

# 振幅変調に基づく音脈分凝の知覚交替特性

山岸慎平<sup>1,a)</sup> 大塚翔<sup>2</sup> 古川茂人<sup>2</sup> 柏野牧夫<sup>1,2</sup>

**概要:** 特徴量の異なる 2 音 (A 音, B 音) と無音 (-) を組み合わせた ABA-音列を聴取すると, 2 つの異なる知覚状態が生じることが知られている. ひとつは 2 音をまとめる聞こえ方で, もうひとつは 2 音を分離する聞こえ方である. この ABA-音列を繰り返し呈示すると聴取者の知覚状態がこれら 2 つの間で切り替わる. この現象は曖昧な聴覚情景 (auditory scene) を探る脳内の処理を反映するものであると考えられる. 我々は, この音脈の知覚交替現象の特性が音脈分凝を引き起こす特徴量の種類に依存するかどうかを調べた. 2 音間の周波数差を手掛かりとした場合と振幅変調の違いを手掛かりとした場合について, 知覚交替の特性を比較した結果, 知覚交替回数が特徴に依存することがわかった.

YAMAGISHI SHIMPEI<sup>1,a)</sup> OTSUKA SHO<sup>2</sup> FURUKAWA SHIGETO<sup>2</sup> KASHINO MAKIO<sup>1,2</sup>

## 1. はじめに

周波数の異なる 2 音 (A 音, B 音) と無音 (-) を組み合わせた ABA- 音列を聴取すると, 2 種類の知覚が生じることが知られている [1], [2]. ひとつは ABA-ABA-... のように 2 音をまとめる聞こえ方 (音脈統合, 1 stream) で, もうひとつは A-A-... と -B-... のように 2 つの音脈が分離する聞こえ方 (音脈分凝, 2 streams) である (図 1). 本研究では前者を S1, 後者を S2 と呼ぶ. この知覚は 2 音の物理特性によって決まることが知られており, 2 音間の周波数差が大きい, あるいは呈示間隔が短い場合, S2 になりやすく, 周波数差が小さいあるいは呈示間隔が長い場合, S1 になりやすい [1]. しかし, 物理特性が一定の場合でも, ABA-音列を繰り返し呈示すると, 聴取者の知覚が S1 と S2 の間で確率的に切り替わる現象が起きる. この現象は曖昧な聴覚情景 (auditory scene) を探る脳内の処理を反映するものであると考えられる. ABA-音列による音脈分凝は, 周波数手掛かりを含まない場合でも起こることが知られている [3]. 今回は, 振幅変調 (amplitude modulation: AM) を手掛かりとした音脈の知覚交替現象に着目した [4], [5]. 搬送周波数は一定で変調周波数の異なる 2 つの AM 音を用いた場合, 音脈知覚の変化は聴覚末梢の出力だけでは説明が

できず, 周波数を手掛かりとした音脈知覚とは異なるメカニズムが関わる可能性が考えられる. 本研究では, 周波数を手掛かりとした場合 (Spectral condition) と振幅変調を手掛かりとした場合 (Temporal condition) の知覚交替の特性を比較することで, 物理的な手掛かりに依存して音脈の知覚メカニズムが異なるのか否かを検討した.

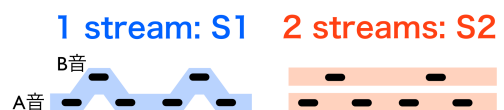


図 1 ABA-音列と 2 種類の知覚.

Fig. 1 ABA- sequence and two types of percepts

## 2. 実験

### 2.1 実験参加者

正常聴力を有する男女 13 名が実験に参加した.

### 2.2 実験刺激

各条件について, 実験で使った音の音圧レベルは約 75 dBA となるようにした. 音の長さは 50 ms, 音のオンセットから次の音のオンセットまでの時間 (Stimulus-onset asynchrony: SOA) は 110 ms であった. ABA-音列一回分の長さは 440 ms であった.

#### 2.2.1 周波数手掛かり条件: Spectral condition

A 音の周波数は 315 Hz に固定し, 2 音間の周波数差は 2 半音 ~ 10 半音まで, 2 半音ずつ変化させた (図 2). すべ

<sup>1</sup> 東京工業大学  
Tokyo Institute of Technology, Yokohama, Kanazawa, Japan  
<sup>2</sup> NTT コミュニケーション科学基礎研究所  
NTT Communication Science Laboratories, NTT Corporation, Atsugi, Kanagawa, Japan  
<sup>a)</sup> yamagishi@u.ip.titech.ac.jp

ての条件において、B音の周波数はA音の周波数を中心とした聴覚フィルタの帯域幅 (ERB: Equivalent rectangular bandwidth) の外側にあることがわかる。ERBは以下の式(1)から求めた [7]。ただし、周波数  $f$  の単位は kHz, ERBの単位は Hz である。

$$ERB(f) = 24.7(4.37f + 1) \quad (1)$$

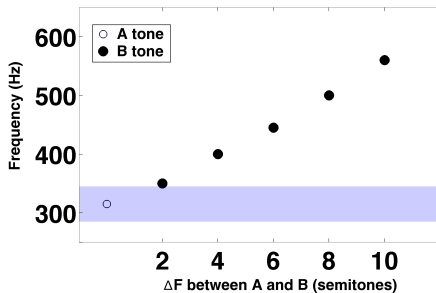


図2 周波数手掛り条件でのA音とB音の周波数。紫のバンドはA音の周波数を中心としたERBを表す。

Fig. 2 The frequency of A and B tones in spectral condition. The colored band represents the ERB at the A-tone frequency.

### 2.2.2 時間手掛り条件: Temporal condition

各音の変調度は100%, 搬送周波数は4 kHzであった。A音の変調周波数は60 Hzに固定し、A音とB音の変調周波数差は0.5~2.5オクターブまで、0.5オクターブ刻みで変化させた(図3)。変調周波数差が2.5オクターブの場合、搬送周波数(4 kHz)を中心とした聴覚フィルタの幅(ERB)から変調周波数成分が出ているが、それ以外の場合には概ねフィルタの中に収まっており、周波数手掛りよりも音の振幅の時間変化の方が主な手掛りとなる。

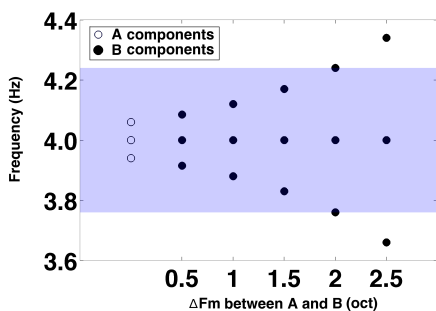


図3 時間手掛り条件でのA音とB音の周波数成分。搬送(キャリア)周波数は4 kHz。紫のバンドは搬送周波数を中心としたERBを表す。

Fig. 3 The frequency components of A and B sounds in temporal condition. The colored band represents the ERB at the carrier frequency (4 kHz).

## 2.3 実験手順

実験参加者は各条件について4~5回ずつ課題を行った。すべての課題はランダムに実施した。1回のセッションで、ABA-音列は400回繰り返し呈示され、参加者はパソコンのキーボードを使って知覚交替のタイミングを報告した。1セッションの長さは約176秒であった。結果は、各条件について4~5回分のデータを平均した値から算出した。

## 3. 結果

### 3.1 知覚確率

周波数手掛り条件、時間手掛り条件どちらについても、特徴量の差が大きくなるほどS2知覚確率が上昇することがわかった(図4)。また、各条件について、音の呈示開始から数十秒かけて徐々にS2知覚確率が上昇している。これはBuildupと呼ばれ、過去の研究結果とも一致している [3], [6]。

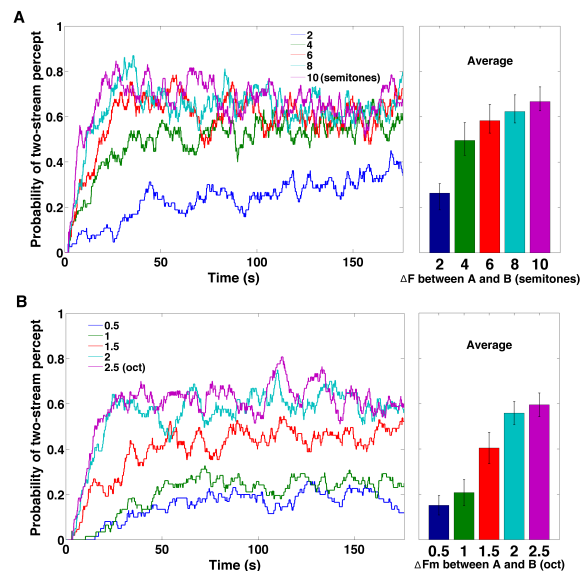


図4 S2知覚確率の時間変化(全参加者の平均値)。(A)周波数手掛り条件 (B)時間手掛り条件

Fig. 4 Time course of probability of S2 percept (grand average). (A) Spectral condition (B) Temporal condition

### 3.2 音脈分凝を引き起こす特徴量の差の効果の比較

同じ知覚をもたらすのに必要な特徴量の差を条件間で比較するため、各特徴量軸(周波数軸と振幅変調周波数軸)について、全参加者の平均値からS2知覚確率が0.3, 0.4, 0.5, 0.6となる時の特徴量の差( $\Delta F, \Delta F_m$ )を求めた(図5AB)。単位は両軸共にオクターブに合わせた。各点の間の値は線形補完によって計算した。また時間手掛り条件では、知覚確率が0.6になる値が得られなかったため、この点のみ線形的に外挿補完をした。周波数手掛り条件では、 $\Delta F_{0.3} = 0.195, \Delta F_{0.4} = 0.27, \Delta F_{0.5} = 0.344, \Delta F_{0.6} = 0.535$ , 時間手掛り条件では、 $\Delta F_{m0.3} = 1.259, \Delta F_{m0.4} =$

1.494,  $\Delta F_{m0.5} = 1.831$ ,  $\Delta F_{m0.6} = 2.559$ であった。これらの関係性を調べるため回帰分析を行った結果、両者は高い相関関係にあり、回帰係数は4.72であった。特徴の差がなければ ( $\Delta F = \Delta F_m = 0$  の場合), S2になる確率はどちらの条件でも0になるはずなので、回帰分析の際、点 (0,0) を追加した。

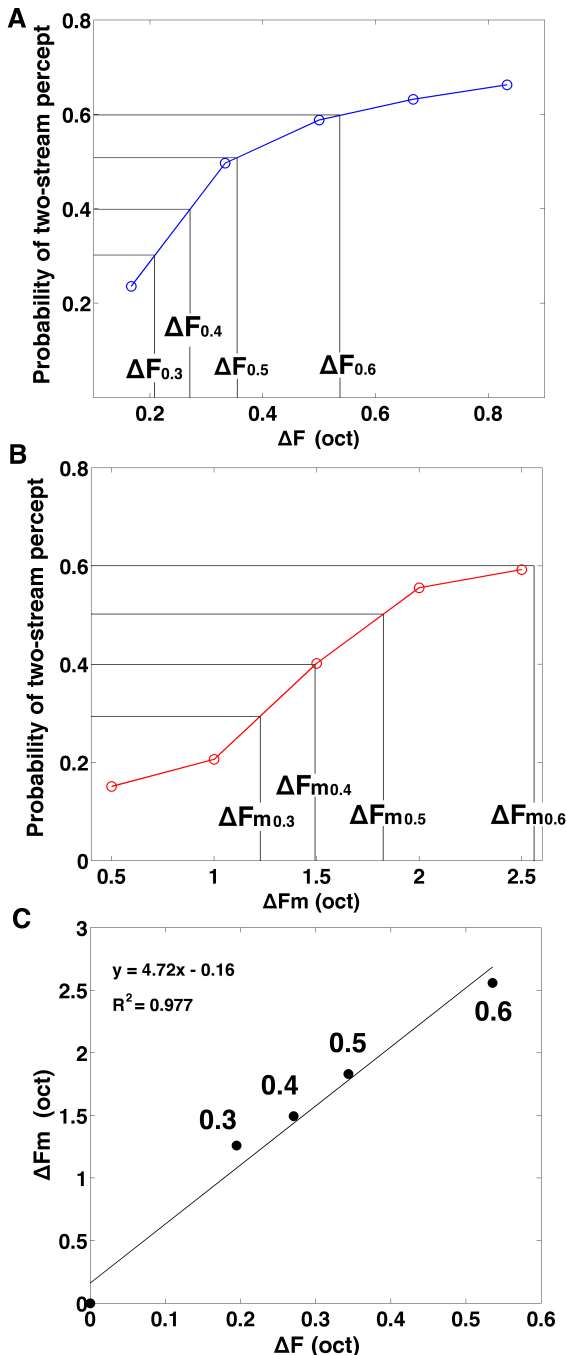


図5 特徴量の差と S2 知覚確率の関係 (全参加者平均). (A) 周波数手掛かり条件 (B) 時間手掛かり条件 (C) 同じ知覚をもたらす特徴量の差 ( $\Delta F$  と  $\Delta F_m$ ) の関係

Fig. 5 The relationship between feature separation and probability of S2 percept. (A) Spectral condition (B) Temporal condition (C) The relationship of feature separation between two conditions

### 3.3 知覚交替の特性

知覚確率が等しくなる範囲内で知覚交替の回数を比較した。具体的には、周波数手掛かり条件については  $\Delta F = 2 \sim 8$  半音の4条件、時間手掛かり条件については  $\Delta F_m = 1 \sim 2.5$  オクターブの4条件について比較を行った (図6)。2 (条件)  $\times$  4 (特徴量の差) の被験者内分散分析 (repeated-measures ANOVA) を行った結果、条件間の主効果 ( $F(3,36)=6.492$ ,  $p=0.0013$ ) が検出された。これは時間手掛かり条件の方が知覚交替回数が少ないことを意味している。交互作用は検出されなかった ( $F(3,36)=1.934$ ,  $p=0.1415$ )。

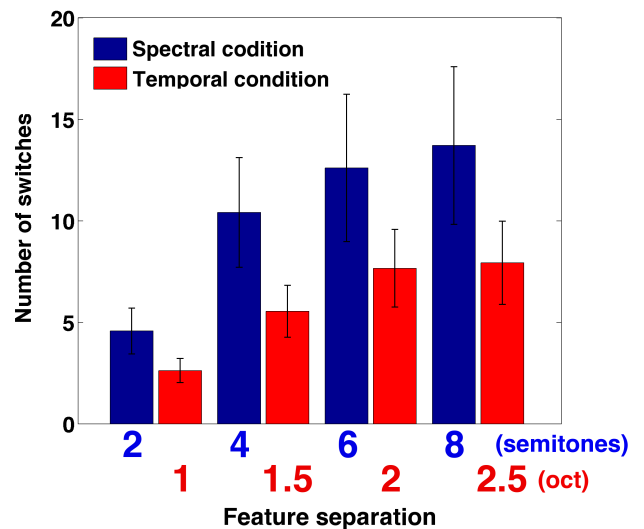


図6 知覚交替回数の比較。青いバーは周波数手掛かり条件、赤いバーは時間手掛かり条件を表す。エラーバーは標準誤差を表す。

Fig. 6 Comparison of the numbers of switching. Blue bars represent spectral condition. Red bars represent temporal condition. Error bars represent the standard error of mean.

## 4. 考察

### 4.1 時間手掛かりによる音脈分凝

周波数手掛かりを含まない振幅変調音によって音脈分凝が起きることが確認された。この結果はこれまでの研究と一致している [4], [5]。

### 4.2 特徴次元による音脈分凝の効果の違い

周波数手掛かりによる音脈分凝は末梢の周波数チャンネル間の距離によってある程度説明できる (peripheral channeling theory: [8], [9])。振幅変調を手掛かりとした音脈分凝はそのような末梢の周波数分解では説明できないが、変調周波数軸でも周波数軸での聴覚フィルタに似たフィルタがあることが報告されている。振幅変調による音脈分凝は、この変調周波数軸でのフィルタによって説明できる可能性がある。これは modulation フィルタと呼ばれ、Q 値

(フィルタの鋭さを表すパラメータ)が1-2の間に収まる  
ことが報告されている [10], [11], [12]. 対して, 本研究の  
周波数範囲内で, 聴覚フィルタのQ値をERBから概算す  
ると5~7と推定できる. つまり, 聴覚フィルタに比べて,  
modulation フィルタは緩く, バンド幅が広い. 本研究で得  
られた知覚的に等価な2つの特徴量の差の間の回帰係数は  
約4.7であったが, これは聴覚フィルタと modulation フィ  
ルタのQ値の比と良く一致している. よって, 特徴次元に  
よる音脈分凝の効果の違いは, 各次元のフィルタの幅の違  
いで説明できる.

#### 4.3 特徴量による知覚交替の回数の違い

各条件について平均知覚確率が等しくなる範囲で比べた  
結果, 時間情報を手掛かりとした場合の方が知覚交替の回  
数が少なかった. この結果は音脈分凝の処理モデルを考え  
るうえで新たな情報となりうる.

#### 参考文献

- [1] van Noorden, L.P.A.S.: *Temporal Coherence in the Perception of Tone Sequences*, Eindhoven University of Technology, Eindhoven, The Netherlands (1975).
- [2] Bregman, A. S.: *Auditory Scene Analysis*, MIT Press (1990).
- [3] Moore, B.C.J., Gockel, H.E.: Properties of auditory stream formation, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 367, 919-931 (2012).
- [4] Grimault, N., Bacon, S.P., Micheyl, C.: Auditory stream segregation on the basis of amplitude modulation rate, *J. Acoust. Soc. Am.*, 111, 1340-1348 (2002).
- [5] Dollezal, L.V., Beutelmann, R., Klump, G.M.: Stream segregation in the perception of sinusoidally amplitude-modulated tones. *Plos One*, 7, e43615 (2012).
- [6] Bregma, A. S.: Auditory streaming is cumulative. *J. Exp. Psychol.: Human Percept. Perf.* 4, 380-387 (1978).
- [7] Glasberg, B.R., Moore, B.C.J.: Derivation of auditory filter shapes from notched-noise data, *Hear. Res.*, 47, 103-138 (1990).
- [8] Beauvois, M.W., Meddis, R.: Computer simulation of auditory stream segregation in alternating-tone sequences, *J. Acoust. Soc. Am.*, 99, 2270-2280 (1996).
- [9] Hartmann, W.M., Johnson, D. Stream segregation and peripheral channeling, *Music Perception*, 9,155-184. (1991).
- [10] Ewert, S.D., Dau, T.: Characterizing frequency selectivity for envelope fluctuations, *J. Acoust. Soc. Am.*, 108, 1181-1196 (2000).
- [11] Sek, A., Moore, B.C.J.: Testing the concept of a modulation filter bank: The audibility of component modulation and detection of phase change in three-component modulators, *J. Acoust. Soc. Am.*, 113, 2801-2811 (2003).
- [12] Moore, B.C.J., Füllgrabe, C., Sek, A.: Estimation of the center frequency of the highest modulation filter, *J. Acoust. Soc. Am.*, 125, 1075-1081 (2008).