二重唱の歌い出しタイミングに対する同時性知覚の刺激閾調査*

☆兵藤弘明 (東大院・情報理工), 高道慎之介 (慶應義塾大/東大院・情報理工),
猿渡洋 (東大院・情報理工)

1 はじめに

重唱は，複数の歌唱者が互いに異なる声部を担当し，声を合わせて歌唱する歌唱形態であり，コミュニケーションや芸術表現の手段として多くの人に親しまれている．一体感のある重唱の実現には，楽曲内の然るべき箇所において，重唱グループを構成する者の歌い出しが互いに揃うことが必要である．しかしながら，具体的にどの程度歌い出しが揃っていれば一体感が生まれるのかは明らかでない．重唱における歌い出しの時間差の許容度合いがわかれば，重唱練習の指針となるだけでなく，許容度合いを歌声の客観評価指標として応用することも可能である．

そこで本研究では，特に二声部からなる重唱（二重唱）の歌い出しに注目し，聴取者が「揃っている」と判断する歌い出しの時間差の閾値，すなわち歌い出しの同時性知覚の刺激閾の推定を行う．本稿では刺激閾の推定手法について述べた後，主観評価実験により刺激閾を推定し，刺激閾の傾向について議論する．

2 関連研究

2.1 音の開始タイミングの決定方法

複数の音の開始タイミングの同時性知覚の刺激閾を推定する上では，各音の開始タイミング間の時間差を算出する必要がある．

これを算出する方法の一つは，各音の開始タイミングを決定し，時間差を直接計算することである．この方法は，開始タイミングの決定が容易なアタックの強い音を扱う場合 [1] や，電子楽器を用いて演奏タイミングを取得する場合 [2] に用いられる．一方，歌声はアタックの強弱のばらつきが大きく，音の開始タイミングが明確でないため，各歌唱者の歌声の開始タイミングを正確に決定することが困難である．

そこで本研究では，はじめに各声部の歌い出しの同時性知覚の主観的同時点 [3]，すなわち二つの刺激が最も同期していると主観的に知覚される際の，二つの刺激間の時間間隔を算出する．次に，算出した主観的同時点を歌い出しが同期する時間間隔であると仮定し，同時性知覚の刺激閾を推定する．

2.2 歌声以外の音における同時性知覚について

合奏における音の同時性知覚については，ピアノの合奏において 100 ms 以内の時間差はずれとして指摘されにくいこと [2] や，時間差が 15 から 20 ms 以上であれば二つの打鍵音の演奏順を指摘できること [4] が明らかにされており，歌声の場合における同時性知覚の刺激閾も類似した値となることが予想される．一方で，楽器音ごとに同時性の知覚が異なること [1] が知られており，歌声においても他の楽器音とは異なる同時性知覚の傾向があることが予想される．

音の同時性知覚の調査においては，音の出だしに着目することが多い．例えば，Hirsh [4] はピアノの打鍵音やクリック音などを用いて音の出だし時刻の同時性知覚を調査している．これに倣い，本研究でも歌声の歌い出し時刻に着目する．

3 実験的調査

本実験では，主旋律と副旋律からなる二重唱音声の歌い出しを対象に調査を行い，同時性知覚の主観的同時点の算出，および歌い出しの同時性知覚の刺激閾の推定を行う．

3.1 実験手順

3.1.1 知覚確率の算出方法

主観的同時点の算出および刺激閾の推定のため，はじめに歌い出しの呈示間隔（Stimulus Onset Asynchrony, SOA）ごとに,「歌い出しが揃っている」と回答された割合（知覚確率）を求める．ここで歌い出しの呈示間隔とは，声部間の歌い出しの時間差を指す．呈示間隔を求めるため，本実験では予め各声部について音素アライメントを実施する．呈示間隔は，音素アライメントから定めた副旋律の歌い出し時刻から主旋律の歌い出し時刻を引いた値とする．

具体的な実験手順は以下の通りである．まず，主旋律の歌声に対し，副旋律の歌い出し時刻を時間方向にずらして加算することで，呈示間隔の異なる二重唱音声を作成する．呈示間隔は複数のパターンを用意し，互いに等間隔になるように設定する．次に，これらの歌声をランダムな順で被験者に提示する．被験者は歌声を聴取し，歌い出しが揃っているか否かの二択を回答する．最後に複数の被験者の回答から,「揃っている」と回答された割合（知覚確率）を呈示間隔ごとにを計算する．以降では知覚確率の値を用いて，主観的同時点の算出と歌い出しの同時性知覚の刺激閾の推定を行う．

3.1.2 主観的同時点の算出手法

主観的同時点の算出方法を説明する．まず， 3.1.1 節で求めた知覚確率を呈示間隔の関数とみなし，正規分布関数でフィッティングを行う．推定した正規分布関数の平均値に対応する呈示間隔を主観的同時点とする．図 1 に，ある二重唱音声に対する被験者の回答結果と，回答結果から算出した主観的同時点を例として示す．この例では，算出した主観的同時点の値は 16 ms である．これは，音素アライメントから定めた歌い出し時刻から 16 ms 遅れた時点に主観的

---

*Investigation of stimulus threshold to simultaneity perception for onset of duet singing. by HYODO, Hiroaki (The University of Tokyo), TAKAMICHI, Shinnosuke (Keio University/The University of Tokyo), SARUWATARI, Hiroshi (The University of Tokyo).

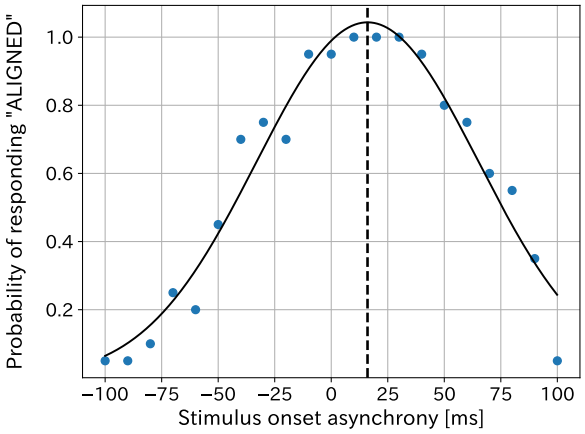

Fig. 1　ある二重唱音声に対する被験者の回答結果．破線は算出した主観的同時点を表す．

同時点があることを意味する．

3.1.3　刺激閾の推定手法

同時性などの知覚刺激の刺激閾を推定する手法はいくつか存在するが，本研究では恒常法 [5] を使用する．

恒常法ではまず，3.1.1 節で述べたように，歌い出しの呈示間隔の異なる二重唱の歌声を被験者に提示し，知覚確率を呈示間隔ごとに計算する．次に，知覚確率を 3.1.2 節で計算した主観的同時点からの時間差の関数とみなし，主旋律が副旋律に対して先行する場合と後行する場合のそれぞれについて，片側正規累積分布関数でフィッティングを行い，時間差と知覚確率の関係を表す関数を得る．

この関数上で知覚確率が 0.5 となる時間差，すなわち被験者の半数が歌い出しが「揃っている」と判断すると予測される時間差を刺激閾の推定値とする．

3.2　実験条件

3.2.1　実験に使用する歌声

本実験では，日本語の重唱音声コーパスである jaCappella コーパス [6] の歌声を使用した．jaCappella コーパスの重唱は計 6 つの声部により構成されるが，本実験ではリードボーカル（主旋律）とソプラノ（副旋律）の二声部の歌声を使用した．歌唱者の性別は全て女性であり，歌唱者の組み合わせは計 4 通りであった．また，これらの歌声のサンプリング周波数は 48000 Hz であった．

実験で使用する歌声は以下のように作成した．はじめに，二名の歌唱者が歌い出しで同じ歌詞を歌唱する箇所に注目し，歌い出しとその前後を各声部の歌声から切り出す．次に，歌い出しに後続する歌詞の影響を排除するため，歌い出しの歌声のうち，先頭のモーラが歌唱されている部分をロングトーンに変換する．ロングトーンへの変換手順は 3.2.2 節に後述する．最後に変換した各声部の歌声のうち，副旋律の歌声を時間方向にずらして主旋律の歌声に足し合わせることで，歌い出し時刻が声部間で異なる二重唱音声を作成する．この際，歌唱部分の時間長が 3 秒となるようにした．また，歌い出し以外の部分が評価に影響することを防ぐため，最後の 1 秒間でロングトーンの歌声をフェードアウトさせた．

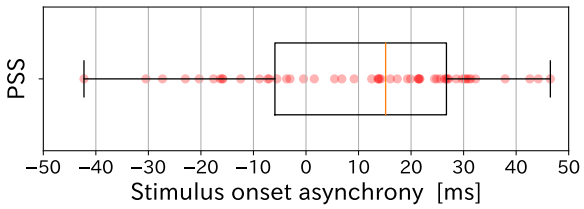

Fig. 2　各二重唱音声の主観的同時点（Point of Subjective Simultaneity, PSS）の算出結果．横軸は音素アライメントにより得た呈示間隔に対応する．

本実験では，計 52 個の二重唱音声ごとに，21 通りの呈示間隔の歌声を作成して使用した．使用した 21 通りの呈示間隔は，−100 ms から 100 ms まで 10 ms 刻みとなるように設定した．

3.2.2　ロングトーンへの変換

歌い出しの歌声のうち，先頭のモーラが歌唱されている部分をロングトーンに変換する方法を説明する．まず WORLD [7] (D4C edition [8]) を用いて歌声の音響特徴量（基本周波数，スペクトル包絡，非周期性指標）の系列を抽出する．次に，歌い出しの歌詞に対応する音素列の中で，最初に出現する母音を歌唱する部分に対応する適当な 1 フレームを時間方向に繰り返すことで，変換後の歌声に対応する音響特徴量を作成する（例えば，歌いだしの歌詞に対応する音素列が /k a/ であれば /a/の部分に対応する音響特徴量を繰り返す）．最後に再度 WORLD を用いて，この音響特徴量から歌声を再合成する．

3.2.3　主観的同時点の算出と刺激閾の推定

作成した二重唱音声を用いて主観評価実験を行い，主観的同時点の算出と同時性知覚の刺激閾の推定を行った．評価者は Lancers[1]で募集したクラウドワーカである．52 個の二重唱音声ごとに 20 名の評価者が参加した．

はじめに主観的同時点の算出を行った後，主旋律が副旋律に対して先行する場合（歌い出しの時間差が正の場合）の刺激閾と，主旋律が副旋律に対して後行する場合（歌い出しの時間差が負の場合）の刺激閾をそれぞれ推定した．

3.3　実験結果

3.3.1　主観的同時点の算出結果

はじめに，音素アライメントにより得た呈示間隔が 0 となる点と，算出した主観的同時点のずれを確認する．このずれの絶対値が極端に大きい場合，刺激閾の推定に用いる呈示間隔の数が極端に少なくなり，刺激閾の推定精度を損ねる恐れがある．図 2 に，各二重唱音声の主観的同時点と，音素アライメントにより得た呈示間隔の関係を示す．この結果より，主観的同時点は約 15 ms を中心とし，±50 ms の範囲内に収まっていることがわかる．従って，本実験の呈示間隔が −100 ms から 100 ms までであることを踏ま

[1] https://www.lancers.jp/

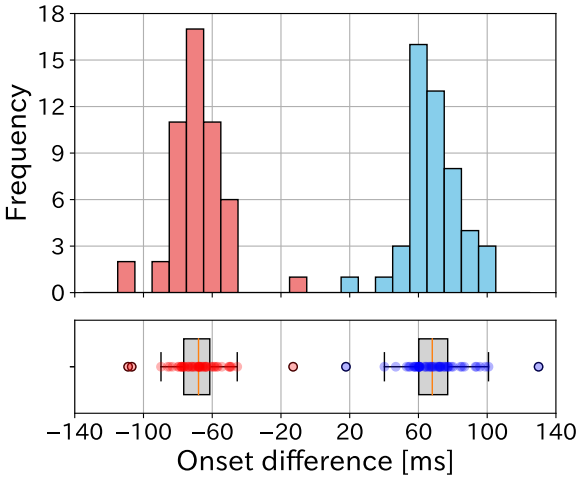

Fig. 3　各二重唱音声の刺激閾の推定結果．主観的同時点の前後の回答から推定した刺激閾をそれぞれ赤色と青色で示した．歌い出しの時間差（Onset differences）は，主観的同時点の呈示間隔を 0 ms としたときの，副旋律の歌い出し時刻と主旋律の歌い出し時刻の差である．

Table 1　音素の分類カテゴリ

音素カテゴリ	音素
母音	/a, e, i, o, u/
無声子音	/ch, f, h, hy, k, ky, p, py, s, sh, t, ts/
有声子音	/b, by, d, g, gy, j, m, my, n, ny, r, ry, w, y, z/

えると，刺激閾は少なくとも 50 ms 以上の時間幅から推定されることとなる．

3.3.2　刺激閾の推定結果

次に，各二重唱音声の刺激閾の推定値を図 3 に示す．主旋律が副旋律に対して後行する場合（歌い出しの時間差が負の場合）の刺激閾はおおむね 50 ms から 90 ms，先行する場合（歌い出しの時間差が正の場合）の刺激閾はおおむね 40 ms から 100 ms の範囲に分布した．

なお，推定した二つの刺激閾の少なくとも一方の値が 0 ms 以下または 1000 ms 以上であった二重唱音声は外れ値とみなし，両方の刺激閾の値を除外した．この処理により，52 個のうち 2 個の二重唱音声に対する分析結果が除外された．

3.3.3　刺激閾の分析

楽器音ごとに同時性の知覚が異なることから，歌い出しの音素の場合も同様に異なる知覚傾向があることが予想される．そこで，歌い出しの音素の種類と刺激閾の関係を調査する．表 1 に従い，歌い出しの音素を母音，無声子音，有声子音に分類し，それぞれの刺激閾を描画したものを図 4 に示す．

まず，歌い出しの音素が母音である場合の刺激閾は比較的狭い範囲に，歌い出しの音素が無声子音である場合の刺激閾は比較的広い範囲に分布する傾向が見られた．また，歌い出しの音素が有声子音である場合，刺激閾の値は他の場合と比べやや小さくなる傾向が見られた．

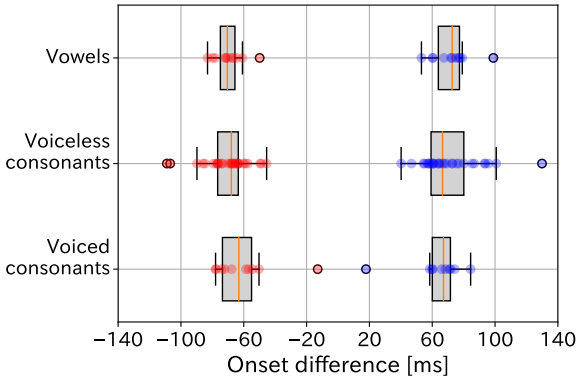

Fig. 4　歌い出しの音素がそれぞれ母音，無声子音，有声子音である場合の刺激閾．歌い出しが母音，無声子音，有声子音である二重唱音声の数はそれぞれ 11, 29, 10 であった．

次に各二重唱音声について，主旋律が副旋律に対して先行する場合と後行する場合の刺激閾の大きさを図 5 に示す．二つの刺激閾間の相関係数は 0.79 であり，正の相関が見られた．このことから，二重唱音声の刺激閾は声部について対称性があると考えられる．

最後に，jaCappella コーパスの二声部の呈示間隔が刺激閾内に含まれるか，すなわち，歌い出しが「揃っている」と判断される時間差の範囲内にあるかを検証する．当該コーパスのミキシングはプロにより行われていることから，コーパスの二重唱音声の歌い出しは十分に揃っていることが期待される．jaCappella コーパスの二声部の呈示間隔が刺激閾の範囲内に含まれるのであれば，本稿で推定した刺激閾には妥当性があると考えられる．各二重唱音声の刺激閾と，jaCappella コーパスにおける歌い出しの時間差を図 6 に示す．まず，全ての二重唱音声について，jaCappella コーパスの歌い出しの時間差は二つの刺激閾の範囲内に含まれることがわかる．また，jaCappella コーパスの歌い出しの時間差の絶対値の平均値は 20.29 ms，最大値は 46.51 ms であり， 3.3.2 項で示した刺激閾の分布範囲より小さい値であった．以上より，本稿で推定した刺激閾の妥当性を確認できる．

4　おわりに

本稿では二重唱音声の歌い出しのタイミングに注目して同時性知覚の刺激閾の推定を行い，刺激閾はおおむね 40 ms から 100 ms の範囲であること，歌い出しの音素の種類により刺激閾の傾向が異なることを明らかにした．また，刺激閾が声部について対称であることを明らかにした．

今後は，得られた知見を元に歌声の客観評価指標を設計し，深層学習を用いた重唱合成モデルの学習・評価への応用に取り組む．また，歌唱者人数が異なる

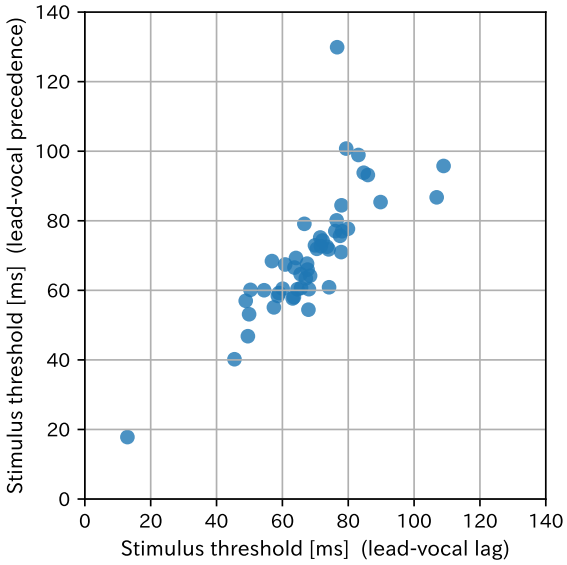

Fig. 5　各二重唱音声に対応する二つの刺激閾．横軸は主旋律が副旋律に対して後行する場合の刺激閾に，縦軸は主旋律が副旋律に対して先行する場合の刺激閾に対応する．

重唱音声の評価や．歌い出し以外の歌唱タイミングについても同様の実験を実施することを検討している．

謝辞　本研究は JST 創発的研究支援事業 JPMJFR226V，JSPS 科研費 23H03418，23K18474 の助成を受けた．

参考文献

[1] R. A. Rasch, “Synchronization in performed ensemble music,” *Acta Acustica united with Acustica*, vol. 43, no. 2, pp. 121–131, 1979.

[2] 堀内靖雄 et al., “二人の人間による演奏の収録と分析,” **情報処理学会研究報告音楽情報科学** *(MUS)*, vol. 1996, no. 53 (1996-MUS-015), pp. 21–26, 1996.

[3] J. Stone et al., “When is now? perception of simultaneity,” *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 268, no. 1462, pp. 31–38, 2001.

[4] I. J. Hirsh, “Auditory perception of temporal order,” *The Journal of the Acoustical Society of America*, vol. 31, no. 6, pp. 759–767, 1959.

[5] 黒田剛士, 蓮尾絵美, “早わかり心理物理学実験 (やさしい解説),” **日本音響学会誌**, vol. 69, no. 12, pp. 632–637, 2013.

[6] T. Nakamura et al., “jaCappella corpus: A Japanese a cappella vocal ensemble corpus,” 2023.

[7] M. Morise et al., “WORLD: A vocoder-based high-quality speech synthesis system for real-time applications,” vol. 99, no. 7, pp. 1877–1884, 2016.

[8] M. Morise, “D4C, a band-aperiodicity estimator for high-quality speech synthesis,” *Speech Commun.*, vol. 84, pp. 57–65, 2016.

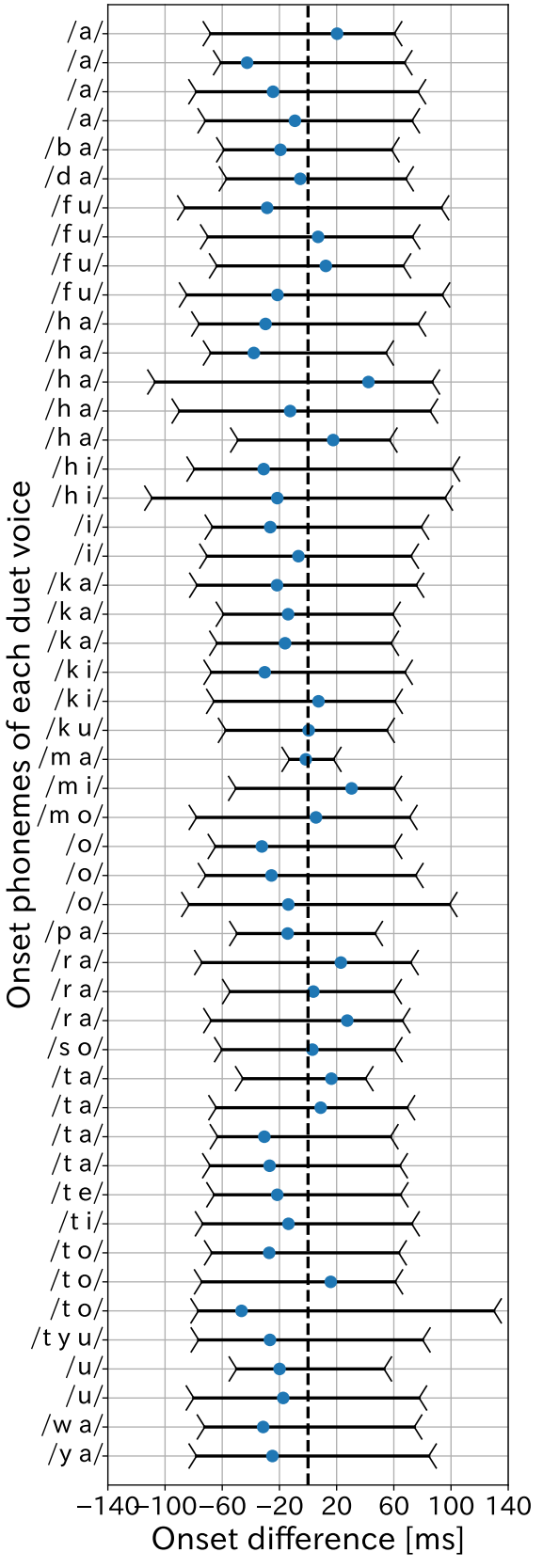

Fig. 6　各二重唱音声の刺激閾と，jaCappella コーパスにおける歌い出しの時間差．数直線は二つの刺激閾間の範囲を表し，歌い出しが「揃っている」と判断される時間差の範囲に相当する．また，青丸は jaCappella コーパスにおける歌い出しの時間差を示す．