3. 未来100年を進む私が目指すもの

3-2 未来の音の収録・再生・編集技術の実現に向けて

The Future of Sound Recording, Reproduction, and Editing

小山翔一

1. はじめに

人間は周囲で起こるあらゆる物理現象を，様々な感覚器を用いて知覚する．周囲の環境は膨大な情報量を持っているにもかかわらず，我々は認識過程を通じて必要な要素を抽出して処理することができる．例えば聴覚においては，雑踏の中でも着目した音声を聞き取ることができる現象（カクテルパーティ効果）がよく知られている[1]．しかしながら，このように意識的に認識する情報以外の「無意識」の情報も，やはり我々に重要な影響を与えている[2]．再び聴覚を例に出せば，音源までの距離の推定は，音の強度だけでなく残響なども手掛かりとして，無意識的に行われる[3]．視覚障害者が障害物を検知するために，聴覚を利用しているという研究報告もある[4]．いわゆる「臨場感」や「雰囲気」を作り出す要素も，このような普段は余り意識しない情報の中に含まれていることは疑いないであろう．

筆者はこれまでに，高い臨場感を持つ音響システムの実現に向けた音響信号処理技術の研究を行ってきた．本稿では，「音」に焦点を当て，未来の音の収録・再生・編集技術に関して，筆者のこれまでの研究についても触れながら，私見を述べたいと思う．

2. 物理現象の記録・伝送と再生

ある時刻，ある場所で生じた物理現象は，本来，二度と起こることはない．物理現象を記録あるいは伝送し，再生しようという技術的な試みは，特に視覚と聴覚を対象として古くから取り組まれてきた．この技術的な進歩が，人々の生活だけでなく，芸術分野をはじめとして文化的にも大きな影響を与えてきたと言えるだろう[5]．現在では，バーチャルリアリティが学術的な分野として確立され，五感を含む様々な感覚を再構成，あるいは拡張する技術の構築を目的としている[6]．

音の記録・伝送と再生技術に関しては，1876年のAlexander Graham Bellによる電話機の発明[7]や，1877年のThomas Edisonによる蓄音機の発明[8]がその起源と考えられる．現在では一般に，マイクロホンを用いて音波を集音し，ディジタルデータに変換した上で，記録メディア上に保存，あるいは遠隔地に伝送される．再生においては，この情報はアナログ信号へと再変換され，スピーカを用いて音波を物理的に再構成する．ところが，このような従来の音の収録・再生の過程は，大幅な情報量の削減を伴う．すなわち，本来は空間と時間の四次元情報であるはずの音の情報を，一次元情報である時間信号（あるいはその複数の組合せ）に変換しているのである．どれだけサンプリング周波数を上げて，時間的な解像度を高くしたとしても，空間的な情報のほとんどは失われたままである．このような方式は古くから標準的となっており，これによって音楽の記録・再生においても，録音芸術が演奏芸術とはある種独立して発展してきたとも言える．その背景には，音響信号がその時間構造において，言語コミュニケーションに十分な情報量を保っていることが理由としてあろう．

しかしながら，人間が空間的な音を知覚する能力を有することは言うまでもない[9]．より現実感の高い音の提示を実現するためには，本来の空間も含めた四次元の音情報を適切な形で記録・伝送，再生することが不可欠となる．これにより，例えばSFの世界で描かれているような高い現実感をもたらす遠隔コミュニケーションシステムや，あたかも同じ空間を共有しているかのような高

小山翔一　正員　東京大学大学院情報理工学系研究科システム情報学専攻
E-mail koyama.shoichi@ieee.org
Shoichi KOYAMA, Member (Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, 113-8656 Japan).
電子情報通信学会誌　Vol.100 No.6 pp.474-478 2017年6月

い臨場感の再生システムなどが実現可能となる．このような技術によって，物理的に移動することなく様々な体験ができるようになるほか，新たな文化を形成する基盤ともなり得るだろう．

3. これまでの空間音響再生技術

空間的な音の提示技術としては，いわゆるステレオ方式が一般的であろう．広く普及している2チャネルステレオだけでなく，5.1チャネルサラウンド[10]や，NHK放送技術研究所による22.2 chマルチチャネル音響[11]などもこれに含まれる．Summing localization[9]と呼ばれる聴覚特性を利用し，各スピーカの振幅差や位相差を制御することで，空間的な音像をデザインし，擬似的に受聴者に提示する[12]．しかしながら，そのような効果が得られるのはスピーカの中心位置，いわゆるスィートスポットに限られる．

両耳位置の信号を模擬することを目的とした，バイノーラル再生技術も古くから研究がなされている[13]．一般的には，人間の頭部を模擬したダミーヘッドによって両耳位置の音を収録するか，音源から両耳までの伝達特性である頭部伝達関数（HRTF：Head-Related Transfer Function）をあらかじめ測定し，これを音源信号に畳み込むことで，バイノーラル信号を得る．再生にはヘッドホンを用いるか，複数のスピーカを用いて受聴者の両耳位置の音圧を制御する方法が用いられる．受聴者の頭部運動に追従してバイノーラル信号を厳密に得ることが困難であるほか，HRTFの個人差が定位感に影響してしまうことが知られている．

4. 高臨場感音響再生のための音場再現技術

筆者らがこれまで取り組んできた音場再現技術は，多数のマイクロホン・スピーカを用いて，音空間を物理的に忠実に再構成することを目的とする（図1）．これにより，広い受聴領域での音空間の提示が可能になる．また，音をデザインする過程を経ることが難しい，遠隔コミュニケーションやライブ配信，ディジタルアーカイブといった応用にも適用できる．両耳位置での信号を模擬するバイノーラル技術と異なり，空間的に音を再現するため，受聴者が自由に動いたとしてもそれに応じた音がそのまま提示できるほか，HRTFの個人差なども影響しない．

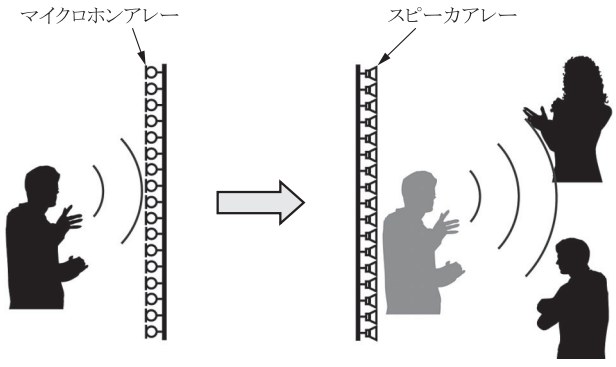


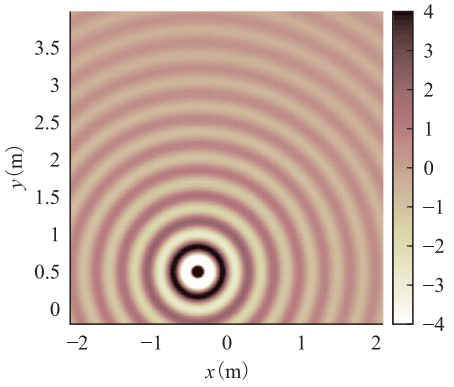
図1　音場再現の概念

このような音場再現の概念自体は古くから存在しており，例えばHuygensの原理に基づいて制御対象空間の境界面上で取得した信号をスピーカによって再合成する方法[14]が1967年に，Kirchhoff-Helmholtz積分方程式または第二種Rayleigh積分と呼ばれる境界面上からの音の伝搬を記述する物理式に基づく波面合成法（WFS：Wave Field Synthesis）[15]が1997年に提案されている．ところが，当時はシステムとして実現することが困難だったことも要因であろうが，最近までこのような技術の理論的な整理はなされてこなかった．現在ではセンサ・トランスデューサの小形化・低コスト化，A-D・D-A変換器の多チャネル化，計算機の高性能化などによってシステムとしての実現可能性が飛躍的に向上し，音響信号処理の一つの分野として国内外で盛んに研究され始めている[16]．

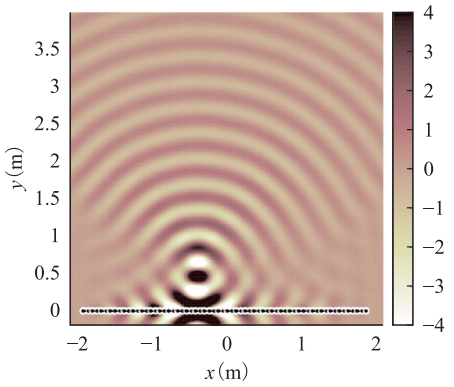
Berkhoutらによって最初に提案されたWFSは，2008年にSporsらによって再考された[17]．先にも述べたKirchhoff-Helmholtz積分方程式は，制御対象領域の境界面上にモノポール及びダイポール特性を持つスピーカを連続的に配置し，それらを所望音場における音圧及び音圧勾配で駆動することで，領域内の音場を再構成するというものであった．Sporsらはこれに対し，境界面を平面状とした場合に第一種Rayleigh積分と呼ばれる式に変形できることを利用し，平面状に配置したモノポール特性のスピーカを所望音場の音圧勾配で駆動することで，その半空間の音場を再現可能であることを示した．

高次アンビソニックス（HOA：Higher-Order Ambisonics）と呼ばれる手法は，音場を球面調和関数領域で表現することに基づく手法である[18]～[21]．1974年にGerzonによって，ある受聴点近傍の音場を再現するアンビソニックスと呼ばれる手法が提案されており[22]，HOAはその広範囲の再現への拡張としてみなせる．HOAでは多くの場合，球状のマイクロホンアレーとスピーカアレーを用いて，球面調和関数領域における音場の分析と合成を行う．それぞれエンコーディング／デコーディングと呼ばれている．

筆者らは，通常のマイクロホンで収録した音場を，スピーカアレーを用いて再構成するため，波面再構成（WFR：Wave Field Reconstruction）フィルタによる信号変換手法を提案している[23]～[25]．ここで注意したいのは，単純に対象領域の境界面上で音を収録し，そのまま

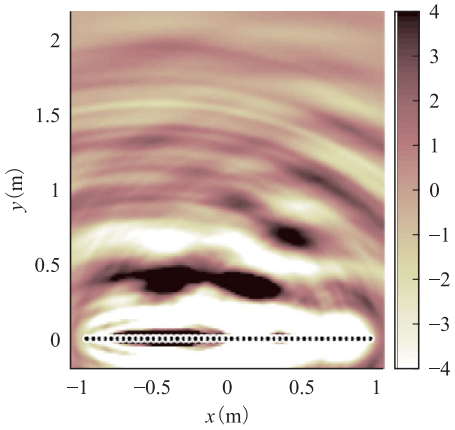


(a) 原音場

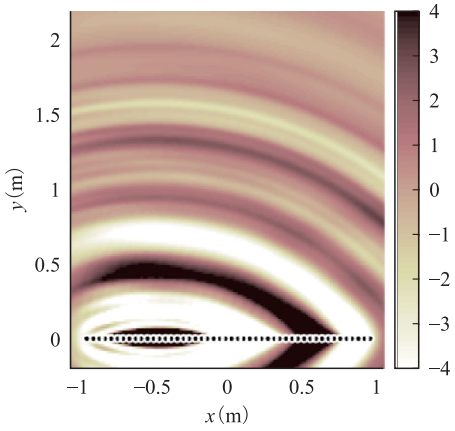


(b) 再現音場

図 2　直線状スピーカアレーによる音場再現　黒い丸印がスピーカ位置を示す．



(a) 従来法[25]



(b) 文献(33)の手法

図 3　音場再現における超解像化により，空間エイリアシングの誤差を抑制した実験例

スピーカによって再生しても，音場を正しく再現することはできないということである．例えば WFS では，所望音場における音圧勾配を取得し，スピーカアレーの駆動信号とする必要があったが，一般的なマイクロホンで取得できるのは音圧のみである．そこで，マイクロホンアレーで取得した音圧分布を空間周波数領域に変換し，解析的な表現に基づいて導出される WFR フィルタを適用することで，スピーカの駆動信号を得る[23]．平面状アレーだけでなく，様々なアレー形状に対してもこの手法を適用できることを示している[24],[25]．また，通常は受聴者に対してスピーカアレーの奥側に仮想的な音源が再現されるが，これをスピーカアレーの前側に再構成することも可能である[26]．図 2 に，本手法を用いて，スピーカアレー前側（ここでは $y>0$ の位置）に点音源を再現したシミュレーション実験の例を示す．また，本手法を用いて，遠隔地へリアルタイムに音場を伝送するシステムの構築も行っている[27]．

5. 未来の音の収録・再生・編集技術に向けて

音場再現技術では，高い周波数帯域まで広い領域を再現するためには，多数のマイクロホンとスピーカが必要になるという欠点がある．高周波数帯域では空間エイリアシングと呼ばれる誤差が生じ，音像の定位に対する影響はそれほど大きくない場合が多いものの[27]，再現された音源信号の音色が大きく劣化してしまう[28]．この問題に対して期待したいのは，更なるデバイスの進化である．最近では MEMS マイクロホンを用いて極めて多数のチャネル数を有するシステムが実現されている[29],[30]．スピーカに関しても同様に小形のデバイスが開発されつつあるが，周波数特性の広帯域化や高音質化などが課題と言えるだろう[31]．このようなデバイスがより簡便かつ安価に利用できるようになれば，空間中に分散配置したマイクロホンやスピーカを用いて音場の記録や再現を行うことも可能になると考えられ，応用範囲も更に広がるだろう．

また，筆者らは信号処理によって空間ナイキスト周波数以上の再現精度を向上する，言わば音場再現における超解像化技術の検討を行っている[32]～[34]．例えばマイクロホン数を減らすために，圧縮センシングの文脈で発展してきたスパース信号表現[35]に基づく手法を提案して

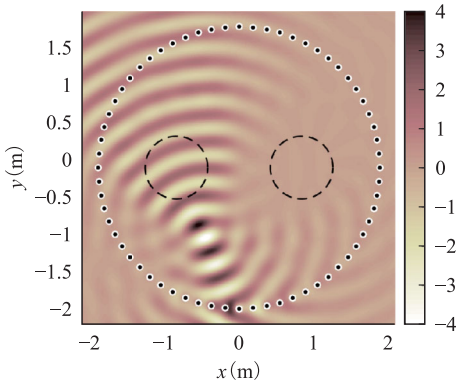


図 4　マルチゾーン音場再現　左側の円形領域に点音源による音場，右側の円形領域に無音の領域を合成した場合．

いる[33]．スピーカ数を減らすためには，受聴エリアの事前情報を利用することが有用と考えている[34]．図 3 は，文献(33)の手法を用いて，マイクロホン数が少数の場合に，音場の超解像化を行った実験例であり，従来法に比べて空間エイリアシングの誤差が抑制されているのが分かる．

ある単一の音場を再現するだけでなく，対象領域内に複数の異なる音場を再現する，マルチゾーン再現の技術も実現されつつある[36]．ある領域では日本語で，別の領域では英語で再生するというような，多言語の再生などにも応用可能な技術である．図 4 は文献(34)の手法を用いたマルチゾーン音場再現のシミュレーション実験例である．

特に音楽メディアにおいては，既存フォーマットである 2 チャネルステレオやモノラルの信号に対して，音の空間性を付与して再生する，いわゆるアップミキシングのような技術も必要となるだろう．複数の音源信号が重畳された信号を各音源に分離する音源分離の技術は，非負値行列因子分解，深層ニューラルネットワークなどの機械学習の手法を利用し，ここ数年で大きく発展を遂げてきた[37),(38]．音の空間的な情報のほとんどは失われているものの，ユーザが自由に空間性を編集・デザインすることができると捉えることもできる．

実環境の音場を収録して再現するのではなく，計算機上で模擬した仮想的な音場を合成することも可能であるが，現在は非常に単純なモデル化によるものに限られている．計算機性能が更に向上すれば，より複雑な物理現象を模擬することや，現実では起こり得ないような現象を合成することも，実時間で実現できるようになるだろう．先に述べた音源分離技術などと組み合わせることで，空間的に非常に自由度の高い音の編集や，音の拡張現実感システムの実現が期待できる．

現状，音場再現技術は物理音響や信号処理の考え方を基本にしているものの，今後は聴覚やマルチモーダルのような研究分野における知見を取り入れて行けるものと期待している．2. で述べたように，これまでは音の空間的な要素をむやみに破棄していた．聴覚やマルチモーダルの観点から臨場感に必要な要素を明らかにすることができれば，本当の意味での音情報のデータ圧縮が可能になると考える．

6. おわりに

冒頭で述べたように，音の臨場感は意識的に得られるものではなく，無意識の中にある情報から得られるものと考えられる．音響現象を音源の信号と位置のようなパラメータによって表現してしまうと，このような情報は失われてしまう．音場再現技術では，このようなパラメータで表現された音場だけでなく，複雑な音響現象も含めて再現できるということに，大きな利点がある．5. で述べたように，ここからいかにデータを圧縮するかは，別の観点から考える必要があるだろう．100 年後の未来に，本稿で取り上げた新たな音の収録・再生・編集技術が，新しい文化を形成していることを期待したい．

文　献

（1） A.W. Bronkhorst, “The cocktail party phenomenon : A review of research on speech intelligibility in multiple-talker conditions,” Acta Acust. United With Acust., vol. 86, no. 1, pp. 117-128, 2000.
（2） 柏野牧夫，“聴覚：環境に適応する無意識の知性(小特集-感性の領域に迫る音処理技術)，”音響誌，vol. 54, no. 7, pp. 508-514, 1998.
（3） D.H. Mershon and L.E. King, “Intensity and reverberation as factors in the auditory perception of egocentric distance,” Perception & Psychology, vol. 18, no. 6, pp. 409-415, 1975.
（4） T. Miura, T. Maruoka, and T. Ifukube, “Comparison of obstacle sense ability between the blind and the sighted : A basic psychophysical study for designs of acoustic assistive devices,” Acoust. Sci. Technol., vol. 31, no. 2, pp. 137-147, 2010.
（5） ヴァルター・ベンヤミン，複製技術時代の芸術，晶文社，1999.
（6） 舘　暲，佐藤　誠，廣瀬通孝，バーチャルリアリティ学，コロナ社，2010.
（7） B.A. Graham, “Improvement in telegraphy,” US Patent 174,465, March 1876.
（8） T.A. Edison, “Improvement in telephones or speaking-telegraphs,” US Patent 203,018, April 1878.
（9） J. Blauert, Spatial Hearing : The Psychophysics of Human Sound Localization, The MIT Press, 1996.
（10） ITU-R, “Multichannel stereophonic sound system with and without accompanying picture,” ITU-R Recommend. BS-775-1, 1994.
（11） K. Hamasaki, T. Nishiguchi, R. Okumura, Y. Nakayama, and A. Ando, “A 22.2 multichannel sound system for ultrahigh-definition TV (UHDTV),” SMPTE Motion Imaging J., vol. 117, no. 3, pp. 40-49, 2008.
（12） F. Rumsey, Spatial Audio, Oxford : Focal Press, 2001.
（13） 平原達也，大谷　真，戸嶋巌樹，“頭部伝達関数の計測とバイノーラル再生にかかわる諸問題，”信学 FR 誌，vol. 2, no. 4, pp. 68-85, April 2009.
（14） M. Camras, “Approach to recreating a sound field,” J. Acoust. Soc. Am., vol. 43, no. 6, pp. 1425-1431, 1967.
（15） A.J. Berkhout, D. de Vries, and P. Vogel, “Acoustic control by wave field synthesis,” J. Acoust. Soc. Am., vol. 93, no. 5, pp. 2764-2778, 1993.
（16） T. Betlehem, W. Zhang, M.A. Poletti, and T.D. Abhayapala, “Personal sound zones : Delivering interface-free audio to multiple listeners,”

IEEE Signal Process. Mag., vol. 32, no. 2, pp. 81-91, 2015.
(17) S. Spors, R. Rabenstein, and J. Ahrens, "The theory of wave field synthesis revisited," Proc. 124th AES Conv., no. 7358, Amsterdam, Oct. 2008.
(18) J. Daniel, "Spatial sound encoding including near field effect : Introducing distance coding filters and a viable, new ambisonics format," Proc. 23rd AES Int. Conf., no. 16, Copenhagen, May 2003.
(19) M. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," J. Audio Eng. Soc., vol. 53, no. 11, pp. 1004-1025, 2005.
(20) J. Ahrens and S. Spors, "An analytical approach to sound field reproduction using circular and spherical loudspeaker distributions," Acta Acust. United With Acust, vol. 94, no. 6, pp. 988-999, 2008.
(21) Y.J. Wu and T.D. Abhayapala, "Theory and design of soundfield reproduction using continuous loudspeaker concept," IEEE Trans. Audio, Speech, Lang. Process., vol. 17, no. 1, pp. 107-116, 2009.
(22) M.A. Gerzon, "Periphony : With-height sound field reproduction," J. Audio Eng. Soc., vol. 21, no. 1, pp. 2-10, Jan. 1973.
(23) S. Koyama, K. Furuya, Y. Hiwasaki, and Y. Haneda, "Analytical approach to wave field reconstruction filtering in spatio-temporal frequency domain," IEEE Trans. Audio, Speech, Lang. Process., vol. 21, no. 4, pp. 685-696, 2013.
(24) S. Koyama, K. Furuya, Y. Hiwasaki, Y. Haneda, and Y. Suzuki, "Wave field reconstruction filtering in cylindrical harmonic domain for with-height recording and reproduction," IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 22, no. 10, pp. 1546-1557, 2014.
(25) S. Koyama, K. Furuya, K. Wakayama, S. Shimauchi, and H. Saruwatari, "Analytical approach to transforming filter design for sound field recording and reproduction using circular arrays with a spherical baffle," J. Acoust. Soc. Am., vol. 139, no. 3, pp. 1024-1036, 2016.
(26) S. Koyama, K. Furuya, Y. Hiwasaki, and Y. Haneda, "Reproducing virtual sound sources in front of a loudspeaker array using inverse wave propagator," IEEE Trans. Audio, Speech, Lang. Process., vol. 20, no. 6, pp. 1746-1758, 2012.
(27) S. Koyama, K. Furuya, H. Uematsu, Y. Hiwasaki, and Y. Haneda, "Real-time sound field transmission system by using wave field reconstruction filter and its evaluation," IEICE Trans. Fundamentals, vol. E97-A, no. 9, pp. 1840-1848, Sept. 2014.
(28) D. de Vries, Wave Field Synthesis, AES Monographs, Audio Eng. Soc., 2009.
(29) I. Hafizovic, C. -I.C. Nilsen, M. Kj. lerbakken, and V. Jahr, "Design and implementation of a MEMS microphone array system for real-time speech acquisition," Appl. Acoust., vol. 73, no. 2, pp. 132-143, 2012.
(30) 坂本修一，松永純平，本郷　哲，岡本拓磨，岩谷幸雄，鈴木陽一，"252 ch リアルタイム音空間情報収音再生システム SENZI の音空間再現精度改善手法の検討，" 信学技報，EA 2012-55, pp. 7-12, Aug. 2012.
(31) F. Kontomichos, A. Koutsioubas, J. Mourjopoulos, N. Spiliopoulos, A. Vradis, and S. Vassilantonopoulos, "Testing and simulating of a thermoacoustic transducer prototype," Proc. 126th AES Conv., no. 7679, Munich, May 2009.
(32) S. Koyama, K. Furuya, Y. Haneda, and H. Saruwatari, "Source-location-informed sound field recording and reproduction," IEEE J. Sel. Topics Signal Process., vol. 9, no. 5, pp. 881-894, 2015.
(33) S. Koyama, S. Shimauchi, and H. Ohmuro, "Sparse sound field representation in recording and reproduction for reducing spatial aliasing artifacts," Proc. IEEE Int. Conf. Acoust., Speech, Signal Process (ICASSP), pp. 4443-4447, Florence, May 2014.
(34) N. Ueno, S. Koyama, and H. Saruwatari, "Listening-area-informed sound field reproduction based on circular harmonic expansion," Proc. IEEE Int. Conf. Acoust., Speech, Signal Process (ICASSP), pp. 111-115, March 2017.
(35) D.L. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1289-1306, 2006.
(36) Y.J. Wu and T.D. Abhayapala, "Spatial multizone soundfield reproduction : Theory and design," IEEE Trans. Audio, Speech, Lang. Process., vol. 19, no. 6, pp. 1711-1720, 2010.
(37) P. Smaragdis and J.C. Brown, "Non-negative matrix factorization for polyphonic music transcription," Proc. IEEE Int. Workshop Appl. Signal Process. Audio Acoust (WASPAA), pp. 177-180, New Paltz, Oct. 2003.
(38) P.-S. Huang, M. Kim, M. Hasegawa-Johnson, and P. Smaragdis, "Deep learning for monaural speech separation," Proc. IEEE Int. Conf. Acoust., Speech, Signal Process (ICASSP), pp. 1562-1566, Florence, May 2014.

（平成 29 年 1 月 3 日受付　平成 29 年 2 月 1 日最終受付）

小山　翔一（こやま　しょういち）（正員）

2007 東大・工・計数卒．2009 同大学院情報理工学系研究科修士課程了．同年，日本電信電話株式会社入社．以来，音響信号処理の研究に従事．2014 東大大学院情報理工学系研究科助教．現在，Paris Diderot University（パリ第 7 大学）客員研究員兼任．博士（情報理工学）．2015 日本音響学会独創研究奨励賞板倉記念受賞など．