

音のピッチ知覚

大串健吾^{†1}

音のピッチ知覚については長い間実験と議論が行われてきた。時代の変遷と共に、聴覚理論は変化してきた。この報告では、聴覚理論の変遷を簡単に紹介する。また複合音のピッチがその基本周波数に対応する純音のピッチとはわずかに異なることを示す実験データを取り上げ、この現象をどのように説明すべきかについて新しい観点から述べる。

On the Perception of Tones

KENGO OHGUSHI^{†1}

Perception of tones has been studied and discussed for a long time. As times change, hearing theories change. In this report, changes in the hearing theory are simply introduced. Further, the experimental fact that the pitch of complex tones is slightly but systematically different from the pitch of pure tone corresponding to the fundamental frequency of the complex tone is explained from a new point of view.

1. はじめに

一般にヴァイオリンやクラリネット等の楽音（調波複合音）は、通常その基本周波数に対応する純音と同じピッチ（音の高さ）をもつ単一音として知覚されることが知られている(e.g., Plomp,1967).

複合音のピッチがどのような聴覚メカニズムによって生じているのかを探求する研究は古くから行われている。この動きをかたんに追ってみて現在でも解決されていない問題をどのように考えるべきかを私自身の研究結果を織り交ぜながら述べて行きたい。

2. 聴覚理論の簡単な紹介

2.1 Seebeck と Ohm の論争

Seebeck (1841-44) は、同心円上に等間隔に穴のあいた回転円盤（サイレン）に対して直角方向から管で空気を吹き付けた。空気流は穴と穴の間で周期的に遮断され、明確なピッチをもつ音を発生した。また回転数を速めるとピッチが上昇することを示した。さらに穴の間隔を、 a, b, a, b, \dots と交互に異なった値にすると、ピッチは穴の間隔が等間隔で $(a+b)$ であるような円盤のピッチと等しくなった。これらのことから、Seebeck は音響波形の周期という時間情報がピッチの感覚を惹き起こす（時間説）ののだと考えた。

これに対して、Ohm は複合音のピッチは音響波形に含まれる基本周波数成分の存在によって生じるののだと考えていた。そこで彼は回転円盤のしかし Seebeck は基本周

波数成分の極めて弱い複合音を作り、この音が基本周波数に対応する明確なピッチをもつことを示した。ここで彼らの論争は解決を見ないままに終わった。

2.2 Helmholtz の場所説

約 20 年後に Helmholtz は、中耳が非線形特性をもちこれが基本周波数に対応する差音を発生させ、基本周波数に対応するピッチを惹き起こしているのだと考えた。

また Helmholtz は、基底膜は横に張られた多数の線維から成り、その線維の長さは基底膜の幅が異なるので共鳴する周波数が異なり、異なった周波数の音に対しては異なった線維が共鳴して振動し、これが、異なった聴神経を刺激して異なったピッチを発生させるのであると考え、Ohm の主張を支持した。Helmholtz の理論は場所説 (place theory) と呼ばれ、Seebeck の時間説 (temporal theory) に対抗した。その後数十年の間、場所説が支配的であった。Helmholtz のこの説は共鳴説とも呼ばれている。

2.3 Schouten のレジデュール理論

Schouten は、差音の発生による場所説に対して否定的な新しい実験結果を発表し、レジデュール理論という画期的な理論を発表した (1938,1940) [1][2]。彼は光学装置とサイレンを組み合わせて周期的パルス列音（調波複合音）を発生する装置を作成した。周期が 5 ms のパルス列音を注意を集中して聴くと、4 つの音から成っているように知覚された。すなわち、200 Hz のピッチをもつ鋭い音色の音と、200, 400, 600 Hz の各純音（基音、第 2 倍音、第 3 倍音）を聴き取ることができた。これらの各音の聴こえ方は注意の集中の強さに大きく依存した。Schouten はこれらの鋭い音色

^{†1} 京都市立芸術大学名誉教授
Kyoto City University of Arts, Professor Emeritus

の音のピッチ (= 基音のピッチ) をレジデューピッチと呼んだ。レジデューは、基音と同じ高さであるが、純音とは違って濁った鋭い音色をもっているため、音色によって区別できたと述べている。

また、基本周波数が 200 Hz でその 9, 10, 11 倍音 (1800, 2000, 2200 Hz) から成る調波複合音の各成分周波数を 40 Hz (中心周波数の 2 %) ずつ上昇させると、各成分は 1840, 2040, 2240 Hz となるが、この非調波複合音のレジデューピッチは 204 Hz (もとのレジデューピッチの 2 % 上昇) となった。もしレジデューピッチが差音によって生じるのなら、200 Hz になり、204 Hz にはならないはずである。このことは明確に差音による場所説を否定したものである。

Schouten ら (1962) [3] はレジデューピッチは音響波形の各周期のピーク間の時間間隔という時間情報 (fine structure) によって生じると考えた。

2.4 パターン認識モデル

このモデルは複合音の基本周波数成分の存在(場所情報)や基底膜の振動パターン(時間情報)という情報から直接的にピッチが抽出されるのではない。パターン認識モデル (Wightman, 1973; Goldstein, 1973; Terhardt, 1974) [4][5][6] は、前半部分として周波数分析器の機能を持ち、複合音中の分解できる各成分の周波数またはピッチを推定する。また後半部分は分析器で分解された複数の分解できる周波数(ピッチ)成分から知覚されるべき複合音のピッチを計算する機能をもつ。これらのモデルのパターン認識メカニズムは、複合音の各周波数成分を、およそ 8 次位までの倍音を並べた倍音列によって作られた鋳型 (template) に嵌まればその基本周波数が知覚されるピッチということになる。鋳型と成分音を当てはめてみることを *template matching* という。これらのモデルは基本周波数をずらした非調波複合音についても説明可能であるが、分解できない高い周波数成分のみから成る複合音のピッチについては説明できない。

3. 調波複合音のピッチと基本周波数のピッチ

上述のように、複合音のピッチはその基本周波数に等しい周波数の純音のピッチと同じであると考えられてきた。ところが、Walliser(1969)[7]は、単極性パルス列を 1.4~2.8 kHz の帯域フィルタを通したときの調波複合音のピッチを調べた。5 人の聴取者が、基本周波数が 230~380 Hz の範囲で比較音である純音とのピッチマッチングを行ったところ、概して 2 % 程度以下ではあるが低くマッチングされていることが見出した。この結果は従来の常識を覆すものである。

また Smoorenburg(1970)[8]は、2 周波複合音(1.8 kHz+2.0 kHz; 2.0 kHz+2.2 kHz) と純音それぞれのピッチを聴取者にマッチングさせたところ、複合音のピッチは基本周波数で

ある 200 Hz よりもわずかに低くなるという結果を得た。

次いで Terhardt(1971)[9]は、基本周波数成分を含む広帯域複合音と純音のピッチマッチング実験を行い、基本周波数が約 800 Hz 以下では純音よりも低くなることを示した。また基本周波数成分を除去すると、純音とのピッチの差は広がることを見出した。

大串(1976a)[10]は、キャリア周波数が 2000 Hz で変調周波数がそれぞれ 333 Hz, 400 Hz, 500 Hz であるような振幅変調音のピッチを純音を比較音として求めたところ、基本周波数に対応する純音よりは振幅変調音の方が系統的に低いことを見出した。さらに大串(1976b)[11], Ohgushi(1978)[12]は、基本周波数がほぼ 200 Hz から 1.5 kHz の範囲では、広帯域複合音(パルス列音)のピッチは基本周波数に対応する純音のピッチよりも低いことを見出した。このように、複合音のピッチはその基本周波数成分と同じ周波数の純音のピッチよりもわずかではあるが低くなるという多くの実験データが報告されている。

この現象を説明する理論が Terhardt (1974)[6]により提案されている。彼の理論によれば、ピッチはスペクトルピッチ (spectral pitch) とバーチャルピッチ (virtual pitch) に分けられ、純音のピッチや聴覚系で分解可能な複合音の成分音のピッチをスペクトルピッチと呼び、複合音のピッチはスペクトルピッチによって合成 (synthesize) されるのである。またその成分音のピッチを知覚する場合に、隣接音間の相互マスキング (mutual masking) によってその成分音のピッチは周波数の等しい純音のピッチとはわずかに異なるのである。とくに基音のピッチは第 2 倍音以上の成分からのマスキングにより、常に純音単独のピッチよりは低くなるのである。この理論は、いわゆる場所説に立った神経興奮パターンのピークに対応してピッチが決定するというものである。

4. 関連する知覚現象

4.1 オクターブ伸長現象

二つの純音を継時的に聴き、ちょうど 1 オクターブだけ離れて知覚されるように周波数を調整すると、二音の周波数比は 2 よりもわずか (3 % 以下) ではあるが大きくなる。つまり心理的オクターブは物理的オクターブよりも広くなる。この現象をオクターブ伸長現象 (octave enlargement phenomenon) と呼んでいる。この現象は周波数が高くなるほど顕著になるが、低い方の周波数が 2.7 kHz を越えると聴取者のオクターブマッチングの作業が困難になる (Ward, 1954) [13]。この理由は周波数がほぼ 5 kHz を超えるとトーンクロマが消失するからである。オクターブ伸長現象の説明は聴神経の純音刺激に対する 1 周期に対応する発火時間間隔が、周波数が高くなるにしたがって神経細胞の不応期のために系統的に周期よりも長くなっていくという生理実験データ (Rose ら, 1967) [14]を用いて行われてい

る(Ohgushi, 1983). また Rose らの生理実験データと同じ傾向の生理データが後に McKinney ら(1999) [15]によって報告されている. また2オクターブおよびもっと広いオクターブ間の伸長現象は個々のオクターブ伸長の和になる(Terhardt,1983) [16].

4.2 複合音の多重ピッチ

Davis, Silverman & McAuliffe (1951) [17]は, 繰り返し周波数(基本周波数)が90~150 Hzの周期パルス列を中心周波数が2 kHzの帯域フィルタに通した調波複合音を作成し, 聴取者にこの音を純音とのピッチマッチングを行うように求めた. 実験結果によれば, 聴取者は基本周波数あるいは2000 Hzに合わせたが, しばしば1オクターブの食い違いがあり, 場合によっては2オクターブの違いもあった.

また大串(1976a) [10]は, 振幅変調音(3成分)と純音のピッチマッチング実験において, 振幅変調音が調波複合音の場合, 基本周波数がほぼ1オクターブあるいは2オクターブ異なる純音ともしばしばマッチングされた. さらに5度だけ離れた純音とマッチングされることもあった. これらの結果は, 複合音のピッチの判断は場合によってはかなりあいまいであり, とくに tone chroma が等しい場合にはしばしば異なる複数のピッチとして判断されることを示している. さらに Terhardt(1987)[19]も複合音のピッチに純音のピッチをマッチングさせたとき, 複数の周波数にマッチングされることを示している.

5. 2成分複合音の実験結果と新しい理論

5.1 ピッチ実験結果の例 2成分音のピッチ

テスト音を2成分複合音とし, 純音を比較音としてピッチマッチングを行った4人の聴取者の実験結果(大串, 1976b) [11]をまとめて図1に示す.

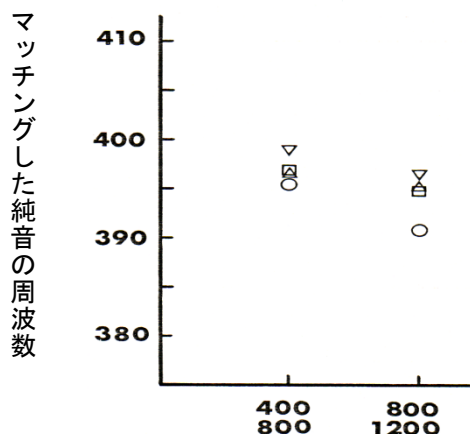


図1 2周波成分複合音のピッチ

図1の結果を要約すると, (1)400 Hzと800 Hzの2周波成分音のピッチは400 Hzの純音のピッチよりも低くなる;

(2)基本周波数成分を除去し, 第3倍音を加えた800 Hzと1200 Hzの2周波成分音のピッチはさらに低くなる. これらの結果は, 周波数成分がたった2つだけであっても複合音のピッチの低下を示している.

5.2 ピッチ知覚のメカニズムに関する理論

上述の結果を説明するために, 次のような理論を提案する. すなわち, 「純音のピッチを調波複合音のピッチにマッチングする場合, 聴取者は複合音の基音成分とのマッチングをするだけでなく, 分解できる(resolved)低次倍音とのマッチングを行い, それらの値の間の何らかの加重平均値をとる. つまり, 第2倍音成分とのオクターブマッチングも行われ, また第3および第4倍音あるいはそれ以上の倍音などとのマッチングも行われている可能が考えられる. 聴取者による実験結果の違いは, おそらく倍音の分解能力などの個人差によるものが大きいと思われる.

5.3 実験的事実の理論による説明

この理論に従えば, 図1の実験的事実は容易に説明できる:

- (1)2成分複合音(400 Hz+800 Hz)のピッチは400 Hzの純音よりわずかに低くなるが, その理由としては, 400 Hz付近の純音は, 基音の400 Hzとマッチングされるだけでなく, 800 Hzの第2倍音とのオクターブマッチングもなされるからである. 純音のオクターブマッチングにおいては, 心理的オクターブの伸長現象が生じるので, 800 Hz部分音の1オクターブ下に感じられる純音の周波数は400 Hzよりも低くなる. したがって, この複合音のピッチは基音とのピッチマッチングおよび第2倍音とのオクターブマッチングとの中間的な値となり, 400 Hzよりも低くなるのである.
- (2)(800 Hz+1200 Hz)の複合音はさらにピッチが低下する. その理由は, 基音成分とのピッチマッチングがなくなり(結合音が生じればピッチマッチングは行われる)オクターブマッチングおよび第3倍音とのマッチングになるからである.

6. 討論

Terhardtは場所説により複合音のピッチシフト現象を説明したが, さらに純音のオクターブ伸長現象も同じ原理で説明している. しかし継時的に提示される純音のオクターブ伸長現象はマスキング現象では説明できない.

Petersら(1983) [19]は, 基音そのもののピッチはマスキングの影響によってわずかに低下するという Terhardtの理論をチェックするために, 部分音(基音を含む)の分析的聴取(analytic listening)の実験を行った. 調波複合音の基音成分に注意を集中すれば(完全に集中できれば)部分音としての基音のピッチは基音の周波数に等しい純音のピッチと等しくなるはずである.

テスト音を、基本周波数を 200 Hz 一定、部分音は 7 次倍音以下の 7 成分音を用いて、聴取者に各部分音のピッチを純音のピッチとマッチングさせた。実験法の一つとしては、テスト用の部分音は最初の 200 ms の間は除去されており、その後の 500 ms の間は加えられた。この途中からの提示により聴取者の注意を引きつけた。実験結果によれば、実験法や提示条件によって本質的には変化はなく、各部分音のピッチは、各成分を単独で提示したときのピッチに非常に近いことが示された。

この結果は Terhardt の理論に対する反論となっており、時間説の方が妥当だと考えられる。

参考文献

- 1) Schouten, J.F. : The perception of subjective tones. Proceedings of Koninklijke Nederlandsche Akademie van Wetenschappen, Vol.41, No.10, pp.1086-1093 (1938).
- 2) Schouten, J.F. : The Perception of pitch. Philips Technical Review, Vol.5, No.10, pp.286-294 (1940).
- 3) Schouten, J.F., Ritsma, R.J. and Cardozo, B.L. : Pitch of the residue. J. Acoust. Soc. Am., Vol.34, No.8,(2), pp. 1418-1424 (1962).
- 4) Wightman, F.L. : The pattern-transformation model of pitch. J. Acoust. Soc. Am., Vol.54, No.2, pp.407-416 (1973).
- 5) Goldstein, J.L. : An optimum processor theory for the central formation of the pitch of complex tones. J. Acoust. Soc. Am., Vol.54, No.6, pp.1496-1516 (1973).
- 6) Terhardt, E. : Pitch, consonance, and harmony. J. Acoust. Soc. Am., Vol.55, No.5, pp.1061-1069 (1974).
- 7) Walliser, K. : Zusammenhänge zwischen dem Schallreiz und der Priodenhöhe. Acustica, Vol.21, pp.319-329 (1969).
- 8) Smoorenburg, G.F. : Pitch perception of two-frequency stimuli, J. Acoust. Soc. Am., Vol.48, No.4, pp. 924-942 (1970).
- 9) Terhardt, E. : Die Tonhöhe Harmonischer Klänge und das Oktavintervall. Acustica, Vol.24, pp. 126-136 (1971).
- 10) 大串健吾 : 複合音の高さの知覚形成のメカニズム. 音響学会誌, Vol.32, No.5, pp. 300-309 (1976a).
- 11) 大串健吾 : 複合音の高さの知覚における時間情報の役割. 音響学会誌, Vol.32, No.11, pp.710-719 (1976b).
- 12) Ohgushi, K. : The origin of tonality and a possible explanation of octave enlargement phenomenon. J. Acoust. Soc. Am., Vol.73, No.5, pp.1694-1700 (1983).
- 13) Ward, W. : Subjective musical pitch. J. Acoust. Soc. Am., Vol.26, No.3, pp. 369-380 (1954).
- 14) Rose, J.E., Brugge, J.F., Anderson, D.J. and Hind, J.E. : Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. J. Neurophysiol., Vol.30, pp.769-793 (1967).
- 15) McKinney, M.F. and Delgutte, B. A possible neurophysiological basis of the octave enlargement effect. J. Acoust. Soc. Am., Vol.106, No.5, pp.2679-2692 (1999).
- 16) Terhardt, E. : Auditory universals in music perception. Musicological conference of world music weeks, pp.1-11 (1983).
- 17) Davis, H., Silverman, S.R. and McAuliffe, D.R. : Some observations on pitch and frequency. J. Acoust. Soc. Am., Vol.23, No.1, pp.40-42 (1951).
- 18) Terhardt, E. : Gestalt principles and music perception. In W.A. Yost & C. S. W. Watson (Eds.), Auditory processing of Complex sounds, (pp.157-166). New Jersey: Lawrence Erlbaum Associates, Publishers (1987).
- 19) Peters, R.W., Moore, B.C.J., and Glasberg, B.R.J. : Pitch of components of complex tones. J. Acoust. Soc. Am., Vol.73, No.3, pp.924-929 (1983).