聴覚情報が同期することにより，被験者の注意力が視覚情報に集中し，聴覚による定位精度が低下する様子も見られた．

なお，今回の前後知覚実験は，スピーカの使用数および視聴覚の非同期・同期の条件において，セッションの順番を固定して行った．そのため，両条件において順序効果が生じている可能性がある．ただし，実験結果を見ると，後から行った 2ch，同期の実験の方が誤答率が高くなっており，仮に慣れの影響があったとしても，それ以上に条件による差が大きく出たと解釈できる．とはいえ，順序効果を差し引いた厳密な差を議論するためには，両条件をランダムにミックスした形での追試を行うことが望ましい．

以上の結果から，サラウンドコンテンツを作成する際に注意すべき項目がいくつか挙げられる．音の定位が重要なシーンでは，日常で良く知っている音を使用することが望ましい．あるいは，それまでのストーリーの中で何度も再生され，視聴者が慣れている音を効率的に使うことも有用であろう．また，視聴者の注意力をうまく配分するために，視覚上の大きな変化のタイミングと，聴覚上の大きな変化のタイミングについて，両者をずらして徐々に立体感を感じさせる，あるいは同時に呈示することで混乱を与えるなどの演出手法が可能になる．

一般的な映像コンテンツでは，音の定位をわかりやすくして，立体感を感じさせるようにすることが重要である．一方で，音の定位をわざと難しくしたいコンテンツも考えられる．例えば，音源方向を頼りにもぐらたたきのようにキャラクターを探すゲームなどでは，上述の知見を逆に使い，音の方向をわかりにくくするということもできる．ゲームだけでなく，ホラー映画などの緊迫したシーンで，あえて定位をわかりにくくし，恐怖感を増幅するということも考えられる．

なお，本研究全体を通じて，サラウンドコンテンツの視聴者は 1 名だけとして，左右 (第 3 節) もしくは前後左右 (第 4 節) の中心にいる視聴者に対する定位のコントロールを検討してきた．しかし，多くの商用システムにおいては，多数の視聴者がいたり，あるいは単独であっても指定の位置からずれた位置で視聴するケースがあったりして，定位の厳密なコントロールができない場合も多い．そうした状況について，本研究では十分に扱っておらず，今後の研究課題となるであろう．

6 おわりに

本研究では，左右方向の定位と前後方向の定位を分けて実験を行い，それぞれの定位に対する視聴覚刺激の時間的不一致の影響を分析した．これらを通じて，視覚情報と聴覚情報が時間的に一致する場合と一致しない場合の違いを明らかにすることができた．不一致があっても違和感が無い場合や，不一致がある方が音源定位が正確になる場合もあり，コンテンツに応じて適切な不一致を使えば，むしろ効果的な演出ができる可能性が示唆された．

本研究では左右定位と前後定位を分けて実験を行ったが, 実際のコンテンツでは, 前後左右を自由に組み合わせて音源を動かすことが求められる．左右方向・前後方向のそれぞれに関する知見を活かし，臨場感の高いコンテンツを作成していくためには，両者を共に含む条件での実験・検討が必要となるであろう．また，鑑賞者の位置を大雑把にしか指定できない場合の対処についても，今後さらなる研究が必要である．実サービスに近い具体的

なコンテンツを用いた研究が進めば，今後のサラウンドコンテンツ制作が更に発展し，聴覚的な臨場感の高いコンテンツが生み出されていくようになると期待したい．

参考文献

[1] Recommendation ITU-R BT.1359, “Relative Timing of Sound and Vision for Boadcasting,” (Question ITU=R 35-4/11)

[2] R. Y. Litovsky, H. S. Colburn, W. A. Yost and S. J. Guzman, “The Precedence Effect,” The Journal of the Acoustical Society of America, Vol.106, No.4, pp.1633–1654 (1999)

[3] M. Schutz and S. Lipscomb, “Hearing Gestures, Seeing Music: Vision Influences Perceived Tone Duration,” Perception, Vol.36, pp.888-897 (2007)

[4] A. Lecuyer, S. Coquillart, A. Kheddar, P. Richard, and P. Coiffet, “Pseudo-Haptic Feedback: Can Isometric Input Devices Simulate Force Feedback?” Proc. IEEE Virtual Reality, pp.83-90 (2000)

[5] 鳴海拓志, 伴祐樹, 藤井達也, 櫻井翔, 井村純, 谷川智洋, 廣瀬通孝, “拡張持久力: 拡張現実感を利用した重量近く操作による力作業支援,” 日本バーチャルリアリティ学会論文誌, Vol.17, No.4, pp.333-342 (2012)

[6] D. P. Kumpik, O. Kacelnik, and A. J. King, “Adaptive Reweighting of Auditory Localization Cues in Response to Chronic Unilateral Earplugging in Humans,” The Journal of Neuroscience, Vol.30, No.14, pp.4883–4894 (2010)

[7] 矢島春香, 大淵康成, 越智景子, “サラウンド音響と動画像のずれに対する許容度の検証,” 第 5 回 ADADA Japan 学術大会, P-12 (2018)

[8] H. Yajima, K. Ochi, and Y. Obuchi, “Learning Effect of Fore-Aft Perception of Familiar and Unfamiliar Sounds,” NICOGRAPH International 2019, Yangling, Shaanxi, China (2019)

[9] 矢島春香, 越智景子, 大淵康成, “音の瞬時的前後知覚における不一致視覚情報の影響,” 日本音響学会春季研究発表会, 1-Q-9 (2020)

[10] H. McGurk and J. MacDonald, “Hearing lips and seeing voices,” Nature, Vol.264, pp.746–748, (1976)

[11] 長田祐介, 鈴木陽一, 近藤公久. “音声と映像の空間的分離がマガーク効果に及ぼす影響に関する考察,” 日本音響学会秋季研究発表会講演論文誌, pp.351–352 (2000)

[12] P. Bertelson and G. Aschersleben, “Automatic Visual Bias of Perceived Auditory Location,” Psychonomic Bulletin & Review, Vol. 5, No. 3, pp.482–489 (1998)

[13] L. Shams, Y. Kamitani, and S. Shimoji, “What you see is what you hear,” Nature, Vol.408, No.6814, pp.788–788 (2000)

[14] K. O. Bshara, J. Grafman, and M. Hallett, “Neural Correlates of Auditory-Visual Simulus Onset Asynchrony Detection,” The Journal of Neuroscience, Vol.21, No.1, pp.300–304 (2001)

[15] D. Alais and S. Carlile, “Synchronizing to Real Events: Subjective Audiovisual Alignment Scales with Perceived Auditory Depth and Speed of Sound,” Proc. National Academy of Sciences of the United States of America, Vol.102, No.6, pp.2244–2247 (2005)

[16] K. Asakawa, A. Tanaka, and H. Imai, “Audiovisual Temporal Recalibration for Speech in Synchrony Perception and Speech Identification,” Kansei Engineering International Journal, Vol.11, No.1, pp.35–40 (2012)

[17] 小宮山摂. “視覚と聴覚による音像知覚,” 日本音響学会誌, Vol.52, No.1, pp.46–50 (1996)

[18] 飯田一博, 森本政之, “空間音響学,” 2.2, コロナ社 (2010)

矢島 春香

2018 年東京工科大学大学メディア学部メディア学科卒業．2020 年東京工科大学大学院バイオ・情報メディア研究科メディアサイエンス専攻博士前期課程修了．在学中は主にサラウンド音響を用いた研究を行う．

越智 景子

2011 年東京大学情報理工学系研究科電子情報学専攻博士課程修了．国立障害者リハビリテーションセンター研究所流動研究員 (2011-2014)．同客員研究員 (2014-2016)．国立情報学研究所特任研究員 (2015-2016)．同特任助教 (2016-2017)．2017 年より東京工科大学メディア学部助教．博士 (情報理工学)．音声合成，韻律，音声分析，音声インターフェース，言語訓練の研究に従事．

大淵 康成

1990 年東京大学大学院理学系研究科物理学専攻修士課程修了．1992 年同博士課程中退．1992 年より 2015 年まで (株) 日立製作所中央研究所および基礎研究所勤務．その間，Carnegie Mellon University 客員研究員 (2002-2003)，早稲田大学客員研究員 (2005-2010)，クラリオン (株)(2013-2015)．2015 年より東京工科大学メディア学部教授．博士 (情報理工学)．