ものは突発性難聴であり，発症早期（遅くても２週以内）に副腎皮質ステロイドを中心とした治療を行う必要がある。また，強大音暴露による音響外傷，アミノ配糖体抗菌薬やシスプラチンなどの抗がん剤による薬剤性難聴，細菌感染（髄膜炎，内耳炎）なども急性難聴をきたす。聴力の再発や変動をきたす代表的なものはメニエール病である。

　一方，慢性の感音難聴は聴力が固定しているものと進行性に分けられる。徐々に進行する感音難聴の代表は老人性難聴であり，ほかに遺伝性難聴，騒音性難聴，特発性難聴などがある。遺伝性難聴の多くは先天性であるが，後天性に進行性の難聴をきたすこともある。日本人の報告をみると，常染色体優性では COCH, KCNQ4, WFS1, TECTA など，常染色体劣性では SLC26A4, CDH23など，ミトコンドリア遺伝子異常では3243位点変異，1555位点変異などがあり，このうち高音域から徐々に悪化するもの（TECTA 遺伝子異常，CDH23遺伝子異常，ミトコンドリア遺伝子1555位変異など）では早期発症の老人性難聴と鑑別が必要となる。

　騒音性難聴は，慢性的に長時間にわたって騒音に曝露されて発症する。日々の騒音下に生じた一過性聴力閾値上昇が蓄積され，完全に回復しないうちに次の騒音が加わることが原因とされ，職業と密接な関係があるため職業性難聴とも呼ばれる。初期の聴覚機能の変化は４kHz を中心に現れることが多く，さらに進行すると広い周波数帯域が侵され，高音漸傾型などに移行

する。騒音下に作業した場合，比較的早期の就業５～15年間での難聴の進行が著明であり，10数年以降には進行が治まる傾向にある。年単位の比較的短い期間では目にみえて難聴をきたさないレベルの騒音であっても，長期的には加齢に伴う感音難聴の進行を悪化させると考えられており，静かな環境に住んでいた老人の難聴では水平型になるが，騒音の影響を受けていた場合には高音域がより障害されると考えられている[1]。

## 老人性難聴

　老人性難聴は，加齢に伴い進行する両側性の感音難聴の総称であり，聴力閾値上昇，（特に騒音下での）語音聴取能の低下，聴覚情報の中枢処理遅延，音源認知の障害などが知られる。この結果，日常会話，音楽聴取，社会生活活動などにおいて困難を感じることになる。加齢に伴う聴覚障害には，大きく分けて３つの要素，すなわち①末梢聴覚の機能低下，②中枢聴覚の機能低下，③認知機能全般の低下が関与している。高齢者が聞き取りの障害を訴える場合には，これら末梢・中枢・認知の３機能が複合的に障害されていると考える必要がある[1]。日常診療でよく耳にする「雑音の中でうまく会話が聞き取れない」という訴えは，若年の感音難聴症例でも同様に雑音下聴取が障害されることから，末梢聴覚の機能低下が主に影響していると考えられる。一方，空間聴や音源分離の機能が脳幹

にあることから，中枢聴覚の障害も（雑音下）聴取を低下させる原因となる。また，認知機能の低下も音源への注意に影響し，雑音下聴取がより困難になると推測されている。

　図４に，30歳代から５歳ごとの年齢別聴力（平均）を示す[2]。この図からわかる特徴は２つあり，１つは，聴力が高周波数域から障害され，難聴が進行するにつれ低音から中音域まで障害されることである。２つ目は，聴力の進行が年をとるほど加速する傾向にあることであり，その傾向は高周波数ほど明らかである。また，この他に３つ目の特徴として，難聴に個人差が大きいこともあげられる。さらに，男女差が存在することもよく知られているが，最近では以前の報告より男女差は縮まりつつあるとされ，男女の生活様式の近似化などの影響も示唆される。

### 1．疫学およびリスク因子

　Yueh ら[3]は，65歳以上の25～40％，75歳以上の40～66％，85歳以上の80％以上において，加齢に伴う難聴があると推定している。韓国の2000年の報告[4]では，500, 1,000, 2,000, 4,000Hz の６分法平均気導聴力レベルが27dBHL 以上の割合は65歳以上で37.8％，41dBHL 以上は8.3％とされている。一般に，40dBHL 以上の中等度難聴になると日常生活上の不自由を強く感じ，補聴器装用を希望することが多い。韓国からの結果は，難聴ありと診断された高齢者の５人に１人が生活上の不自由を感じていることを示唆している。内田ら[5]は「国立長寿医療
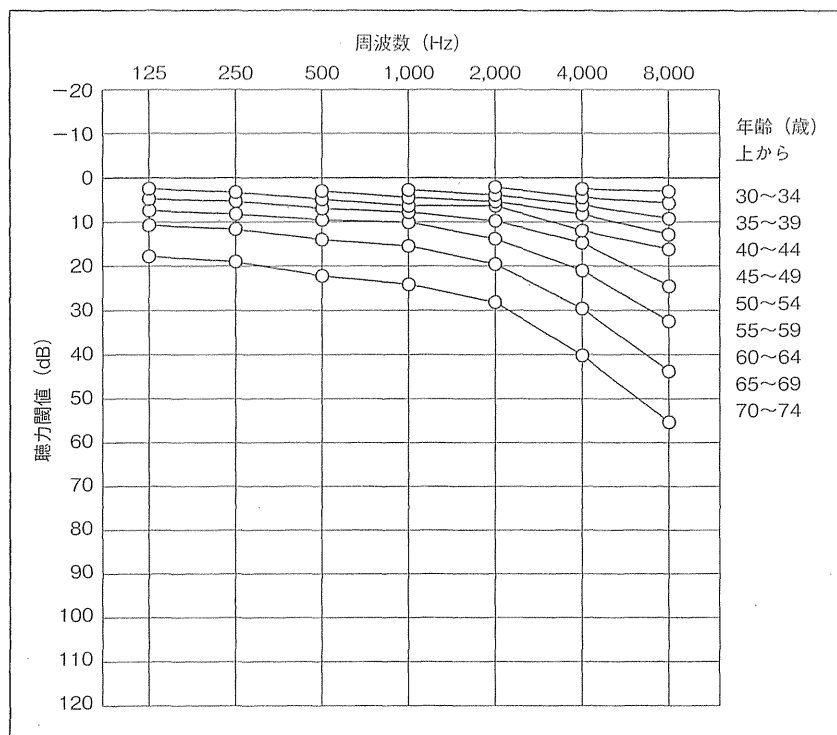
難聴の基礎と臨床



図4．年齢別平均聴力

（文献2より引用改変）

進国における老人性難聴は，純粋な老化に加え，環境要因が大きく影響していることが示唆される。

一般に，老人性難聴の発症や程度に影響する因子としては，遺伝要因のほか，人種差，騒音曝露歴，喫煙，飲酒，糖尿病・循環器疾患などの合併，性ホルモンなどがあげられる[7)8)]。遺伝的要因の関与は0.35～0.55と推定されており，また黒人のほうが白人より難聴の程度が軽いことも知られている[8)]。NILS-LSA においては，動脈硬化や肥満に関与する遺伝子多型の関与が示唆されている[9)]。遺伝子多型ではエンドセリン1，脂肪酸結合蛋白2，ミトコンドリア脱共役蛋白2などが報告されている[10)]が，その多くが動脈硬化や酸化ストレスと関与するものであることは興味深い。遺伝外要因については，NILS-LSA では騒音暴露歴，内頚・網膜動脈の動脈硬化，糖尿病などの関与が報告されている[9)]。文献的には，高血圧，心血管疾患，脳血管疾患，喫煙，糖尿病，騒音曝露などとの相関が報告されている[7)8)]。

**2．老人性難聴の発症機序**

加齢に伴う老化現象の一般的な発症機序については諸説があるが，現在では酸化ストレスとこれに伴うミトコンドリア DNA の損傷集積が主に考えられている。老人性難聴の発症機序についても，基礎研究のデータが十分ではないものの，この説を支持する所見が得られてきている[8)]。

たとえば，CBA マウスの蝸牛では，酸化ストレスの痕跡である hydroxy-

センター　老化に関する長期縦断疫学研究（NILS-LSA）」データを検討し，WHO の聴力障害基準に従い，500，1,000，2,000，4,000Hz の会話音域4周波数平均気導聴力レベルを基準とした良聴耳聴力レベルが25dBHL を超えた場合を「難聴あり」として，第6次調査（2008～2010年）参加者の難聴有病率を集計した。その結果を図5[5)]に示すが，難聴有病率は60～64歳までは徐々に増加し，65歳以上で急速に増加する傾向が見て取れる。なお，いずれの年代においても男性の有病率が女性より高いが，有意差は65～69歳でのみ見られている。この結果をもとに65歳以上の全国難聴有病者数を推計する

と1,655万程度であり，また「耳疾患の既往なし」「職場騒音の就労歴なし」と答えた者のみの集計結果から計算すると1,569万人超と推計されるという。上記の韓国の結果を考えると，300万人以上が生活に不自由を感じる中等度以上の難聴に罹患していると推測される。

一方，この高い難聴の頻度は先進国にのみ見られる可能性も高い。Rosenら[6)]は，スーダンに住む Mabaan 族の聴力検査を行い，高齢に至るまで聴力が明らかによく保たれていることを報告した。この地域は特に静かな場所であり，人種の差（肌色の差）や遺伝的素因の影響もあると思われるが，先
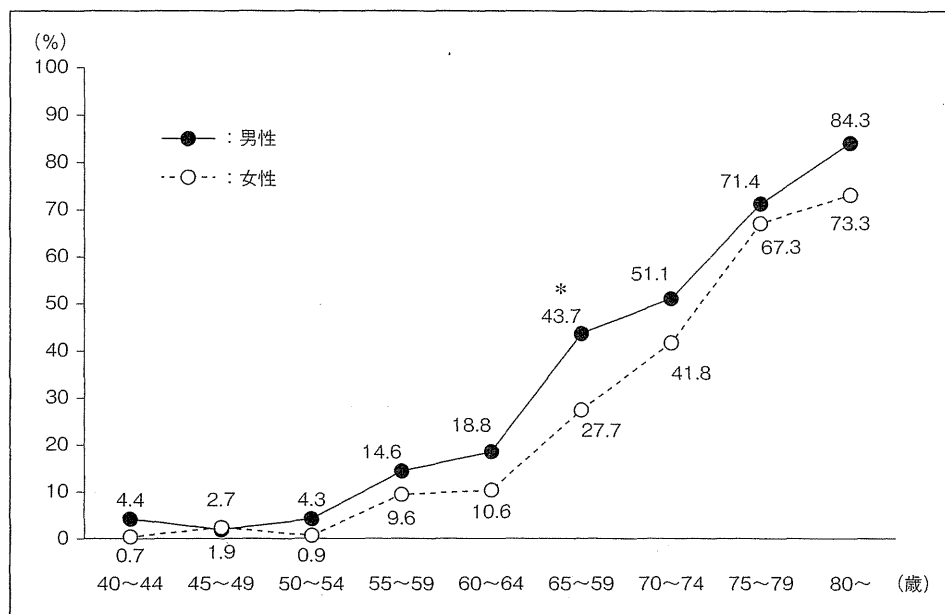
図5．NILS-LSA における難聴の有病率
* ：p＜0.05

（文献5より引用改変）

nonenal や nitrotyrosine が加齢に伴って増加し，内在性アンチオキシダントである superoxide dismutase（SOD）2が低下する[11]。SOD1をノックアウトしたマウスでは野生型より早期に老人性難聴が出現する[12]。我々は，カタラーゼをミトコンドリア内に過剰発現させた遺伝子改変マウスでは，野生型に比べて老人性難聴の発症が遅れることを見出した[13]。これらの結果は，内在性のアンチオキシダントが老人性難聴の進行抑制に関与していることを示唆している。また，17種類のフリーラジカルスカベンジャーを C57BL/6マウスに投与して老人性難聴の発症を抑制できるか調べたところ，コエンザイム Q10，α リポ酸，ＮアセチルＬシステインが老人性難聴発症を抑制して

いた[13]。

一方，ミトコンドリア機能とミトコンドリア DNA 障害の影響についてみると，高齢者の側頭骨病理標本において，難聴のあるヒトの標本では正常聴力者に比べてミトコンドリア遺伝子4,977塩基対の欠失が多くみられることが示されている[14]。また，DBA/2J マウスでも老人性難聴をきたした場合には，蝸牛でのミトコンドリア機能が著明に低下していた[15]。ミトコンドリア DNA の変異が増加する遺伝子改変マウス（POLG マウス）を作製し，その聴力を調べたところ，野生型では難聴の生じない数ヵ月頃から難聴が生じた[16][17]。これらの結果から，ミトコンドリア遺伝子の障害蓄積が老人性難聴の進行に関与していると考えられる。

なお，この一連の研究において，老人性難聴の発症には蝸牛での Bak 遺伝子の発現亢進が関与することが示唆された。老人性難聴をきたした蝸牛に比べ，遺伝子改変やアンチオキシダント投与で難聴の発現が抑えられた蝸牛では，Bax 遺伝子の発現に差はないが，Bak 遺伝子の発現は抑制されていた。そこで Bax と Bak をノックアウトしたマウスの聴力経過を調べると，野生型と Bax ノックアウトマウスには差がなかったが，Bak ノックアウトマウスでは加齢による難聴の発現が有意に抑えられていた。また，蝸牛の培養細胞にパラコートを加えた場合，Bak の発現が上昇して細胞死が起こるが，Bak ノックアウトマウスの蝸牛培養細胞では細胞死が有意に抑えられて

いた[13]。この結果は，酸化ストレスが Bak の発現を介して蝸牛の細胞死を誘導していることを示唆している。

カロリー摂取制限は線虫から霊長類まで老化の出現を抑制することが知られ，ラットやいくつかの strain のマウスではカロリー制限による老人性抑制効果が示されている。我々がC57BL/6マウスに通常量の食餌を与えたところ，15ヵ月齢までに中等度の難聴が出現したが，26％カロリー摂取を制限した場合は聴力が正常に維持され，組織変性はほとんど生じなかった[18]。ミトコンドリアに発現する Sirt3 のノックアウトマウスでは，このカロリー摂取制限による老人性難聴抑制効果がみられないことから，

Sirt3 がカロリー制限の老人性難聴予防効果に重要なこと，酸化ストレスの抑制が鍵であることが報告されている[19]。

ヒトの老人性難聴の発症機序は動物とは必ずしも同じでないと考えられるが，以上のような動物実験の結果およびヒトの疫学データから，加齢に伴う蝸牛内の酸化ストレスによりミトコンドリア DNA 変異が蓄積，ミトコンドリア機能が悪化し，その結果，有毛細胞，ラセン神経節細胞，血管条など聴覚機能に重要な細胞が障害を受けて脱落し，難聴が進行性に生じると考えられる（図6）[8]。

## 3．老人性難聴の予防と治療

上記の仮説に従うと，蝸牛内における過剰なフリーラジカル産生の予防が重要ということになり，不必要な強大音への曝露を若いうちから極力避けることが勧められる。また，騒音環境では蝸牛に必要な血流が増加するが，動脈硬化があると十分な血流の供給が障害され，相対的虚血，さらにその後の再灌流によるフリーラジカルの過剰産生を引き起こすと考えられる。この意味からも動脈硬化を予防することは重要であり，高脂血症，糖尿病，高血圧症などを持病にもつ患者または予備軍では内科的治療による介入も勧められる。

フリーラジカルを除去する物質やそのサプリメント，特にミトコンドリア内で作用するものの摂取も予防効果が期待されるが，ヒトにおける老人性難聴予防効果のエビデンスは十分ではない。アンケートを用いた食事内容調査から検討した海外の報告では，たとえば，カロテン，ルテイン，リコペン，ビタミン A，C，E などいくつかのフリーラジカルスカベンジャーの摂取量と老人性難聴の程度との相関が報告されているが，その信憑性は十分ではない。このような個々の物質を調べるのではなく，食事の健全性に着目したほうがよいという考えもある。Healthy Eating Index は 5 つの食材群（肉，乳製品，果実，野菜，穀物），4 つの栄養素（脂肪摂取割合，飽和脂肪摂取割合，コレステロール，塩分），食事の variety をそれぞれ10点満点として評価するものであるが，この評価が高
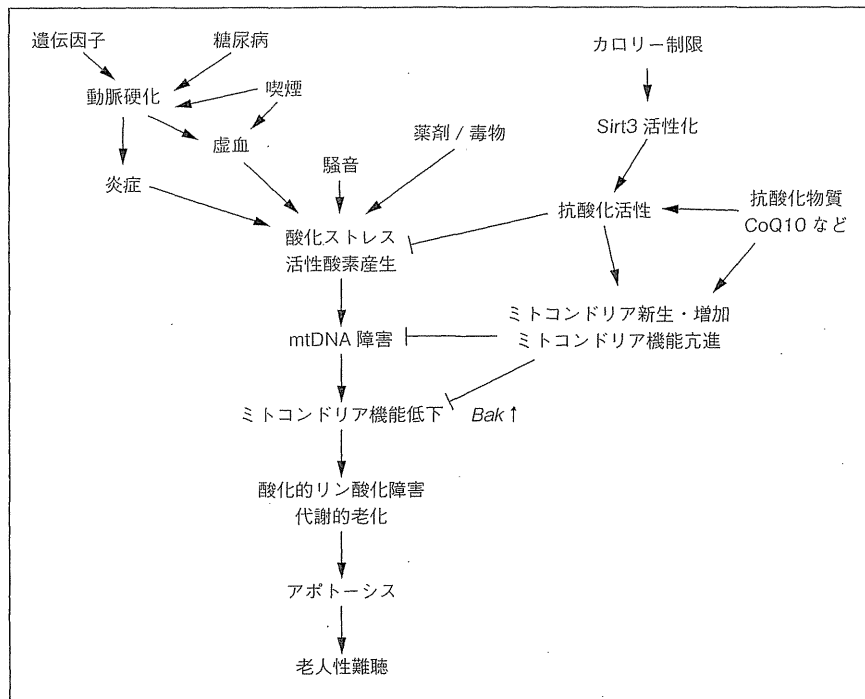


図6．老人性難聴の発症機序（仮説）

（文献8より引用改変）

い群ほど老人性難聴が軽かったという報告がされている[20]。

老人性難聴がすでに生じており，生活の不具合を感じる場合，すなわちコミュニケーション障害がある場合，まずは補聴器の活用が勧められる。米国の退役軍人を対象として補聴器装用後1年間の経過をみた検討では，補聴器装用は認知機能，社会性，感情，うつ傾向，コミュニケーションに明らかに有益であったとされている[21]。横断研究である Blue Mountains Eye Study に参加した60歳以上1,328名の検討でも，両側軽度難聴（500〜4,000Hz の平均が25〜40dBHL）者でうつの傾向が強く，1日1時間以上補聴器を装用する者では有意にうつ傾向は抑制されたと報告されている[22]。このような補聴器に付帯する有用性についてはいうまでもないが，特に「騒音下での聞き取りが悪い」ということに配慮し，ノイズリダクション機能を活用することが重要である。また，会話においては，時間分解能が障害されていることを念頭に置き，「ゆっくりと話してもらう」ように指導する。また，顔がよく見える位置で会話すると視覚情報も活用できる。

時間分解能の機能低下は補聴器では補えないが，聴覚訓練で改善する可能性がある。Anderson ら[23] は，55〜70歳の67例を対象に，トレーニング群では家庭で8週間の Brain Fittness cognitive training 聴覚訓練を行い，対照群は科学や歴史などの教育 DVD を見て多項選択式問題（multiple choice question：MCQ）に答える（注

意して DVD を見るため）こととし，評価は6つの formant をもつ170ms の ［da］を刺激音とした脳幹反応（静寂下と騒音下（S/N 比＋10dB）で記録），騒音下の単語聴取（S/N 比0〜25dB），短期記憶などで行った 。その結果，formant transition に反応する時間が騒音下で早くなり，短期記憶が向上し，騒音下聴取成績が改善し，脳幹反応のピークの騒音負荷によるずれも聴覚訓練群で減少したと報告している。この結果は，聴覚に基づく認知訓練が，加齢に伴う時間分解能低下をある程度改善させうる可能性を示唆している。

●文　献
1 ) Gates GA, Mills JH : Presbycusis. Lancet 366 : 1111-1120, 2005
2 ) 立木　孝，一戸孝七：加齢による聴力悪化の計算式. Audiology Japan 46 : 235-240, 2003
3 ) Yueh B, Shapiro N, MacLean CH, et al : Screening and management of adult hearing loss in primary care: scientific review. JAMA 289 : 1976-1985, 2003
4 ) Kim HN, Kim SG, Lee HK, et al : Incidence of presbycusis of Korean populationin Seoul, Kyunggi and Kangwon provinces. J Korean Med Sci 15 : 580-584, 2000
5 ) 内田育恵，杉浦彩子，中島　務，他：全国高齢難聴者数推計と10年後の年齢別難聴発症率. 老化に関する長期縦断疫学研究（NILS-LSA より）. 日本老年医学会雑誌 49：222-227，2012
6 ) Rosen S, Beragman M, Plester D, et al : Presbycusis study of a relatively noise-free population in the Sudan. Ann Otol Rhinol Laryngol 71 : 727-

743, 1962
7 ) Van Eyken E, Van Camp G, Van Laer L : The complexity of age-related hearing impairment: contributing environmental and genetic factors. Audiol Neurootol 12 : 345-358, 2007
8 ) Yamasoba T, Lin FR, Someya S, et al : Current concepts in age-related hearing loss: epidemiology and mechanistic pathways. Hear Res 303 : 30-38, 2013
9 ) 下方浩史：高齢者の聴力に個人差が大きいのは何故か. 全身の老化との関係において. Audiology Japan 51 : 177-184, 2008
10) Uchida Y, Sugiura S, Ando F, et al : Molecular genetic epidemiology of age-related hearing impairment. Auris Nasus Larynx 38 : 657-665, 2011
11) Jiang H, Talaska AE, Schacht J, et al : Oxidative imbalance in the aging inner ear. Neurobiol. Aging 28 : 1605-1612, 2007
12) McFadden SL, Ding D, Burkard RF, et al : Cu/Zn SOD deficiency potentiates hearing loss and cochlear pathology in aged 129, CD-1 mice. J Comp Neurol 413 : 101-112, 1999
13) Someya S, Xu J, Kondo K, et al : Age-related hearing loss in C57BL/6J mice is mediated by Bak-dependent mitochondrial apoptosis. Proc Natl Acad Sci U S A 106 : 19432-19437, 2009
14) Bai U, Seidman MD, Hinojosa R, et al : Mitochondrial DNA deletions associated with aging and possibly presbycusis: a human archival temporal bone study. Am J Otol 18 : 449-453, 1997
15) Someya S, Yamasoba T, Prolla TA, et al : Genes encoding mitochondrial respiratory chain components are profoundly down-regulated with

aging in the cochlea of DBA/2J mice. Brain Res 1182 : 26-33, 2007

16) Someya S, Yamasoba T, Kujoth GC, et al : The role of mtDNA mutations in the pathogenesis of age-related hearing loss in mice carrying a mutator DNA polymerase gamma. Neurobiol. Aging 29 : 1080-1092, 2008

17) Kujoth GC, Hiona A, Pugh TD, et al : Mitochondrial DNA mutations, oxidative stress, and apoptosis in mammalian aging. Science 309 : 481-484, 2005

18) Someya S, Yamasoba T, Weindruch R, et al : Caloric restriction suppresses apoptotic cell death in the mammalian cochlea and leads to prevention of presbycusis. Neurobiol Aging 28 : 1613-1622, 2007

19) Someya S, Yu W, Hallows WC, et al : Sirt3 mediates reduction of oxidative damage and prevention of age-related hearing loss under caloric restriction. Cell 143 : 802-812, 2010

20) Spankovich C, Le Prell CG : Healthy diets, healthy hearing: National Health and Nutrition Examination Survey, 1999-2002. Int J Audiol 52 : 369-376, 2013

21) Mulrow CD, Tuley MR, Aguilar C : Sustained benefits of hearing aids. J Speech Hear Res 35 : 1402-1405, 1992

22) Gopinath B, Wang JJ, Schneider J, et al : Depressive symptoms in older adults with hearing impairments: the Blue Mountains Study. J Am Geriatr Soc 57 : 1306-1308, 2009

23) Anderson S, White-Schwoch T, Parbery-Clark A, et al : Reversal of age-related neural timing delays with training. Proc Natl Acad Sci USA 110 : 4357-4362, 2013

**著者プロフィール**

山岨　達也

東京大学医学系研究科外科学専攻耳鼻咽喉科学分野教授
　1983年 東京大学医学部医学科卒業。竹田綜合病院，日立総合病院にて研修。亀田総合病院部長を経て，1993年に東京大学耳鼻咽喉科講師，1999年に助教授，2007年より教授。1996年から1998年までミシガン大学クレスギ聴覚研究所に留学。
専門：耳科学，聴覚医学，人工内耳

# Factors that account for inter-individual variability of lateralization performance revealed by correlations of performance among multiple psychoacoustical tasks

## Atsushi Ochi[1,2], Tatsuya Yamasoba[2] and Shigeto Furukawa[1]*

[1] Human Information Science Laboratory, NTT Communication Science Laboratories, NTT Corporation, Atsugi, Japan
[2] Department of Otolaryngology, Faculty of Medicine, University of Tokyo, Tokyo, Japan

This study explored the source of inter-listener variability in the performance of lateralization tasks based on interaural time or level differences (ITDs or ILDs) by examining correlation of performance between pairs of multiple psychoacoustical tasks. The *ITD*, *ILD*, *Time*, and *Level* tasks were intended to measure sensitivities to ITD; ILD; temporal fine structure or envelope of the stimulus encoded by the neural phase locking; and stimulus level, respectively. Stimuli in low- and high-frequency regions were tested. The low-frequency stimulus was a harmonic complex ($F_0 = 100$ Hz) that was spectrally shaped for the frequency region around the 11th harmonic. The high frequency stimulus was a "transposed stimulus," which was a 4-kHz tone amplitude-modulated with a half-wave rectified 125-Hz sinusoid. The task procedures were essentially the same between the low- and high-frequency stimuli. Generally, the thresholds for pairs of ITD and ILD tasks, across cues or frequencies, exhibited significant positive correlations, suggesting a common mechanism across cues and frequencies underlying the lateralization tasks. For the high frequency stimulus, there was a significant positive correlation of performance between the ITD and Time tasks. A significant positive correlation was found also in the pair of ILD and Level tasks for the low- frequency stimulus. These results indicate that the inter-listener variability of ITD and ILD sensitivities could be accounted for partially by the variability of monaural efficiency of neural phase locking and intensity coding, respectively, depending of frequency.

**Keywords: interaural time difference, interaural level difference, level discrimination, correlation, temporal fine structure, phase locking**

## INTRODUCTION

Performance in lateralization tasks based on interaural time and level differences (ITDs or ILDs), the major cues for horizontal sound localization, often varies markedly among listeners. Lateralization behavior is a product of multiple stages of auditory processing, and thus the listener's performance should reflect the efficiencies of the individual processes by varying degrees. We consider that the processing of the ITD or the ILD in the auditory system consist of two or more stages. The earliest is the peripheral stage, in which the auditory information is processed individually for different ears. At this stage, the temporal structure and intensity of sounds at each ear are encoded to neural signals in the form of the timing and number of auditory nerve firings. The outputs of this stage of processing are fed to processes at the binaural interaction stage, where the relative timing and number of neural firings for the two ears are compared. This binaural interaction stage is followed by the subsequent higher-order processes.

The present study aimed to evaluate the relative contributions of these processing stages to the inter-listener variabilities in lateralization performance. We measured listeners' monaural sensitivities to the temporal structure and intensity of a sound stimulus, as well as their ITD and ILD sensitivities. The hypothesis was that the lateralization performance based on ITD is predominantly determined by the efficiency of temporal structure coding by neural phase-locking at the peripheral processing stage. If this is true, we would expect that the ITD-based lateralization performance correlates with the performance of a non-lateralization task, which reflects sensitivity to the temporal structure of the stimulus that is presumed to be represented by phase locking. A similar hypothesis and prediction are possible in terms of the relationship between ILD-based lateralization and peripheral intensity coding.

The authors are not aware of a study examining the extent to which monaural intensity (or level) encoding efficiency could account for individual differences in ILD sensitivity. On the other hand, the above hypothesis on the relationship between temporal structure coding and ITD sensitivity is supported by studies on the effects of aging and/or hearing-impairment. Groups of aged listeners (Strouse et al., 1998; Hopkins and Moore, 2011) with sensorineural hearing impairment (Strelcyk and Dau, 2009; Hopkins and Moore, 2011) and those with auditory neuropathy (Zeng et al., 2005) exhibited degraded performance more or less specific to the ITD-based lateralization task and to tasks that measure monaural sensitivity to temporal structure, in comparison to control groups. Within-listener correlation between the two types of tasks has also been reported. Strelcyk and Dau

(2009) found a positive correlation between the FM detection threshold (considered to be indicative of sensitivity to the temporal fine structure, TFS) and ITD-based lateralization threshold for hearing-impaired listeners (there was no report for normal-hearing listeners). A similar relationship between the monaural sensitivity to the TFS and the binaural sensitivity to interaural phase differences was also reported for a pooled population of young and aged listeners with and without hearing impairment (Hopkins and Moore, 2011). Nevertheless, it is uncertain whether the positive correlation could be applicable also to the population of normal-hearing listeners. A possibility is that a long-term impairment of a single mechanism (i.e., peripheral TFS coding) affects the efficiency of another independent mechanism (i.e., central binaural processing), leading to an apparent correlation of performance. Strouse et al. (1998) found a strong positive correlation between the monaural temporal-gap detection threshold and ITD discrimination threshold for a group of normal-hearing young listeners, although such a positive correlation was not found for aged listeners. It should be noted, however, that the gap detection task is considered to focus on the sensitivity to the temporal envelope, rather than on that to the cycle-by-cycle TFS of the stimulus.

A secondary aim of the present study was to examine whether mechanisms for processing the ITD (and ILD) are essentially the same across operating frequency regions. It has been argued that essentially the same binaural mechanism is involved in processing ITDs at low and high frequencies, and apparent differences in ITD sensitivities between the frequency regions reflect differences in input to the system (Van De Par and Kohlrausch, 1997; Bernstein, 2001): When high-frequency "transposed stimulus" (see Material and Methods) is used so that the pattern of neural phase locking to the envelope of the stimulus resembles that to TFS of a low-frequency stimulus, listeners' performance for ITD-related tasks should be comparable. Furukawa (2008), however, found that the degree of ITD and ILD cue interaction in lateralization tasks was smaller for low- than for high-frequency regions, even when the inputs to the binaural system were made comparable by using low-frequency tones and high-frequency transposed stimuli. This implies that a more-or-less independent ITD processor exists in the low frequency region, whereas in the high-frequency region, ITD is processed by a mechanism that is common for ILD processing. In this study, we used low- and high-frequency stimuli and examined the relationship between the lateralization tasks and the monaural temporal/intensity-related tasks for each type of the stimuli. Qualitatively different results between the stimulus types would imply the involvement of separate binaural mechanisms in lateralization depending on stimulus frequency.

## MATERIALS AND METHODS
### LISTENERS AND APPARATUS

Twenty-two adults (10 males and 12 females; 19–43 years old, mean 32.0) participated in the experiment as listeners. All gave written informed consent, which was approved by the Ethics Committee of NTT Communication Science Laboratories. The listeners showed normal audiometric thresholds (<25 dB HL) at frequencies of 250, 500, 1000, 2000, 4000, and 8000 Hz. They had no symptoms of hearing loss and had never been diagnosed as having hearing loss by medical examination. All testing took place in a double-walled sound booth. The listener was seated in front of a computer monitor, which displayed indicators for observation intervals of the forced-choice task and buttons for responses (described later).

Stimuli were digitally synthesized by a personal computer (sampling frequency: 44.1 kHz) and generated by using a digital-to-analog converter with a resolution of 24 bits (M-AUDIO, Transit USB). The signals were amplified and presented to the listener through Sennheiser HDA200 headphones.

MATLAB (Mathworks, Inc.) software was used for stimulus synthesis, experimental control, and data analyses.

### STIMULI

The low- and high-frequency stimuli were designed to assess the listener's ability to use information based on neural phase-locking to the stimulus TFS and envelope, respectively, in the ITD and Time tasks. Essentially the same stimuli were used also in the ILD and Level tasks (See section Procedures for the descriptions of the four tasks).

The low-frequency stimulus was a spectrally shaped multi-component complex (SSMC), which was a harmonic complex with a fundamental frequency ($F_0$) of 100 Hz, consisting of the 7th to 14th harmonics. The components were added in the sine phase. We adopted stimulus parameters as in Moore and Moore (2003) to prevent the listener from using spectral cues (or the excitation-pattern cues) when conducting the tasks: The spectral envelope had a flat passband and sloping edges (5 × $F_0$ centered at 1100 Hz).The overall level of the complex was 54 dB SPL. Threshold equalizing noise (TEN, Moore et al., 2000), extending from 125 to 15000 Hz, was added to mask combination tones and help ensure that the audible parts of the excitation patterns evoked by the harmonic and frequency-shifted tones were the same in the Time task (described later). The TEN level at 1 kHz was set at 30 dB/ERBN, which was 15 dB below the level of the 1100-Hz component.

The high-frequency stimulus was a "transposed stimulus," which was a 4-kHz tone carrier amplitude-modulated with a half-wave rectified 125-Hz sinusoid. It is considered that the auditory-nerve firing is phase locked to the modulator waveform, which provides the cue for judging the ITD and modulation rate of the stimulus (Van De Par and Kohlrausch, 1997; Bernstein, 2001). For the present stimulus, the modulation frequency of 125 Hz was chosen because that was the frequency with which human listeners exhibited the highest ITD sensitivity in the study by Bernstein and Trahiotis (2002).The overall level of the transposed stimulus was set to 65 dB SPL. A continuous, low-pass filtered Gaussian noise (cutoff frequency 1300 Hz; spectrum level 20 dB SPL) was added to prevent the listener from using any information at low spectral frequencies (e.g., combination tones).

### PROCEDURES
#### General procedure

A two-interval two-alternative forced-choice (2I-2AFC) method was used to measure the listener's sensitivities to stimulus parameters. The listener was instructed to choose the "signal" interval by mouse-clicking one of two buttons displayed on a computer

monitor or by pressing a corresponding key on a keyboard. Feedback was given to indicate the correct answer after each response. The two-down/one-up adaptive tracking method was used to estimate discrimination thresholds, corresponding to 70.7% correct (Levitt, 1970). One session of adaptive tracking lasted until twelve turnpoints were obtained. The first two sessions of each task and stimulus type were performed as practice sessions. When the tracking results appeared unstable for a listener with a task, two or three additional practice sessions were added for the listener/task/stimulus. A total of 8–10 sessions besides the practice sessions were conducted for each listener/task/stimulus. The thresholds were computed as the average of all the non-practice sessions. One session set consisted of two consecutive sessions for one task/stimulus. The order of session sets for tasks and stimuli were randomized for each subject in order to reduce the influence of the training and/or order effect.

### Task specific procedures

*ITD task.* In a 2I-2AFC trial, stimuli in the two intervals had ITDs of $+\Delta ITD/2$ and $-\Delta ITD/2$ $\mu s$, respectively (positive and negative ITDs indicate right and left advances in time, respectively). Each stimulus was 400-ms long, including 100-ms raised-cosine onset and offset ramps. The raised cosine ramps at the onset and offset of the stimulus were synchronized between the two ears. Signal and non-signal intervals were separated by a 200-ms silent gap. The listeners were required to indicate the direction of the ITD change between the two intervals on the basis of the laterality of sound images. In each tracking session, $\Delta ITD$ started from 100 to 400 $\mu s$, for low- or high-frequency stimuli, respectively. For the first four turnpoints, $\Delta ITD$ was increased or decreased by a factor of $10^{0.2}$ after one incorrect response or two consecutive incorrect responses, and for the following eight turnpoints, the factor was reduced to $10^{0.05}$. The threshold for the session was computed as the geometric mean of the $\Delta ITD$ at the last eight turnpoints.

*ILD task.* In a 2I-2AFC trial, stimuli in the two intervals had ILDs of $+\Delta ILD/2$ and $-\Delta ILD/2$ dB, respectively (positive and negative ILDs indicate higher and lower levels in the right ear, respectively). Each stimulus was 400-ms long, including 20-ms raised-cosine onset and offset ramps. The listeners were required to indicate the direction of the ILD change between the two intervals on the basis of the laterality of sound images. In each tracking session, $\Delta ILD$ started from 2.5 dB. For the first four turnpoints, $\Delta ILD$ was increased or decreased by 0.5 dB after one incorrect response or two consecutive incorrect responses, and for the following eight turnpoints, the step size was reduced to 0.25 dB. The threshold for the session was computed as the mean of the $\Delta ILD$ at the last eight turnpoints. Other details were the same as in the ITD task.

*Time task.* For the low-frequency stimulus, the listeners were required to detect a common upward frequency shift ($\Delta f$ Hz) imposed on the individual components of the SSMC with the spectral envelope remaining unchanged. The stimulus parameters and measurement methods for a detection threshold for the frequency shift was in accordance with the "TFS1" test developed

by Moore and Sek (2009). It has been reported that such a shift in component frequencies is accompanied with shift in pitch (De Boer, 1956; Schouten et al., 1962; Moore and Moore, 2003). This pitch change was considered to be largely the result of changes in the TFS, since individual frequency components were only intermediately resolved in the auditory periphery (Moore and Moore, 2003) and frequency spacing (corresponding to the periodicity of the envelope) was unchanged. In addition, frequency shifts around a typical threshold value are expected to alter the peripheral excitation pattern by a negligible amount (Moore and Sek, 2009). Therefore, we adopted this task for evaluating the efficiency of neural phase locking to TFS. It should be noted that the pitch of the frequency-shifted SSMC is often ambiguous and listeners could base their judgments not on pitch shifts but on inharmonicity when conducting the tasks (De Boer, 1956; Schouten et al., 1962), and that it was not our intention to use this task for evaluating the pitch mechanism. The "signal" and "non-signal" intervals in the 2I-2AFC method contained RSRS and RRRR sequences, respectively, where R indicates a harmonic complex (i.e., original SSMC) as the reference and S indicates a frequency-shifted SSMC. The listener was required to indicate the signal interval (RSRS).

To assess the peripheral efficiencies of nerural phase locking to stimulus envelope at a high frequency, we adopted a task to measure discriminability of the transposed stimuli with modulation frequencies of 125 Hz and $125 + \Delta f_m$ Hz, referred to as R and S, respectively. Similarly to the low-frequency stimulus, the listener was required to indicate the signal interval (RSRS) as opposed to the non-signal interval (RRRR). When preforming this task, the listeners could base their judgments on changes in pitch associated with the modulation frequency, although the pitch sensation of the transposed stimulus is generally weak and ambiguous (Oxenham et al., 2004).

Commonly for the low- and high-frequency stimuli, an R or S tone had a duration of 100 ms, including 20-ms raised-cosine ramps. There were 100-ms silent intervals between the tones within a sequence in one interval, and there was a 300-ms silent gap between the intervals. In one session of adaptive tracking, $\Delta f$ or $\Delta f_m$ was increased or decreased by a factor of $2^{0.5}$ after one incorrect response and after two consecutive correct responses, respectively, for the first four turnpoints. The factor was reduced to $2^{0.25}$ for the following eight turnpoints. The geometric mean of $\Delta f$ or $\Delta f_m$ was computed across the last eight turnpoints, which represented the threshold for the session.

The maximum frequency shift, $\Delta f$, was limited to 50 Hz (i.e., $0.5 \times F_0$ Hz) in the adaptive tracking for the low-frequency stimulus. For three listeners, the adaptive tracking failed to converge within the maximum $\Delta f$ limit (50 Hz) for at least one session. For those listeners, their performance was evaluated by the method of constant stimuli, instead of the adaptive method. Subjects were given the same instructions as for the adaptive procedure. A session consisted of 20 trials, and subjects completed five sessions. The $\Delta f$ was fixed at the maximum value, 50 Hz. The proportion of correct responses was derived from the pooled responses across 10–12 sessions, and converted to $d'$ (Hacker and Ratcliff, 1979). To make the results comparable to the measures obtained by the adaptive method, the threshold was derived on the assumption

that $d'$ is proportional to the frequency shift (Hopkins and Moore, 2007) and that the adaptive procedure tracked the 70.7% correct point on the psychometric function, which corresponds to a $d'$ of 0.77 with a 2AFC task. This method sometimes yielded values of the threshold greater than the maximum $\Delta f$ limit of 50 Hz. Although such large values of thresholds could not be measured empirically, they could be taken as indicators of the listeners' performance.

**Level task.** In a 2I-2AFC trial, the listeners were required to indicate an interval containing a 400-ms-long SSMC or a transposed stimulus whose central 200-ms portion (including 20-ms raised-cosine ramps) was incremented in level by $\Delta L$ dB, while the other non-signal interval contained an original SSMC or a transposed stimulus. In one session of adaptive tracking, $\Delta L$ started with 6 dB and was increased or decreased by a factor of 2.68 after one incorrect response and after two consecutive correct responses, respectively, for the first four turnpoints. The factor was reduced to 1.67 for the following eight turnpoints. The geometric mean of $\Delta L$ was computed across the last eight turnpoints, which represented the threshold for the session.

## RESULTS

Threshold data for individual tasks and listeners are summarized in **Figure 1**. Each symbol and error bar represents the mean and standard error of thresholds of one listener obtained from multiple sessions. Within each task, the listeners are sorted according to the mean threshold. It should be noted that for the ITD and Time tasks, the means and standard errors are represented on a logarithmic scale. Note also that the thresholds for the low- and high-frequency Time tasks are expressed as fractions to $F_0$ (100 Hz) and modulation rate (125 Hz), respectively. The number in each panel indicates the average across the listeners. One listener (listener number: 10) exhibited an extremely large threshold in the high-frequency Level task (see the rightmost data in the corresponding panel). In the following sections, we report the results of correlation and multiple-regression analyses with and without this listener when they are related to the high-frequency Level task.

**Figures 2–4** show scatter plots comparing individual listeners' thresholds between pairs of tasks. Each panel in the figures shows the data for one combination of tasks, representing 22 listeners with data points. For the Time and ITD tasks, we converted the thresholds to a logarithmic scale when plotting the data and computing the Pearson correlation coefficients.

### LOW-FREQUENCY STIMULUS

Focusing on the results for the low-frequency stimulus (**Figure 2**), one can see statistically significant positive correlations for pairs of ITD and ILD tasks ($r = 0.55$; $p = 0.008$) and of ILD and Level tasks ($r = 0.67$; $p = 0.001$). The pair of Time and ITD tasks showed a weak negative correlation ($r = -0.26$), which was, however, not statistically significant ($p = 0.252$).

We used a multiple linear regression analysis to further explore the factors that might account for inter-individual variability in the lateralization tasks, which might not be revealed by the single correlation analysis. For a given lateralization task of interest

("target task"; i.e., ITD or ILD task), we regarded the threshold for that task as the dependent variable and the thresholds for the remaining three tasks as the explanatory variables. A significant partial correlation of an explanatory task would suggest that the performance of that explanatory task is a good predictor of the performance of the target task. The size of partial correlation coefficient for each explanatory variable could be interpreted as indicating the size of the effect of the variable (or of mechanisms behind the variable) on the performance of the target task, given the values of the other variables are fixed.

The regression analyses were conducted on the threshold data which had been transformed to $z$ scores (i.e., having a mean of 0 and a standard deviation of 1), for individual tasks. Estimated values of partial correlation coefficients are summarized in **Table 1**, along with $p$ values indicating whether the coefficient was significantly different from zero. For the ITD task as the target, the partial correlation coefficient was significant for the ILD task ($p = 0.015$). As for the ILD task as the target, the coefficients for the ITD and Level tasks were significant ($p = 0.015$ and $0.008$, respectively).
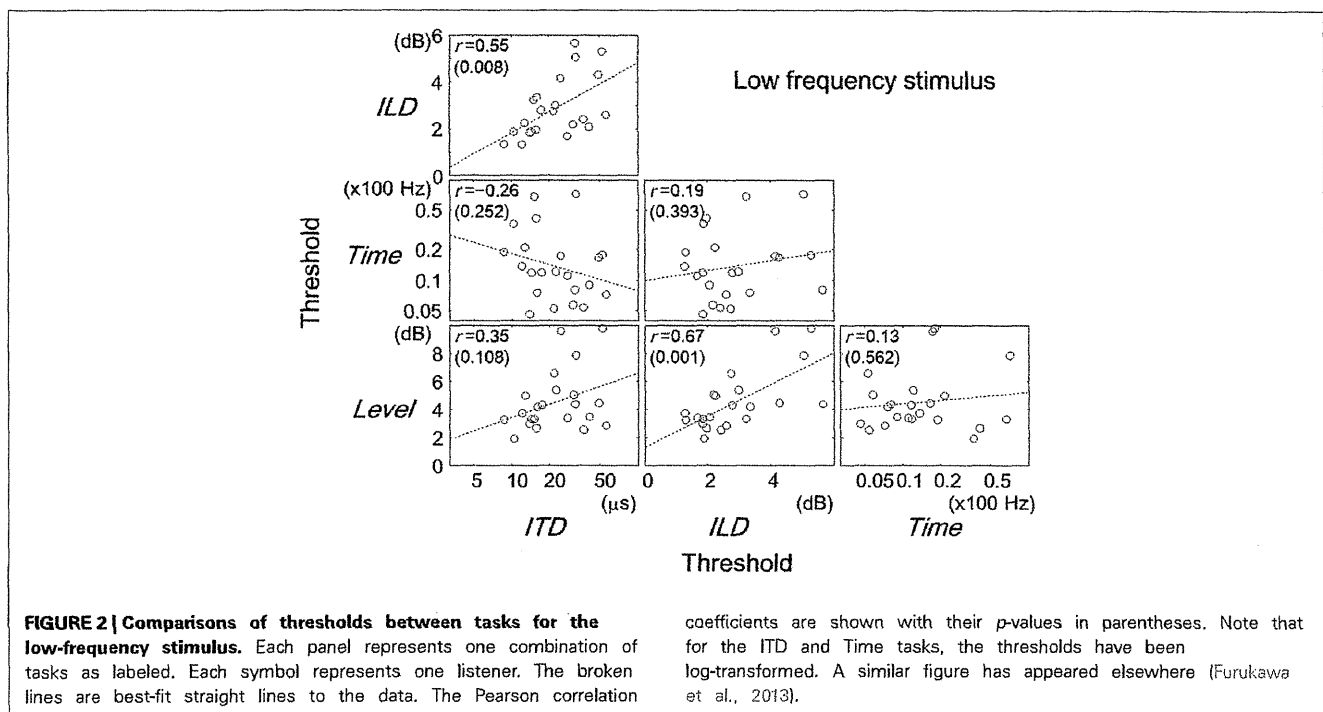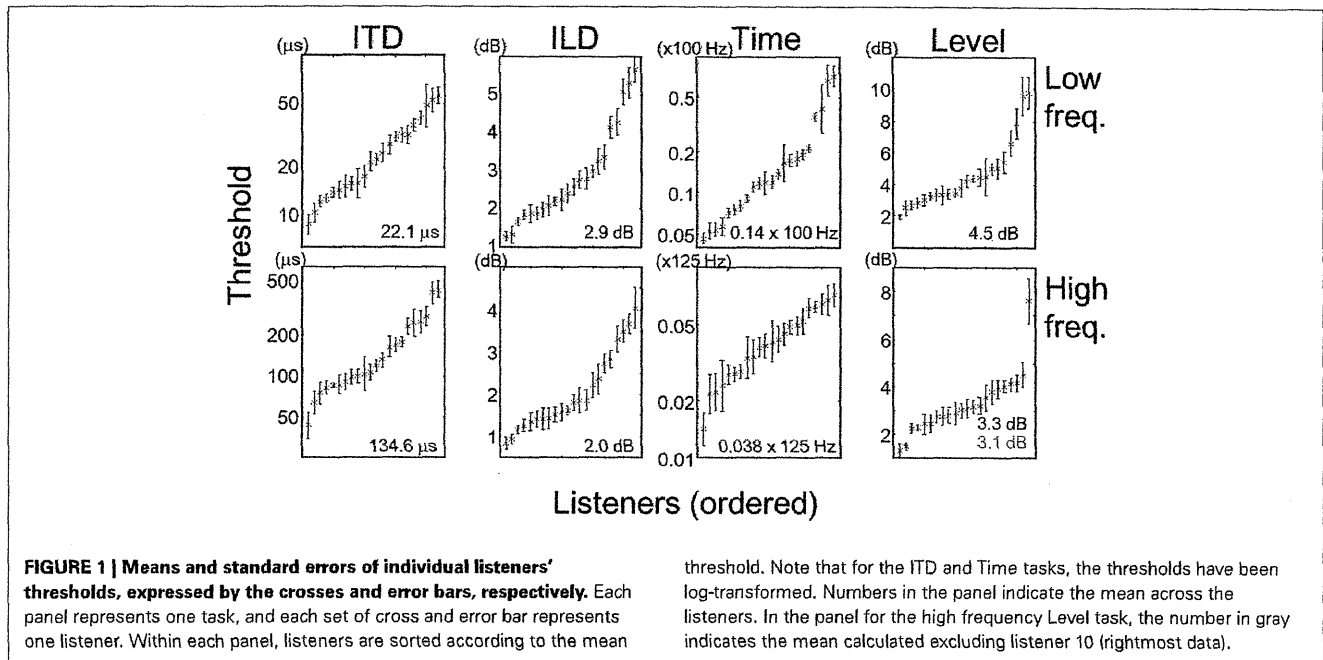
### HIGH-FREQUENCY STIMULUS

Comparisons between the thresholds of the task types for the high-frequency stimulus are represented in **Figure 3**. Significant correlation were found for pairs of the ITD and ILD tasks ($r = 0.66$, $p = 0.001$), and of the ITD and Time tasks ($r = 0.43$, $p = 0.045$). The correlation of the ILD and Level tasks was not significant ($r = 0.41$, $p = 0.056$; $r = 0.24$, $p = 0.295$, when listener 10 was excluded).

The results of the multiple linear regression analysis are shown in **Table 1**. Consistent with the results of the single correlation analysis, the partial correlation coefficients of the ILD and Time tasks were significant when the ITD task was the target ($p = 0.001$ and $0.026$, respectively). The coefficint of the ITD task was significant when the ILD was the target task ($p = 0.001$). Exclusion of listener 10 did not affect the general conclusions of the analysis.

### ACROSS-FREQUENCY COMPARISONS

The correlation of task performance across frequencies can be examined in **Figure 3**. When comparing the thresholds for the same task type, one can see that the correlations were significant for all the tasks except the Time task ($r = 0.56$, $p = 0.007$ for ITD; $r = 0.57$, $p = 0.005$ for ILD; $r = 0.08$, $p = 0.721$ for Time; $r = 0.57$, $p = 0.006$ for Level). The correlation for the Level tasks, however, became non-significant when listener 10 was excluded ($r = 0.31$, $p = 0.165$). A significant correlation for different task types was found in the combination of low-frequency ITD and high-frequency ILD tasks ($r = 0.67$, $p = 0.001$). Significant correlations across frequency regions imply an across-frequency factor that determined the performance of a given task for a frequency region.

Here again, we conducted a multiple linear regression analysis using thresholds (in $z$ scores) of all the combinations of task and stimulus as independent variables. In this analysis, we were specifically interested in the extent to which the performance of one lateralization task could be accounted for by the performances of other tasks, whether the stimuli were in the same or remote

FIGURE 1 | Means and standard errors of individual listeners' thresholds, expressed by the crosses and error bars, respectively. Each panel represents one task, and each set of cross and error bar represents one listener. Within each panel, listeners are sorted according to the mean threshold. Note that for the ITD and Time tasks, the thresholds have been log-transformed. Numbers in the panel indicate the mean across the listeners. In the panel for the high frequency Level task, the number in gray indicates the mean calculated excluding listener 10 (rightmost data).



FIGURE 2 | Comparisons of thresholds between tasks for the low-frequency stimulus. Each panel represents one combination of tasks as labeled. Each symbol represents one listener. The broken lines are best-fit straight lines to the data. The Pearson correlation coefficients are shown with their p-values in parentheses. Note that for the ITD and Time tasks, the thresholds have been log-transformed. A similar figure has appeared elsewhere (Furukawa et al., 2013).

frequency regions, and in identifying tasks where the performance could predict the performance of the target. We used Akaike's information criterion (AIC) as a basis for selecting most effective combination of variables for the regression while avoiding overfitting (Burnham and Anderson, 2002, p. 63). The AIC values were obtained individually for models with all possible combinations of explanatory variables using the LinearModel.fit function of MATLAB. The combination of variables exhibiting the lowest

AIC was employed for constructing the linear model. The results of the analysis are summarized in **Table 2**.

The linear model could account for a relatively large fraction of the variance of the threshold in a target task ($R^2$ ranged between 0.525 and 0.632). In addition, the results of the variable selection were generally in accordance with the findings described earlier: For a given target task and stimulus frequency, the other lateralization task at the same frequency was selected as an
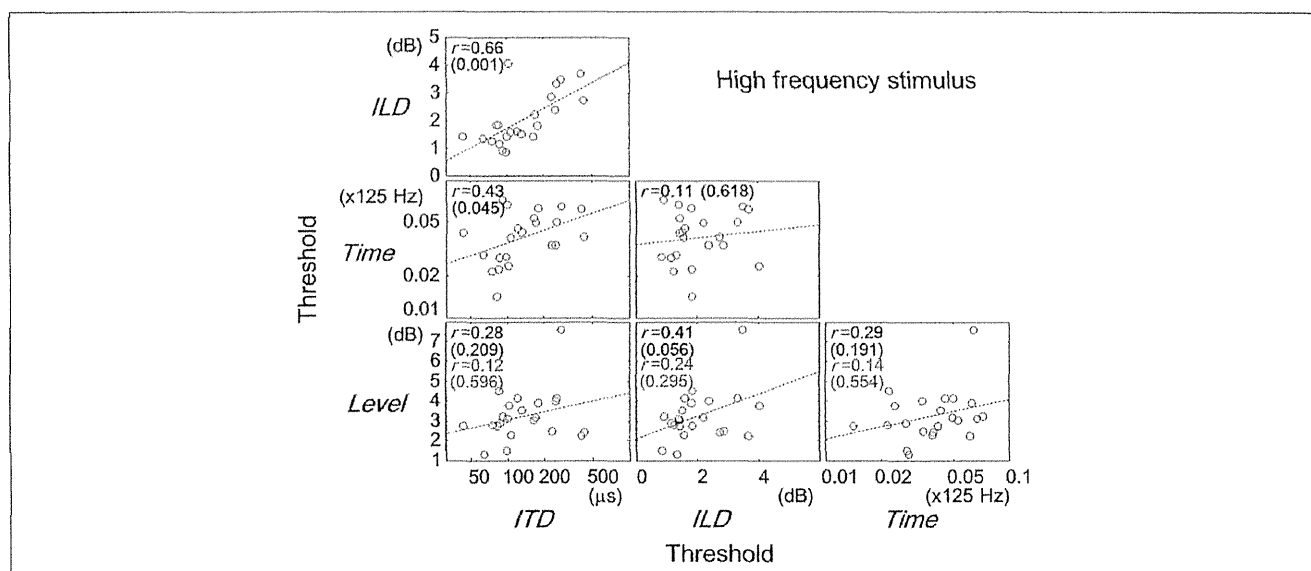
**FIGURE 3 | Same as Figure 2 but for the high-frequency stimulus.** The correlation coefficients and p-values in gray indicates values when listener 10 was excluded.
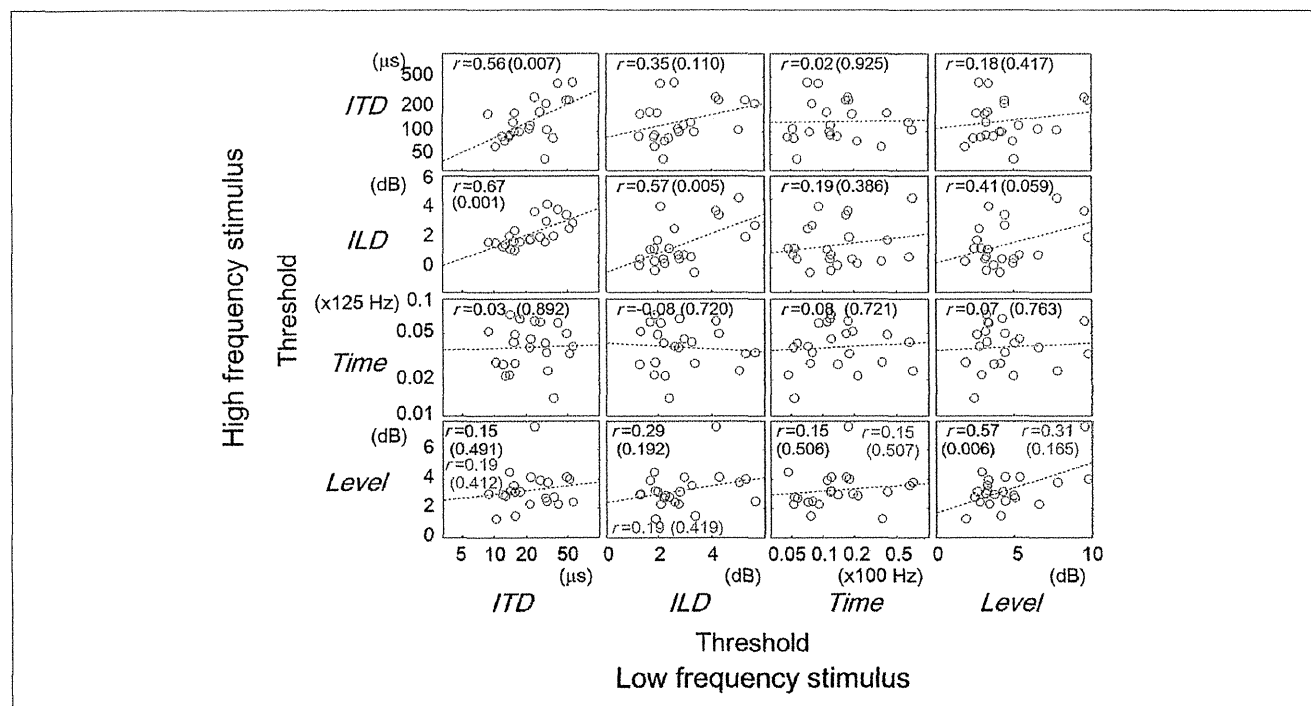


**FIGURE 4 | Comparisons of thresholds between tasks for different frequency regions.** The panels are arranged so that the horizontal and vertical axes represent the data for the low- and high- frequency stimuli, respectively. Other conventions are the same as in **Figures 2, 3.**

explanatory variable (e.g., for the target of the low-frequency ITD task, the low-frequency ILD task was selected), although the coefficients were not always significantly different from zero. It was also confirmed that for target tasks of ITD and ILD tasks, selected explanatory variables included the Time and Level tasks, respectively. The partial correlation coefficient for the low-frequency

ITD task was significant and negative for the target task of low-frequency Time ($-0.425$; $p = 0.007$). Exclusion of listener 10 affected the result for the target task of high-frequency ILD: the high-frequency Level task was no more selected, and the partial correlation coefficient for the low-frequency Time task became significant ($0.346$; $p = 0.030$).

**Table 1 | Summary of multiple regression analyses for low- and high-frequency stimuli.**

| Freq. | Target task | Explanatory tasks (p-value) | | | | Corrected $R^2$ (p-value) |
|---|---|---|---|---|---|---|
| | | ITD | ILD | Time | Level | |
| Low | ITD | – | **0.644** | −0.374 | −0.030 | 0.347 |
| | | | (0.015) | (0.052) | (0.901) | (0.013) |
| | ILD | **0.444** | – | 0.242 | **0.482** | 0.550 |
| | | (0.015) | | (0.139) | (0.008) | (0.001) |
| High | ITD | – | **0.665** | **0.389** | −0.109 | 0.509 |
| | | | (0.001) | (0.026) | (0.541) | (0.001) |
| | | | 0.645 | 0.337 | −0.085 | 0.477 |
| | | | (0.001) | (0.031) | (0.623) | (0.003) |
| | ILD | **0.700** | – | −0.276 | 0.298 | 0.483 |
| | | (0.001) | | (0.137) | (0.090) | (0.002) |
| | | 0.720 | – | −0.286 | 0.190 | 0.495 |
| | | (0.001) | | (0.147) | (0.290) | (0.008) |

*Partial correlation coefficients and the p-values are shown for individual explanatory tasks. Note that the analyses were conducted on the z scores of the threshold data. The bold characters indicate statistically significant correlation (p < 0.05). The rightmost column shows the multiple coefficients of determination (adjusted for degrees of freedom) and their p-values. For the high frequency stimulus, the results obtained when listener 10 was excluded are also shown in gray.*

**Table 2 | Summary of multiple regression analyses on all tasks.**

| Derived formula (p-value for corresponding partial coefficient) | Corrected $R^2$ (p-value) |
|---|---|
| $ITD_L = 0.304 \cdot ILD_L - 0.425 \cdot Time_L + 0.575 \cdot ILD_H$ <br> (0.089)      (0.007)      (0.003) | 0.603 <br> (<0.001) |
| $ILD_L = 0.444 \cdot ITD_L + 0.242 \cdot Time_L + 0.482 \cdot Level_L$ <br> (0.015)      (0.139)      (0.008) | 0.550 <br> (0.001) |
| $ITD_H = 0.623 \cdot ILD_H + 0.362 \cdot Time_H$ <br> (0.001)      (0.027) | 0.525 <br> (<0.001) |
| $ILD_H = 0.545 \cdot ITD_L + 0.297 \cdot Time_L + 0.298 \cdot ITD_H + 0.201 \cdot Level_H$ <br> (0.005)     (0.051)     (0.091)     (0.168) | 0.632 <br> (<0.001) |
| $ILD_H = 0.632 \cdot ITD_L + 0.346 \cdot Time_L + 0.272 \cdot ITD_H$ <br> (0.002)     (0.030)     (0.129) | 0.619 <br> (<0.001) |

*For each target task, explanatory variables (or tasks) were selected based on the AIC (see text). Each symbol (e.g., $ITD_L$) represents the threshold (in z score) of the corresponding task and stimulus (subscripts of L and H represent low- and high-frequency stimuli, respectively). Partial correlation coefficients and p-values are shown for individual explanatory tasks. The bold characters indicate statistically significant coefficients (p < 0.05). The rightmost column shows the multiple coefficients of determination (adjusted for degrees of freedom) and their p-values. For the high frequency ILD task ($ILD_H$), a different result of variable selection was obtained when listener 10 was excluded (indicated in gray).*

## PRINCIPAL COMPONENT ANALYSIS

So far, we have examined associations across tasks through single correlation and the multiple linear regression analyses. Interpretations of the coefficients, however, are often difficult when there are marked correlations among the explanatory
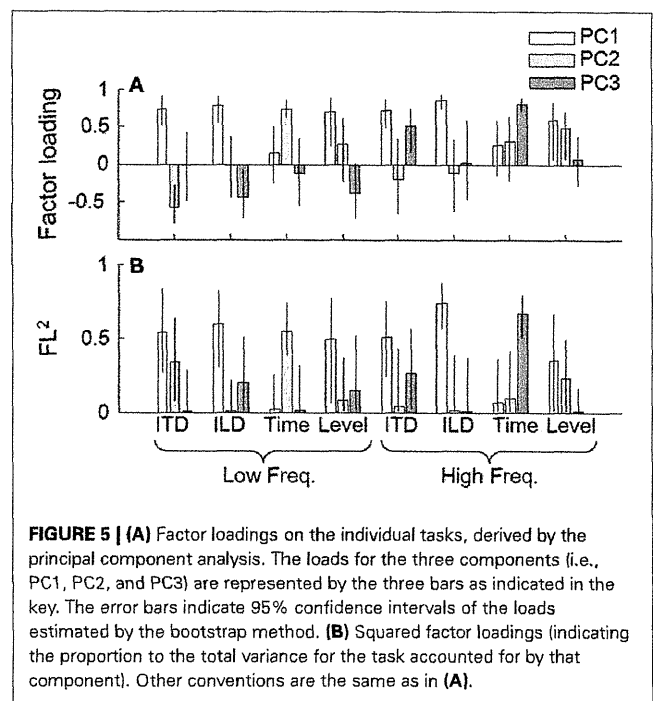


**FIGURE 5 | (A)** Factor loadings on the individual tasks, derived by the principal component analysis. The loads for the three components (i.e., PC1, PC2, and PC3) are represented by the three bars as indicated in the key. The error bars indicate 95% confidence intervals of the loads estimated by the bootstrap method. **(B)** Squared factor loadings (indicating the proportion to the total variance for the task accounted for by that component). Other conventions are the same as in **(A)**.

variables, which was often the case in the present study. It was possible that the performance of the tasks evaluated in the present study could be explained by one or more common underlying factors. To examine this, we conducted a principal component analysis (PCA) on vectors of the eight tasks obtained from the 22 listeners. Before running the analysis, the threshold data were transformed to a logarithmic scale (for the Time and ITD tasks only) and then to z-scores (all the measures). The results indicated that the data could be accounted for well by the first three principal components (PCs; from PC1 to PC3), which had eigenvalues of 3.33, 1.34, and 1.30, respectively. These three PCs accounted for 74.6% of total variance. The factor loadings (FLs) of the three PCs (indicated by gray-scaled bars) and their squared values ($FL^2$s) are shown in **Figures 5A,B**, respectively. The $FL^2$ for a given task by a given PC indicates the proportion to the total variance for the task accounted for by that component.

For all four lateralization tasks, the $FL^2$ values by PC1 were above 0.5. PC1 had positive loads on all the tasks (**Figure 5B**), implying that PC1 reflects the general ability of the listeners to conduct psychophysical tasks. Note, however, that the loads on the low- and high-frequency Time tasks were relatively small. Also, there were marked contributions of PC2 and PC3, depending on the task. For the low-frequency ITD task, PC2 could account for more than 30% of the variance. An examination of FLs revealed that PC2 was associated predominantly with the low-frequency Time task (**Figure 5B**), and the FLs on the low-frequency ITD and Time tasks had opposite signs (**Figure 5A**). This implies that PC2 reflects a factor that had opposing effects on the Pitch and ITD tasks at low frequency. PC3 had appreciable contributions to low-frequency ILD and high-frequency ITD tasks. PC3 was associated with the high-frequency Pitch task,

which had the same sign as the FL on the high-frequency ITD task. To a lesser degree, PC3 also showed some association with the low-frequency Level task, which had the same sign as the FL on the low-frequency ILD task. Exclusion of listener 10 did not alter the general conclusions of the analysis.

## DISCUSSION

The major findings of the present study were: positive correlations between the performance of pairs of lateralization tasks (i.e., ITD and ILD tasks) both within and across stimulus frequencies; a negative correlation for the low-frequency ITD and the Time tasks, revealed by the multiple-regression analysis; a positive correlation for the high-frequency ITD and the Time tasks; and a positive correlation for the low-frequency ILD and the Level tasks.

The mean thresholds obtained in the present study were generally at the same levels of those obtained by earlier comparative studies: ITD: Bernstein and Trahiotis (2002), Furukawa (2008); ILD: Grantham (1984), Furukawa (2008); Time: Plack and Carlyon (1995), Moore and Sek (2009); Level: Moore et al. (1997). Thresholds in the ITD task for the high frequency stimulus were greater than those for the low frequency stimulus by an order of magnitude. This quantitative difference is likely due to the difference in the tone and modulator frequencies and does not immediately indicate mechanistic difference between the frequencies: Typical threshold ITD for the 125-Hz tone, which is considered to be equivalent to the present transposed stimulus in terms of the peripheral phase locking, is comparable to the threshold for the transposed stimulus (see Bernstein and Trahiotis, 2002).

The significant positive correlations generally found between the performance of pairs of lateralization tasks indicate that some degree of inter-individual variation of performance could be accounted for by a common factor or mechanism that underlies lateralization based on both ITDs and ILDs over frequency regions. This notion is supported further by the fact that PC1 found in the PCA had large contributions to all the lateralization tasks. Furukawa (2008) found that the degree of ITD and ILD interaction is greater at high frequency than at low frequency, indicating that the dominance of a common mechanism depends on stimulus frequency or that different mechanisms for ITD and ILD processing are involved for low- and high-frequency stimulus. The present analyses regarding ITD-ILD relations, however, provided no indication of frequency-dependent processes for ITDs and ILDs: The correlation coefficients for the ITD and ILD pairs were not significantly different between low- and high-frequency stimuli ($p = 0.581$; $t$-test after the Fisher transformation of the correlation coefficients). One candidate for such a mechanism is a binaural mechanism that can process both ITDs and ILDs and can operate across frequency regions. Unfortunately, the present study cannot rule out another candidate, which is a non-sensory, higher-order factor related to the experimental procedure. It is possible that the inter-listener variability in the lateralization performance reflected predominantly the difference in procedure-specific skills. It was common across all the lateralization tasks that the listener had to identify the direction in which (toward left or right) intracranial images of

two successive stimulus intervals changed. In the other tasks, on the other hand, the listener was asked to choose the interval that would contain changes in stimulus attributes.

The performance of the ITD task for the high frequency stimulus showed a significant positive correlation with that of the Time task. The following multiple-regression analyses also indicated a significant contribution of the high-frequency pitch task performance to account for the individual variability of the ITD performance. This tendency was captured in PC3 revealed by the PCA, suggesting that this positive correlation reflects a factor that is independent of another non-task-specific factor that determines the listener's overall psychophysical performance (expressed as PC1) or a factor that reflects the relationship of ITD and Time tasks (expressed as PC2; described later). This finding supports our initial hypothesis that the efficiency of neural phase locking to envelope of high frequency stimulus has a significant contribution to ITD-based lateralization performance.

For the low frequency stimulus, however, we failed to observe a positive correlation in the ITD and Time task pairs for the low frequency stimulus. This failure may be attributable to difference in the order of magnitude required for the two tasks: In the low-frequency Time task, a typical threshold of 10-Hz frequency shift of our SSMC stimulus is considered to correspond to difference in peak-to-peak time of TFS by about $100\,\mu s$ (see Moore, 2012 pp. 220-223), which is an order of magnitude greater than a typical ITD threshold of $20\,\mu s$. For the high frequency, on the contrary, a typical threshold $\Delta f_m$ of 4 Hz corresponds to change in the peak-to-peak interval of the modulation by about $250\,\mu s$, which falls in the range of ITD thresholds.

It is interesting that the across-frequency multiple-regression analysis with a variable selection procedure (**Table 2**) revealed that the low-frequency Time-task performance was a significant predictor of the low-frequency ITD-task performance, and it had a *negative* contribution. This negative relationship was observed also as the opposite signs of the FLs for the two tasks in PC2, an independent factor (**Figure 5**). This negative relationship not only was unexpected on the basis of our initial hypothesis but also appears to contradict to earlier reports on hearing-impaired or aged listeners (Strelcyk and Dau, 2009; Hopkins and Moore, 2011). This discrepancy among studies could be explained by postulating two factors that determine the listener's sensitivities to ITDs and the TFS: One factor, associated with the negative correlation, is dominant for normal-hearing listeners. As hearing impairment progresses, the other factor would dominate, resulting in a positive correlation in a population of normal- and hearing-impaired listeners.

One might be concerned about the listener's use of the excitation-pattern or spectral cue as a confounding factor for this negative relationship. Although the change in the excitation level for a typical threshold value (around $\Delta f/F_0 = 0.1$) was expected to be negligible (Moore and Sek, 2009), listeners who exhibited relatively high threshold might rely on the excitation pattern cue, which was usable for frequency shifts near their thresholds. Those listeners might be simply insensitive to the TFS information or might have adapted to placing more weights on the spectral cue than on the temporal cue in pitch judgments through their

long-term experience (McLachlan et al., 2013). However, it is difficult to explain the negative correlation in terms of the use of the excitation-pattern cue: Listeners with general insensitivities to TFS would be expected to be insensitive to ITD also, leading to a positive correlation. We cannot think of obvious association between larger weighting on the place over the temporal cues and better (or poorer) performance in the ITD task.

One explanation for the puzzling negative correlation is that the listeners could use two types of ITD cues when conducting the ITD task, namely, envelope and TFS-based ITDs (since ITDs were imposed on both of those properties), and the performance depended on the relative weights placed on the two cues by individual listeners. It is possible that the envelope ITD of our stimulus was more reliably coded in the auditory system than the TFS-based ITD was. In the Time task, on the other hand, the TFS information could be the main cue for the judgments (although other types of information, such as distortion products by cochlear non-linearity and the excitation pattern, are also arguably potential cues, Oxenham et al., 2009; Micheyl et al., 2010), while the temporal envelope of the stimulus provided no useful cue, since it always had the same repetition rate (100 Hz). Therefore, a listener who places a greater weight on the envelope cue would tend to exhibit better and poor performance in the ITD and Pitch tasks, respectively. It should be noted that this explanation assumes that individual listeners applied more or less the same relative weights on the envelope and TFS invariantly in the Time and ITD tasks.

As for the relationship between the ILD and Level tasks, a significant positive correlation for the low-frequency stimulus supports our initial hypothesis that, at least for the low frequency stimulus, the inter-individual variability of ILD performance reflects the difference in the efficiency of intensity coding at a processing stage earlier than binaural interaction. One might be concerned that the listeners in the ILD task based their judgments primarily on the change of stimulus level within a single ear, and thus the ILD task measured essentially monaural sensitivity to level change. However, this is not likely, as supported by the suggestion of Bernstein (2004) that the listener's judgment is likely to be based on changes in the position of an intracranial image, not on the monaural cues.

## AUTHOR CONTRIBUTIONS

## ACKNOWLEDGMENTS

## REFERENCES

Bernstein, L. R. (2001). Auditory processing of interaural timing information: new insights. *J. Neurosci. Res.* 66, 1035–1046. doi: 10.1002/jnr. 10103

Bernstein, L. R. (2004). Sensitivity to interaural intensitive disparities: listeners' use of potential cues. *J. Acoust. Soc. Am.* 115, 3156–3160. doi: 10.1121/1. 1719025

Bernstein, L. R., and Trahiotis, C. (2002). Enhancing sensitivity to interaural delays at high frequencies by using "transposed stimuli." *J. Acoust. Soc. Am.* 112, 1026–1036. doi: 10.1121/1.1497620

Burnham, K. P., and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* New York, NY: Springer-Verlag.

De Boer, E. (1956). Pitch of inharmonic signals. *Nature* 178, 535–536. doi: 10.1038/178535a0

Furukawa, S. (2008). Detection of combined changes in interaural time and intensity differences: segregated mechanisms in cue type and in operating frequency range? *J. Acoust. Soc. Am.* 123, 1602–1617. doi: 10.1121/1. 2835226

Furukawa, S., Washizawa, S., Ochi, A., and Kashino, M. (2013). "How independent are the pitch and interaural-time-difference mechanisms that rely on temporal fine structure information?," in *Basic Aspects of Hearing*, eds B. C. J. Moore, R. D. Patterson, I. M. Winter, R. P. Carlyon, and H. E. Gockel (New York, NY: Springer), 91–99.

Grantham, D. W. (1984). Interaural intensity discrimination - insensitivity at 1000 Hz. *J. Acoust. Soc. Am.* 75, 1191–1194. doi: 10.1121/1.390769

Hacker, M. J., and Ratcliff, R. (1979). A revised table of d' for M-alternative forced choice. *Percept. Psychophys.* 26, 168–170. doi: 10.3758/BF03208311

Hopkins, K., and Moore, B. C. (2007). Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information. *J. Acoust. Soc. Am.* 122, 1055–1068. doi: 10.1121/1.2749457

Hopkins, K., and Moore, B. C. (2011). The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise. *J. Acoust. Soc. Am.* 130, 334–349. doi: 10.1121/1.3585848

Levitt, H. (1970). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* 49, 467–477. doi: 10.1121/1.1912375

McLachlan, N. M., Marco, D. J., and Wilson, S. J. (2013). The musical environment and auditory plasticity: hearing the pitch of percussion. *Front. Psychol.* 4:768. doi: 10.3389/fpsyg.2013.00768

Micheyl, C., Dai, H., and Oxenham, A. J. (2010). On the possible influence of spectral- and temporal-envelope cues in tests of sensitivity to temporal fine structure. *J. Acoust. Soc. Am.* 127, 1809–1810. doi: 10.1121/1.3384106

Moore, B. C., Huss, M., Vickers, D. A., Glasberg, B. R., and Alcantara, J. I. (2000). A test for the diagnosis of dead regions in the cochlea. *Br. J. Audiol.* 34, 205–224. doi: 10.3109/03005364000000131

Moore, B. C. J. (2012). *An Introduction to the Psychology of Hearing.* Bingley: Emerald.

Moore, B. C., Peters, R. W., Kohlrausch, A., and Van De Par, S. (1997). Detection of increments and decrements in sinusoids as a function of frequency, increment, and decrement duration and pedestal duration. *J. Acoust. Soc. Am.* 102, 2954–2965. doi: 10.1121/1.420350

Moore, B. C., and Sek, A. (2009). Development of a fast method for determining sensitivity to temporal fine structure. *Int. J. Audiol.* 48, 161–171. doi: 10.1080/14992020802475235

Moore, G. A., and Moore, B. C. (2003). Perception of the low pitch of frequency-shifted complexes. *J. Acoust. Soc. Am.* 113, 977–985. doi: 10.1121/ 1.1536631

Oxenham, A. J., Bernstein, J. G., and Penagos, H. (2004). Correct tonotopic representation is necessary for complex pitch perception. *Proc. Natl. Acad. Sci. U.S.A.* 101, 1421–1425. doi: 10.1073/pnas.0306958101

Oxenham, A. J., Micheyl, C., and Keebler, M. V. (2009). Can temporal fine structure represent the fundamental frequency of unresolved harmonics? *J. Acoust. Soc. Am.* 125, 2189. doi: 10.1121/1.3089220

Plack, C. J., and Carlyon, R. P. (1995). Differences in frequency modulation detection and fundamental frequency discrimination between complex tones consisting of resolved and unresolved harmonics. *J. Acoust. Soc. Am.* 98, 1355–1364. doi: 10.1121/1.413471

Schouten, J., Ritsma, R. J., and Cardozo, B. (1962). Pitch of the Residue. *J. Acoust. Soc. Am.* 34, 1418–1424. doi: 10.1121/1.1918360

Strelcyk, O., and Dau, T. (2009). Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing. *J. Acoust. Soc. Am.* 125, 3328–3345. doi: 10.1121/1.3 097469

Strouse, A., Ashmead, D. H., Ohde, R. N., and Grantham, D. W. (1998). Temporal processing in the aging auditory system. *J. Acoust. Soc. Am.* 104, 2385–2399. doi: 10.1121/1.423748

Van De Par, S., and Kohlrausch, A. (1997). A new approach to comparing binaural masking level differences at low and high frequencies. *J. Acoust. Soc. Am.* 101, 1671–1680. doi: 10.1121/ 1.418151

Zeng, F. G., Kong, Y. Y., Michalewski, H. J., and Starr, A. (2005). Perceptual consequences of disrupted auditory nerve activity. *J. Neurophysiol.* 93, 3050–3063. doi: 10.1152/jn.00985.2004

# Distinct neural mechanisms of tonal processing between musicians and non-musicians

Kaoru Amemiya [a,*], Shotaro Karino [b], Tomohiro Ishizu [c], Masato Yumoto [d], Tatsuya Yamasoba [a,b]

[a] Department of Sensory and Motor Neuroscience, Graduate School of Medicine, University of Tokyo, Tokyo, Japan
[b] Department of Otolaryngology – Head and Neck Surgery, Faculty of Medicine, University of Tokyo, Tokyo, Japan
[c] Wellcome Laboratory of Neurobiology, University College London, London, United Kingdom
[d] Department of Laboratory Medicine, Faculty of Medicine, University of Tokyo, Tokyo, Japan

## ARTICLE INFO

## HIGHLIGHTS

- The sense of a melody's completion is influenced by the combination of the preceding melody and the ending tone.
- Musicians and non-musicians assessed the sense of a melody's completion equally.
- The neural responses differed between musicians and non-musicians, suggesting the neural processes involved in the assessment may be distinct between groups.

## ABSTRACT

Objective: Both behavioral and neural responses to deviant melody endings can be enhanced through musical training. Yet it is unknown whether there are any differences in the neural responses of musicians and non-musicians given no difference in their behavioral responses. It is also unknown whether the melody preceding the fixed final tone influences the sense of completion.
Methods: We recorded neuromagnetic responses in ten musicians and ten non-musicians while they were evaluating the sense of completion associated with melodies.
Results: The sense of a melody's completion was influenced by the combination of the preceding melody and the ending tones. The N1 had shorter latency in musicians, while the sustained field had larger amplitudes in non-musicians.
Conclusions: Musicians and non-musicians rated the sense of completion similarly, yet neural responses differed between the groups.
Significance: These findings suggest that neural processes in musicians and non-musicians may be distinct even when the sense of completion is assessed similarly in both groups. In other words, there might be specific tonal processing available to non-musicians which can compensate for their lack of musical training.
© 2013 International Federation of Clinical Neurophysiology. Published by Elsevier Ireland Ltd. All rights reserved.

## 1. Introduction

In western tonal music, the notes within an octave are hierarchically organized according to the degree of consonance to the most essential 'tonic' tone, i.e., C in the key of C major (Fig. 1A). The tone of C is highest in the hierarchy for C major/minor keys because of its perfect overlap in frequency with the tonic tone (e.g., between two Cs one octave apart). According to the theory of harmonics, the tonic tone or tonic chord (e.g., C, E and G in the

* Corresponding author. Address: 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8655, Japan. Tel.: +81 3 5800 8665; fax: +81 3 3814 9486.
E-mail address: caorain-tky@umin.ac.jp (K. Amemiya).

key of C major) has the most important role and should be located at the beginning and end of the melody. Therefore the tonic chord strongly shapes the tonality of the starting tone and the ending pattern, especially in the ending pattern, and influences the sense of completeness and stability that is conveyed by the ending.

Previous experiments have demonstrated that the sense of completion is automatically attenuated when the melody ends with a non-tonic tone or chord, e.g., a dissonant chord or a subdominant tone or chord (e.g., F, A and C in the key of C major) (for a review: Koelsch and Siebel, 2005). Neurophysiological experiments have shown that this phenomenon is observed as early negative responses to infrequent non-tonic stimuli in the oddball paradigm (early right anterior negativity: ERAN, early negativity, EN) or
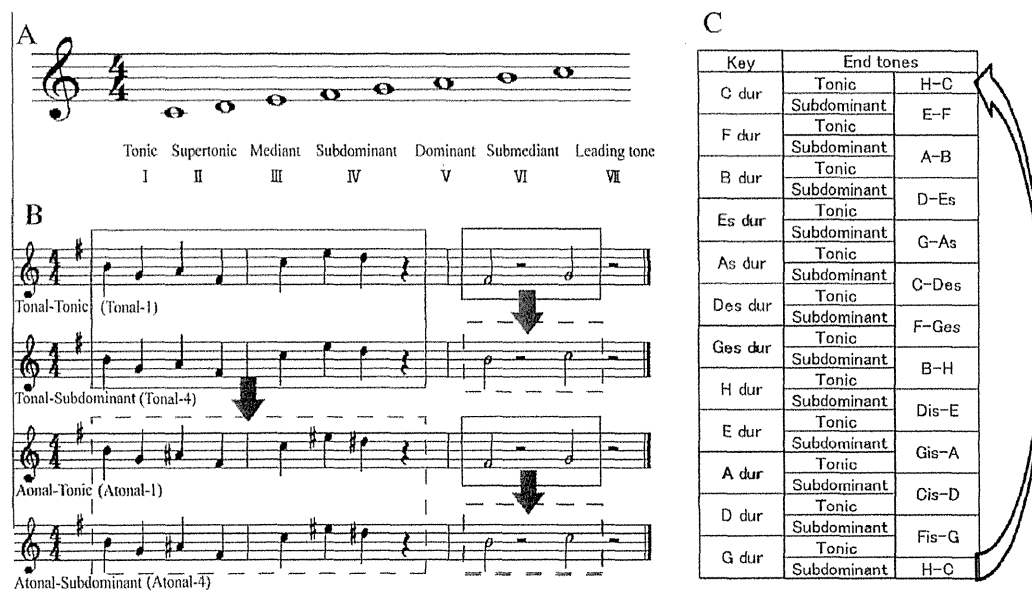
Fig. 1. The nomenclature of the musical scale (A), examples of the four stimulus types (B) and a table showing the correspondence between melodies and ending patterns (C). The tonal-tonic melody (Tonal-Tonic) was based on harmonic theory and used as a prototype melody; the other three melodies were made by modifying Tonal-Tonic. The tonal-subdominant melody (Tonal-Subdominant) was the same melody except with a subdominant ending. The same two ending patterns were also paired with an atonal melody (Atonal-Tonic and Atonal-Subdominant, respectively). In this example in the key of G major, the tonic ending pattern was composed of F#-G. This ending pattern (F#-G) was also used as the subdominant ending pattern for the melodies in the key of D major. Likewise, the subdominant ending pattern (B-C) for the key of G major was also used as the tonic ending pattern for the key of C major. In this way, each tonic and subdominant ending pattern was used with both a tonal and an atonal melody.

enhanced N1/N1m, P2/P2m amplitudes in response to non-tonic stimuli in the non-oddball paradigm applying a simple comparison of iso-frequent endings with tonic or non-tonic chords (Hashimoto et al., 2000; Otsuka et al., 2008).

Specific responses to melody endings with non-tonic tones or chords were evoked consistently, regardless of whether subjects were adults or young children, whether they were musically trained, and whether they were paying attention to the music (Koelsch et al., 2003; Oberecker et al., 2005; Brattico et al., 2006). These findings suggest that the ability to detect unusual endings may be inherent or that only minimal exposure to music might be sufficient to induce this ability. The ability to detect unusual endings and the strength of the corresponding neural responses has not only been reported to be automatic but has also been demonstrated to be enhanced by musical training (Tervaniemi et al., 2001). Although not all previous experiments have explicitly shown behavioral differences between musicians and non-musicians, there is an a priori assumption that music training enhances tonal processing ability. However, it remains unknown whether musicians and non-musicians exhibit different neural responses when their behavioral performances are equally controlled.

The present experiment aims to elucidate the difference between musicians and non-musicians in terms of the neural mechanisms of tonal processing (Madsen and Geringer, 1990; Fujioka et al., 2004). It has been suggested that musical training can enhance the predictive aspect of listening (Baumann et al., 2008; Strait et al., 2010). Based on previous studies, we hypothesized that musicians process melody with less effort than non-musicians, while non-musicians process melody using extra effort such as attention control. Therefore, we further hypothesized that the relatively earlier component of tonal processing, N1m, which reflects automatic processing, training effect, expectation, or focal attention, would be dominant in musicians because musicians can listen to music while also expecting and predicting future ending. Non-musicians, in contrast, would rely more heavily on the relatively

later and more sustained component known as sustained field (SF), which is known to be related to sustained attention (Gunji et al., 2003), because they have to listen to music with less expectation of ending so that the cognitive burden would be heavier than non-musicians. If there are any differences between musicians and non-musicians in relatively early responses, this effect might be related to training, more specifically, to the musicians' greater experience in listening to piano tones. If any differences exist in the SF component, sustained attention or cognitive burden might have contributed to them (Minati et al., 2008, 2009, 2010). There are many reports which demonstrate that the attention modulates much earlier component (such as Auditory Steady-State Evoked Responses: ASSR). However, auditory attention differentially affects the generators of ASSR, N1m, and SF (Okamoto et al., 2011). While the physical aspects of a tone affect the early component (ASSR), sustained attention affects the later component (SF). N1m is influenced by both physical aspects and focused attention. In the present study, we used an identical stimulus set, allowing us to focus on N1m and SF. There is another important component of attention, P3a, but there was little consistent activity in the MEG at the time when the N2 and P3 components of the ERP occurred (Siedenberg et al., 1996). Accordingly, we intend to focus on N1m and SF, both of which are shown in the experiment reporting the effect of attention.

To explore the listener's recognition of a sense of completion, previous studies have compared differences between the expected final chord and unexpected final chords while the preceding melody remains unchanged. However, potentially different neural activities are also elicited by different final chords or various components of chords. In order to avoid these potential differences, we used a single tone melody without the oddball paradigm. Additionally, we sought to investigate changes in neural responses that occur not when the final tone/chord is replaced but when the syntactic relationship between the preceding melody context and the final tone is changed. Therefore, we manipulated the preceding melody while keeping the final tone fixed.

The aim of the present study is twofold. The first is to examine whether there are any differences in the neural responses of musicians and non-musicians to tonic and non-tonic (subdominant) melody endings given no difference in their behavioral responses. The second is to examine whether the sense of completion is context-dependent or not, that is, whether the context provided by the tonality of the preceding melody influences the sense of completion even when the final tone is fixed.

## 2. Experimental procedure

### 2.1. Participants

Ten music experts (age: 20–29, average: 22.1) and ten musically untrained people (age: 21–29, average: 24.6) were enrolled in this study. All of them were right-handed (Edinburgh inventory) and had normal hearing (Oldfield, 1971). All of the music experts had majored in piano at a college of music and had received classical piano instruction for more than ten years (range: 15–25). Non-musicians had no formal musical training. Written informed consent was obtained from all participants after the purpose and procedures of the experiment had been fully explained. The experiment was approved by the Ethics Committee of the University of Tokyo.

### 2.2. Stimuli

Usually, a tonal melody ends with its tonic tone (1: C in C major; for names and numbers in tonal order, see Fig. 1A); for this reason, a melody ending with a subdominant tone (such as 4: F in C major) elicits a sense of incompleteness. We presented melodies with tonic endings and melodies with subdominant endings to the subjects. The melodies themselves were of two types, one tonal melody and one non-perfect tonal (atonal) melody, in order to elucidate the effect of tonality within the melody preceding the ending note. Thus our stimuli consisted of four patterns: two melodies (tonal vs. atonal) × two endings (tonic ending vs. subdominant ending). All twelve major keys were used.

Fig. 1B depicts how we generated the experimental melodies. First, a prototype melody was prepared for each key (12 melodies in all); this was a tonal melody ending with a tonic tone (Audio S1: Tonal-Tonic). Next the final tone was replaced with a subdominant tone without changing the preceding melody (Audio S2: Tonal-Subdominant). Atonal melodies were generated by modifying the tonal melodies (for details, see following paragraph), and the tonic and subdominant endings were applied to the atonal melody for each key (Atonal with tonic ending: Audio S3: Atonal-Tonic; Atonal with subdominant ending: Audio S4: Atonal-Subdominant).

On the basis of the theory of harmonics, the leading tone (subtonic tone: B in C major) was placed before the final tonic tone, so that the final tone was always preceded by the semitone below. The ending of T-1 consisted of continuous tones: leading tone to tonic tone (H–C in C major). To generate a physically similar ending pattern for T-4, the mediant tone (half tone lower than the subdominant: E in C major) was placed before the subdominant (ending) tone (E–F in C major). This ensured that the ending always featured a half-tone relationship. Because these four tones were originally included in the major key, neither of these two endings deviated from its preceding melody.

Since the semitone relationship is physically identical between the tonic (subtonic-tonic) and subdominant (mediant-subdominant) endings, the subtonic-tonic relationship in one key could be the mediant-subdominant in another key. In C-major, for example, the subdominant ending is F and the preceding tone is E; this ending is the same as the tonic ending in the key of F-major

(subtonic: E followed by tonic tone: F). In this way, every ending pattern was used as both a dominant and a subdominant ending pattern. The only difference determining whether an ending was dominant or subdominant was the preceding melody (priming melody) (Fig. 1C).

The procedure by which we generated our atonal melody was as follows. Because each of the mediant-subdominant or subtonic-tonic tone sequences was used as the ending of a tonal melody, the same sequences were also used as the endings of atonal melody patterns. Considering the effect of the listeners' short term memory, these four tones could not be changed in the atonal melodies. Excepting these four tones, therefore, the super tonic (2: D in C major), dominant (5: G in C major) and submediant (6: A in C major) were each played one semitone up (Dis, Gis, Ais in C major each) to convert the tonal melody into an atonal melody.

To ensure that the 12 melodies – one for each key – were equivalent, we adopted the following criteria. Each melody consisted of seven tones in its key (in C major, for example, C, D, E, F, G, A and H) and each tone was used only once per melody. The starting tone was one of the tones in the tonic triad (C or E in C major); these were randomly but equally assigned for all keys. Although the dominant tone is included in the tonic triad, we did not use this tone as a starting tone because it was modified in the atonal version of the melody.

Stimuli were generated using digital piano tones by Score Maker 3 (Kawai, Japan), and pseudo-randomly sequenced using a STIM2 system (Compumedics Neuroscan, El Paso, TX, USA). The duration of each melody tone was 300 ms and that of each ending tone was 600 ms. The average interval between the melody and the ending tone was 600 ms; this interval was randomly varied within a range of ±10% to prevent contamination from the late components of the response to the previous stimulus. The inter-trial interval was 2000 ms regardless of subjects' response. Stimuli were delivered binaurally at a sound pressure of 60 dB through ER-3A earphones (Etymotic Research, Elk Grove Village, IL, USA).

### 2.3. Task

We employed a two-alternative forced-choice task. When participants judged that a melody had ended with a sense of completeness, they were required to press the button on a response pad on their left with their left index finger. When they judged that a melody had ended with a sense of incompleteness, they were required to press a button on their right with their right index finger. After several training sessions allowing participants to grasp the timing of the melodies' endings and the interval between sequences, they were instructed to press either the left or the right button within the interval between sequences to exclude the possibility of contamination related to movement preparation. These behavioral data were also collected using the STIM2 system.

### 2.4. MEG data acquisition

MEG data were acquired in a magnetically shielded room using a helmet-shaped SQUID (superconducting quantum interference device) system with 204 first-order planar gradiometers at 102 measuring sites and 102 magnetometers (Elekta-Neuromag, Helsinki, Finland). Signals were digitized on-line with a bandwidth of 0.1–200 Hz at a sampling rate of 1000 Hz. Epochs with artifacts exceeding 3 pT/cm in any MEG channel were discarded. At least 90 artifact-free epochs were recorded per stimulus-type per participant, and epochs for the same melody category were averaged for analysis.

External noise was removed from the MEG data (including empty-room data) using the Signal-Space Separation temporal extension method (Taulu et al., 2005; Taulu and Hari, 2009)

implemented with the Neuromag software (Elekta-Neuromag, Helsinki, Finland) (Wehner et al., 2008; Henson et al., 2009). We used one of two offline low-pass filters: either (1) the averaged data from the analysis time between 0 and 300 ms were filtered with an off-line low-pass filter of 40 Hz, or (2) we applied another low-pass filter (<5 Hz) for the later component (>300 ms) instead, in order to elucidate the auditory evoked sustained field (Siedenberg et al., 1996; Okamoto et al., 2011).

### 2.5. Statistical analysis

For behavioral data, subjects' judgments of completeness were gathered for each melody type, and these data were analyzed using repeated ANOVA (Subjects × Tonality × End-type). Post-hoc tests were performed according to Ryan's method.

MEG recording was triggered by the onset of the last tone of each melody, and analysis time ranged from −100 to 700 ms. The mean of the waveforms in the 100-ms period before onset served as the baseline. For analysis, MEG data were root mean-squared (RMS) to yield the RMS amplitude of the magnetic field of the predetermined perisylvian channels, and repeated ANOVA was conducted using these RMS values. This subset of sensors, one for each hemisphere, was centered over the auditory cortex and typically used 22 planar sensor pairs (44 channels), especially for analysis on acoustic studies (Yumoto et al., 2005, 2007). For analysis, we used only pairs of gradiometers, because gradiometers and magnetometers measure different values, so that we could calculate RMS values using only gradiometers.

The first positive peak, which occurred around 40–70 ms, was labeled P1m, the first negative peak, which occurred around 80–150 ms, was labeled N1m, the next positive peak, which occurred around 160–200 ms, was labeled P2m, and the last clear negative sustained activation, which occurred around 300–500 ms, was labeled SF. All were clearly observed in most participants for each stimulus condition. Their polarities were judged based on dipole orientation. Since not all participants showed a clear response, especially in the P1m component (in two or three non-musicians, peak for P1m were not obtained for each condition especially in the left hemisphere), we analyzed only the N1m, P2m and SF components in the present study. For each participant, the highest/lowest detectable peaks within each window were picked up for N1m and P2m, respectively. SF was calculated as the mean amplitude between 300 and 500 ms.

Three-way repeated ANOVA ('Group: Musicians vs. Non-musicians', 'Tonality: Tonal vs. Atonal' and 'End-type: Tonic tone vs. Subdominant tone') was performed on the peak latencies and amplitudes of mean perisylvian RMS values of N1m, P2m, and SF. The main effect of 'Group' shows the difference between musicians and non-musicians, while that of 'Tonality' indicates whether the melody was tonal or atonal and that of 'End-type' indicates whether the melody ended with a tonic or subdominant tone.

For dipole estimation, the sources of each component were modeled, separately for each hemisphere, as single equivalent current dipoles (ECD) in a spherical conductor, independently for the averaged data of each category in each participant. The ECDs were calculated at the peak latencies from the same predefined perisylvian channels. The estimated ECDs were described according to a head-based coordinate system. The x-axis passes through the two preauricular points digitized before data acquisition with the positive direction defined as the right. The y-axis passes through the nasion and is normal to the x-axis. The z-axis points upward according to the right-handed rule and is normal to the x–y plane. ECD data were excluded if they did not fulfill the limit of goodness-of-fit (>75%) and the upper limit of confidence volume (<1000 mm$^3$).

To allow for the mapping of each individual brain, high-resolution T-1 weighted MR images (repetition time, 30 ms; field of view, 192 × 192 mm$^2$; resolution, 1 × 1 × 1 mm$^3$) were obtained for all participants (Trio Tim, Siemens, Erlangen, Germany), then overlaid with ECD.

## 3. Results

### 3.1. Behavioral results

The completeness ratings assigned to each stimulus were averaged. Fig. 2 presents the proportion of ratings in the forced choice of both musicians and non-musicians. Significant main effects were obtained both for tonality and end-type (for Tonality: $F(1, 16) = 66.8, p < 0.001$; for End-type: $F(1, 16) = 46.8, p < 0.001$). Significant interaction was also found between end-type and tonality ($F(1, 1) = 32.0, p < 0.001$). The following post hoc simple main effects of interaction were significant: tonic for tonality ($F(1, 32) = 98.7, p < 0.001$), subdominant for tonality ($F(1, 32) = 12.5, p < 0.005$), tonic for end-type ($F(1, 32) = 77.0, p < 0.001$), and subdominant for end-type ($F(1, 32) = 8.7, p < 0.01$). There was no main effect for group and no interaction between group and either of the other two variables, even though there seems to be a difference especially in the Tonal IV condition.

ANOVA results revealed that tonal melodies elicited a significantly stronger sense of completeness than atonal melodies did (Fig. 2B) and that tonic tone endings were judged to give more completeness to stimuli than subdominant endings did (Fig. 2C).

### 3.2. Amplitude and latency of MEG responses

Fig. 3A shows grand-averaged MEG waveforms in musicians and predetermined perisylvian channels (see Methods). Both sets of waveforms exhibited a clear N1m and P2m followed by a large negative SF for each stimulus condition and for each group. The RMS peak latencies and amplitudes of the N1m and P2m peaks and the mean amplitude of SF were measured in the predetermined temporal channels for each subject.

#### 3.2.1. N1m
3.2.1.1. N1m latency. The main effect of Group (group) on N1m latency in the right hemisphere was significant ($F(1, 16) = 5.93; p < 0.05$). Specifically, the latency of N1m was shorter in musicians than in non-musicians. Interaction between group and tonality was obtained but was not significant (Fig. 3B).

In contrast, neither a main effect of any the three factors nor any interaction between them was found in the left hemisphere.

3.2.1.2. N1m amplitude. With respect to peak RMS amplitude, a significant interaction between group and tonality was present in the right hemisphere only ($F(1, 1) = 5.89, p < 0.05$), although a significant post hoc main effect was not observed. No main effect of any of the factors was significant in the right hemisphere. In the left hemisphere, neither main effect nor interaction was found Figs. (3B and 4).

#### 3.2.2. P2m
For P2m peak latency, there were no main effects of interactions between tonality, end-type or group in either hemisphere. Similarly, there were no main effects and no interactions affecting the P2m amplitude in either hemisphere.

#### 3.2.3. SF
Neither main effect nor interaction was found in the left hemisphere. In the right hemisphere, there was a significant interaction