

映像と音の時間ずれに対する違和感は 曲の種類によって異なるか？

—3種のピアノ曲による検討—

**Does the Discomfort of a Time Lag between Vision
and Audition Vary Based on Music Type?:
An Examination with Three Types of Piano Music.**

林 真由佳
Mayuka HAYASHI

(日本女子大学大学院人間社会研究科 心理学専攻博士課程前期)

要 約

本研究ではピアノ演奏動画においてどれぐらいの時間ずれのときに視聴覚の時間ずれに気づくことができるのか、また時間ずれは違和感の抱き方や好ましさに影響を及ぼすのか、そしてそれらは曲によって異なるのかを検討した。異なる3曲のピアノ演奏動画の音と映像をシステマティックにずらした時間ずれ5条件を設定し、印象評価実験を実施した。実験の結果、時間ずれ判断には曲による違いが認められなかった一方で、違和感判断と好ましさ判断においては曲による違いが認められた。一般に、映像と音にずれがあるから違和感や不快感を覚えるのだと考えられることが多いが、本研究の結果においては映像と音の時間ずれ判断と、違和感判断や好ましさ判断が必ずしも同じではないことが示された。

[Abstract]

This study examined piano-performance videos to determine whether time lags between the audio and video were perceptible, whether such time lags affect our perception of discomfort, and whether judgements regarding time lags and discomfort differ by music type. We conducted an impression evaluation experiment under five time-lag conditions, systematically shifting the piano performances' audio and visual elements in videos of three different songs. The results indicated that the judgment of time lag did not significantly vary depending on the music piece. However, differences in the judgments of discomfort were observed depending on the music. Though it is generally believed that a time lag between image and sound causes discomfort, the findings of this study demonstrate that the judgment of the time lag between image and sound is not necessarily equal to the judgment of discomfort.

はじめに

我々は、日々、複数の感覚モダリティからの情報を統合して世界を捉えている。例えば人と話すときには話者の視線や口の動きを見ながら、聞こえてくる音も併せて情報処理している。しかしながら、時に、視覚情報と聴覚情報のタイミングが一致しないことがある。近年、情報通信技術の目覚ましい発展により、オンラインでの視聴覚コンテンツやリアルタイム配信などが普及しているが、それと同時に、通信環境による遅延といった視覚情報と聴覚情報の時間ずれに晒されることとなった。筆者もYouTubeでピアノ演奏のライブ配信を視聴していた際にタイムラグが生じ、見える手の動き(映像)と聞こえるメロディ(音)の不一致に違和感を覚え、視聴をやめた経

験がある。ピアノ演奏に限らず、例えば戸巻ら(2020)はデジタルゲームの操作におけるラグは違和感や不快感を生じさせることを明らかにしている。

では時間ずれによる違和感や不快感は、どれほどのずれであるときに生じるのだろうか？

そもそも視聴覚の時間ずれに対する検知限は、刺激タイプによって異なることが知られている。例えばDixon & Spits (1980)は、ハンマーで釘を打っている刺激と人が話している刺激(スピーチ刺激)を比較して、ハンマーで釘を打っている刺激の方がずれを検知しやすく、スピーチ刺激の方がずれを検知しにくいことを示している。

楽音については、Vatakis & Spence (2006)が、スピーチ刺激、シンプルなピアノ刺激(単音と複音)、シンプルなギター刺激(単音と複音)を比較して、楽音刺激の方がスピーチ刺激よりもさらにずれが検知しにくくなったと報告している。一方で、池田 & 森下(2015)は12音または16音からなる音列パターンを持つピアノ刺激と尺八刺激を用いてスピーチ刺激と比較を行い、その結果視聴覚統合の時間窓がスピーチよりも楽器(ピアノ、尺八)の方が狭くなったと報告している。このように、楽音の場合の時間ずれ知覚については、一貫した結果は得られていない。

Vatakis & Spence (2006)の場合も、池田 & 森下(2015)の場合も、刺激は実験のために厳密に統制されたものであり、その楽器を自然に演奏した場合のものではなかった。そのため、筆者がピアノ演奏のライブ配信で感じたような実際の演奏における映像と音の時間ずれや違和感を調べるためには、自然な演奏を刺激として用いる必要がある、また、もし刺激タイプによって視聴覚ずれの検知限が異なるのであれば、単独の曲ではなく、複数の曲を比較する必要がある。

そこで本研究では、3つの異なるピアノ曲を刺激として用いて実験を行い、曲の違いによって視聴覚の時間ずれ知覚が異なるかを調べた。また、時間ずれ判断だけでなく違和感判断と好ましさの判断を併せて行い、視聴覚の時間ずれがあるから違和感が生じるのか、それとも、時間ずれを感じていなくても違和感や不快感を抱く場合があるのかについて調べた。

方法

1. 実験参加者

女子大学生18名が実験に参加した。

2. 実験刺激作成概要

表1に本研究で用いたピアノ曲とその特徴を示す。

表1. 本研究で用いたピアノ曲3種

曲の種類	拍子	速度
『エーデルワイス』	3	普通 (約 96BPM)
『きらきら星変奏曲』	4	遅め (約 84BPM)
『名探偵コナン～メイン・テーマ～』	4	速め (約 108BPM)

刺激の動画には、拍子(3拍子もしくは4拍子)や速度の異なるピアノ曲3種(『エーデルワイス』、『きらきら星変奏曲』、『名探偵コナン～メイン・テーマ～』)を用いた。この3曲は右手のメロディが単旋律であることを条件に選定された。これは映像(指の動き)と音(ピアノの音)の対応づけの難易度を難しくし過ぎないためである。また、3曲の演奏における演奏者の熟達度が同程度であることも選定の条件とした。ピアノ演奏は、アマチュア奏者(筆者、ピアノ演奏歴8年以上)が行った。演奏はヤマハのクラビノーバ(CLP-130)を用いて行った。ピアノ演奏をビデオカメラ(Sony, HDR-PJ540)で撮影・録音した。なお、奏者の手の動きおよび鍵盤を真上から撮影するため、三脚(Kenko, DVC-0302T)を使用した。

作成した動画は動画編集ソフト(Adobe premiere pro)で編集を行い、演奏の冒頭と最後が時間ずれ判断の手掛かりとならないよう、すべての動画のはじめとおわりにフェードイン・フェードアウトをつけた。動画はそれぞれの曲において任意の部分を10秒分切り取って作成した。

藤崎(2006)によると視聴覚の同時性の時間窓は成人の場合、前後100～200ms程度であるということ参考に、動画の時間ずれは、音声を8フレーム(約266ms)早くしたもの(「266ms音が先」条件)、4フレーム(約133ms)早くしたもの(「133ms音が先」条件)、映像を8フレーム(約266ms)早くしたもの(「266ms映像が先」条件)、4フレーム(約133ms)早くしたもの(「133ms映像が先」条件)、そして時間ずれのない条件(「時間ずれなし」条件)を設定し、計5条件とした。従って、刺激には合計で15種類(3曲それぞれ5条件)の動画を用いた。作成した動画はAdobe premiere pro からH.264形式で書き出してYouTubeに貼り、それをGoogle Formに埋め込んだ。

3. 動画の評価項目

動画の評価は7件法で行い、音と映像のどちらが先かを答える時間順序判断(「1. 非常に音が先－7. 非常に映像が先」)、違和感がどの程度であるかを答える違和感判断(「1. 非常に違和感がある－7. 非常に違和感がない」)、動画がどの程度好ましかを答える好ましき判断(「1. 非常に好ましくない－7. 非常に好ましい」)の3項目とした。

4. 手続き

実験はコロナ禍による感染対策を考慮してGoogle Formを用いてオンライン上で実施した。実験参加者の課題は、動画を視聴するごとにその視聴した動画の印象評価(時間順序判断、違和感判断、好ましき判断)を行うことであった。一つの動画を視聴したら動画に対して抱いた印象が一番近いところに7件法(1から7のいずれか)でそれぞれの項目にチェックをつけてもらった。この一連の流れを動画15個分で行った。動画は何度でも再生可能であり、すべての動画への印象評価が終わったら回答を送信してもらうことで実験の終了とした。

刺激は、『エーデルワイス』の時間ずれ5条件分、『きらきら星変奏曲』の時間ずれ5条件分、『名探偵コナン～メイン・テーマ～』の時間ずれ5条件分の順で提示した。Google Formの都合上、実験参加者による刺激の提示順序の違いはなかったが、各動画がどの時間ずれ条件であるかは伏せて提示した。

集めたデータはExcelに書き出し、各曲・各時間ずれ条件で実験参加者の評定値(1-7)の平均値と標準偏差を算出し、分析を行った。

結果

時間順序判断

横軸に時間ずれ条件、縦軸に時間順序判断の評定の平均値を図1に示す(1. 非常に音が先 - 7. 非常に映像が先)。系列は曲の種類を示している。「266ms 音が先」条件の時、他の条件に比べて評定値が低かった。評定値の平均が5を下回っていることから、映像が先という評価はされていなかった。また、曲の種類による大きな違いは見られず、似た推移を示していることが見て取れた。

曲の種類と時間ずれ条件の二要因分散分析を行った結果、曲の種類の主効果は有意でなく ($F(2, 34) = 0.774, n.s.$)、時間ずれ条件の主効果は有意であった ($F(4, 68) = 31.01, p < .001$)。曲の種類と時間ずれ条件の交互作用は有意でなかった ($F(8, 136) = 0.49, n.s.$)。

時間ずれ条件について Bonferroni 法で多重比較を行ったところ、「266ms 音が先」条件はすべての時間ずれ条件との組み合わせにおいて1%水準の有意差がみられ、「133ms 音が先」条件では「時間ずれなし」条件と「133ms 映像が先」条件との組み合わせの間に1%水準の有意差があった。従って、音が先であるときの条件において他の時間ずれ条件よりも音が先だと判断されていたことが示された。

曲の主効果がなかったことから、曲の拍子や速度、曲調などが変わっても(少なくとも本研究で用いた曲においては)視聴覚の時間ずれの気づきやすさには変化がなかったといえる。

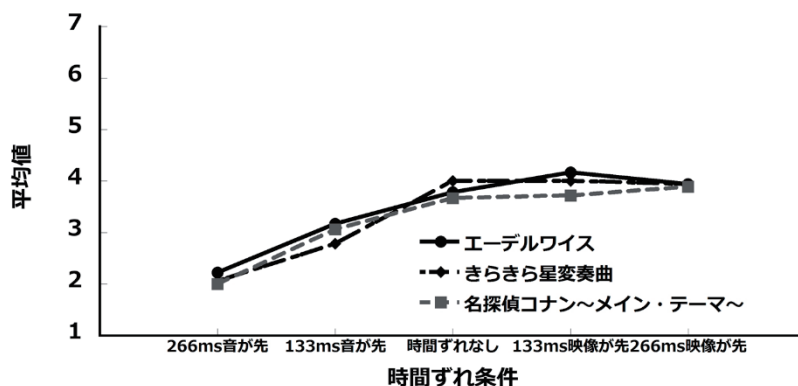


図1. 各時間ずれ条件における曲の種類ごとの時間順序判断の評定平均値

違和感判断

横軸に時間ずれ条件、縦軸に違和感判断の評定の平均値を図2に示した(1. 非常に違和感がある - 7. 非常に違和感がない)。系列は曲の種類を示している。どの曲の種類でも時間ずれ条件により評定値の差がみられた。『きらきら星変奏曲』と『名探偵コナン～メイン・テーマ～』は非常に似た推移を示しており、『エーデルワイス』は他の条件に比べてやや違和感がない判断を得られていた。

曲の種類と時間ずれ条件の二要因分散分析を行った結果、曲の種類の主効果 ($F(2, 34) = 8.12, p < .001$) と、時間ずれ条件の主効果 ($F(4, 68) = 40.89, p < .001$) は有意であったが、交互作用は有意でなかった ($F(8, 136) = 1.79, n.s.$)。

Bonferroni法で多重比較を行った結果、曲の種類では『エーデルワイス』が他の2曲(『きらきら星変奏曲』と『名探偵コナン～メイン・テーマ～』)との間に5%水準の有意差がみられ、より違和感がないと評価されていた。

時間ずれ条件では、「266ms音が先」条件において他のすべての時間ずれ条件との組み合わせの間に1%水準の有意差があり、より違和感があると評価されていた。そして、「133ms音が先」条件では「時間ずれなし」条件と「133ms映像が先」条件との間に1%水準で有意であり、より違和感があると評価されていたことが示された。

「時間ずれ」の評価は曲による違いがなかったが、「違和感」には曲による違いがあった(『エーデルワイス』が他2曲より違和感が少ない)。このことは、時間ずれの知覚と違和感の知覚が同一ではないことを示しており興味深い。

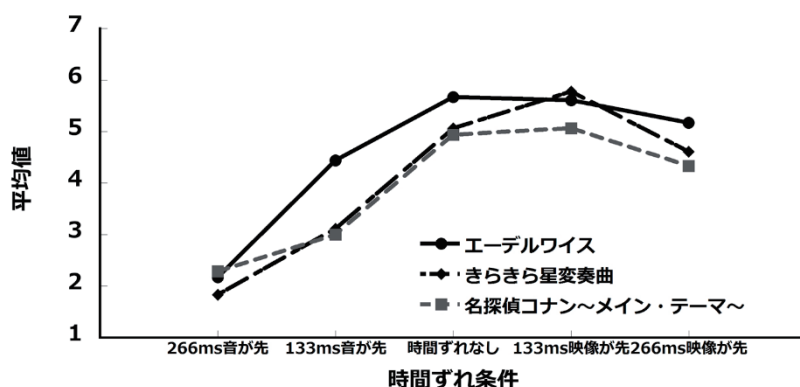


図2. 各時間ずれ条件における曲の種類ごとの違和感判断の評定平均値

好ましさ判断

横軸に時間ずれ条件、縦軸に好ましさ判断の評定の平均値を図3に示した(1. 非常に好ましくない～7. 非常に好ましい)。系列は曲の種類を示している。好ましさ判断の時間ずれ条件による違いは、違和感判断と非常によく似た推移であった。

曲の種類と時間ずれ条件の二要因分散分析を行った結果、曲の種類の主効果($F(2, 34) = 8.82, p < .007$)と、時間ずれ条件の主効果($F(4, 68) = 28.16, p < .001$)が有意であったが、交互作用は有意でなかった($F(8, 136) = 1.56, n.s.$)。

さらにBonferroni法で多重比較を行ったところ、『エーデルワイス』と『名探偵コナン～メイン・テーマ～』の間に5%水準の有意差があり、『エーデルワイス』がより好ましいと評価されていたことが示された。

時間ずれ条件の多重比較では、「266ms音が先」条件とその他すべての時間ずれ条件の組み合わせの間に1%水準で有意差があり、「266ms音が先」条件が最も好ましくないと評価されていた。「133ms音が先」条件では、「時間ずれなし」条件と「133ms映像が先」条件との組み合わせで1%水準の有意差があり、より好ましくないと評価されていたことが示された。また、「133ms映像が先」条件と「266ms映像が先」条件との間には5%水準で有意差がみられ、「266ms映像が先」条件のほ

うが好ましくないと評価されたことが伺えた。

「時間ずれ」の評価は曲による違いがなかったが、「好ましさ」には「違和感」と同様に曲による違いがあった(『エーデルワイス』が『名探偵コナン～メイン・テーマ～』より好ましい)。なお、「違和感」と「好ましさ」の評価は非常によく似ていることがグラフから伺える。

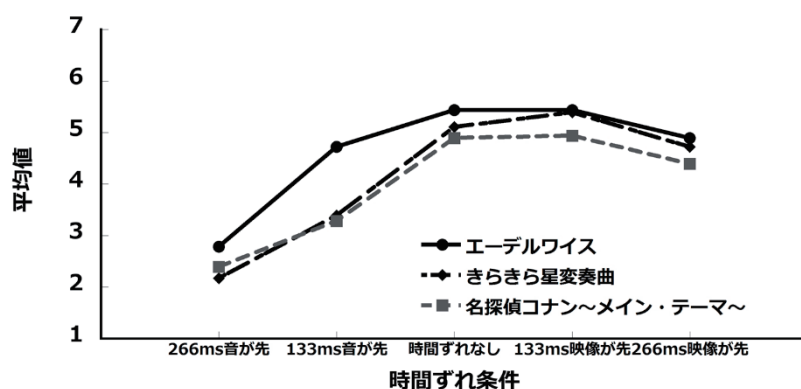


図3. 各時間ずれ条件における曲の種類ごとの好ましさ判断の評定平均値

考察

本研究の目的は、ピアノ演奏動画においてどれぐらいの時間ずれのときに視聴覚の時間ずれに気づくことができるのか、また時間ずれは違和感の抱き方や好ましさに影響を及ぼすのか、そしてそれらは曲によって異なるのかを検討することだった。このことを調べるために、異なる3曲のピアノ演奏動画の音と映像をシステムティックにずらした時間ずれ5条件を設定し、印象評価実験を実施した。

その結果、時間ずれ判断には曲による違いが認められなかったが、違和感判断と好ましさ判断においては曲による違いが認められた(図1-3)。

一般に、映像と音にずれがあるから違和感や不快感を抱くものだと考えがちであるが、本研究の結果においては映像と音の時間ずれ判断と、違和感判断や好ましさ判断が必ずしも同一ではないことが示された。曲の種類による違いがみられた違和感判断と好ましさ判断において、違和感判断では『エーデルワイス』が他2曲よりも違和感がないという評価をされており、好ましさ判断では『名探偵コナン～メイン・テーマ～』と比べると『エーデルワイス』がより好ましいという評価がされていた。

本研究では時間ずれの検知について時間順序判断(映像が先か/音が先か)を用いて調べたが、この方法は(明らかにどちらが先行しているかの判断がつくような大きな時間ずれをもつ刺激でない場合)同時性判断(同時か/非同時か)よりも難しい。映像が先か音が先かという順序の判断では曲による違いは生じなかったが、同時か非同時かという別の判断方法にして時間順序は分からなくてもずれていること自体に気づいているのかを調べれば、ずれの気づきやすさに曲による違いが生じる可能性もあり、今後の検討課題であるといえる。

また、時間ずれ条件による全体的な傾向として、「266ms音が先」条件の場合に、最も「音が先」だと検知され、違和感が生じ、かつ好ましくないと評価されることが判った。加えて、小さいずれ(約

133ms)であっても音が先行しているときの方が映像が先行しているときよりもずれに気づきやすく、違和感や好ましくないという評価をされやすいことが明らかになり、非対称性がみられることが判った。まとめると、音が先行しているときにずれに気づかれやすく、時間ずれに気づいているときその動画に対して違和感を抱き好ましくないという印象をもつことが示唆された。

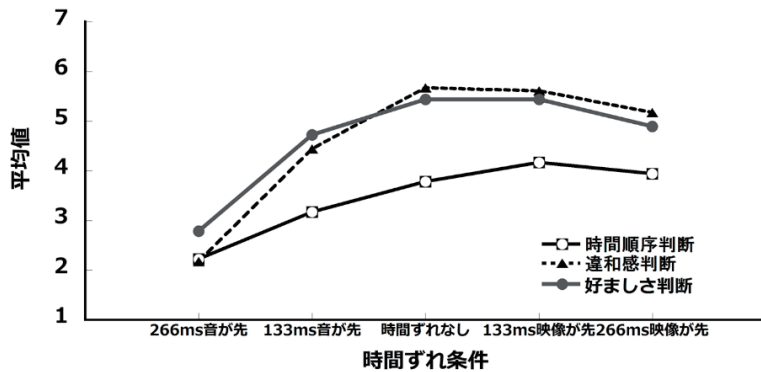


図4. 各時間ずれ条件における判断項目ごとの『エーデルワイス』への評定平均値

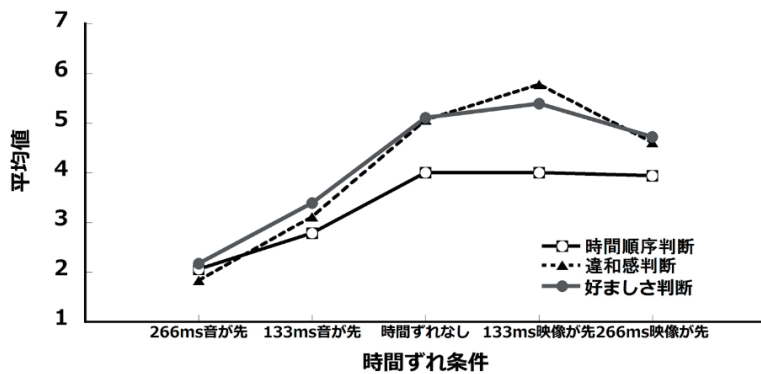


図5. 各時間ずれ条件における判断項目ごとの『きらきら星変奏曲』への評定平均値

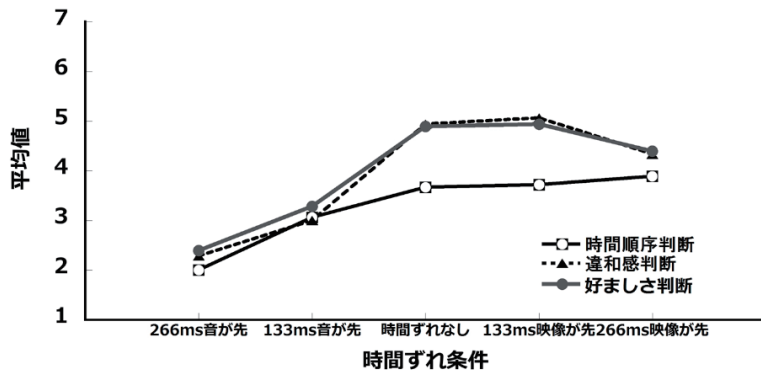


図6. 各時間ずれ条件における判断項目ごとの『名探偵コナン～メイン・テーマ～』への評定平均値

今回みられた非対称性については、題材としたピアノの構造上、鍵盤を押す前に音が鳴るのは通常ではあり得ないため、本研究の刺激に対して「音が先」だと気づきやすいのではないかという因果性の影響が考えられる。Dixon & Spits(1980)はハンマーとスピーチのどちらの刺激でも「音が先」の方が検知しやすい(「映像が先」の方に同時性の時間窓がずれる)ことを示した。一方で Van Eijk et al. (2007)は、「音がしてから視覚イベントが発生する」という因果性を持つ「ニュートンのゆりかご」と呼ばれる振り子のおもちゃを模倣した刺激を用いた実験において「映像が先」の方が検知しやすくなる(「音が先」の方に同時性の時間窓がずれる)ことを示している。

本研究の限界として、本研究はコロナ禍によりオンラインで行ったため、実験参加者や時間ずれ条件が少なく、一人あたりの試行回数も少なかった。また刺激動画は各自の電子端末で再生してもらったため、実験環境を精密にコントロールすることができなかったことも問題であった。今後はオンラインではなく対面で、精密にコントロールされた実験室実験として、引き続き検討していく必要があるだろう。また、時間順序判断だけでなく同時性判断や同期・非同期弁別といった手法を用いて検討していく必要があるだろう。

さらに、ピアノ演奏動画のずれ判断には、ピアノの演奏歴やどれほどピアノ演奏を目にしているかという実験参加者の経験も影響を与えると考えられる。今回は被験者数が少なく経験による群分けができなかったが、経験で群を分けてずれへの気づきやすさや違和感の生じ方を調べることは、例えば視聴者(ターゲット)の動画視聴の満足度を上げるために役立つと考えられる。また熟練の効果についてピアノ演奏動画以外の刺激を用いた場合でも調べていく必要があるだろう。

文献

- Dixon, N. F., & Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception*, 9(6), 719-721.
- 藤崎 和香 (2006). 聴覚情報処理のフロンティア研究と情報通信技術への応用[Ⅱ]—視聴覚の情報統合と同時性知覚—. 電子情報通信学会誌, 89(10), 906-911.
- 池田 維・森下 正修 (2015). 視聴覚同期知覚における刺激種類の効果, 日本心理学会大会発表論文集, 79, 584.
- 戸巻 晃大・遠藤 雅伸 (2020). デジタルゲームの操作におけるラグの不快感に関する研究. 映像情報メディア学会技術報告, 44(10), 151-153.
- Van Eijk, R. L. J., Kohlrausch, A. G., Juola, J. F., & van de Par, S. L. J. D. E. (2007). Causal relationships affect audio-visual asynchrony detection : opposite trends for different stimuli. In *Proceedings of the 8th Annual International Multisensory Research Forum Meeting (IMRF 2007)*, July 5-7, University of Sydney.
- Vatakis, A., & Spence, C. (2006). Audiovisual synchrony perception for speech and music assessed using a temporal order judgment task. *Neuroscience Letters*, 393, 40-44.