Vol.2018-MUS-119 No.34
Vol.2018-SLP-122 No.34
2018/6/16

クラウドソーシング音像定位実験における参加者信頼度と実験デザインの検討

森川大輔[†1]　高道慎之介[†2]

概要：本報告では，バイノーラル再生での音像定位実験をクラウドソーシングで行った．そして，参加者の信頼度指標として，知覚方位角が明らかに誤りと考えられる回答の割合，使用したイヤホン，参加者自身の回答に対する自信，実験時の受聴音圧レベルを用いることができないか検討した．その結果，知覚方位角が明らかに誤りと考えられる回答の割合と，参加者自身の回答に対する自信が信頼度の向上に効果があることがわかった．

キーワード：クラウドソーシング，音像定位，参加者信頼度，バイノーラル

Study of crowd workers' credibility and experimental design on sound localization experiment by crowdsourcing

DAISUKE MORIKAWA[†1]　SHINNOSUKE TAKAMICHI[†2]

Keywords: Crowdsourcing, Sound localization, Crowd workers' credibility, Binaural

1. はじめに

音像定位実験は，ヒトの聴知覚特性を調べたり，立体音再生システムを評価したりするため，古くからよく行われている[1]．信頼できる実験結果を得るためには，多数の受聴者を集め，大規模な実験を行う必要になるケースも多い．しかし，人的・時間的コストの観点から小規模な実験に留まる場合が多く見られるのも現状である．一方で，クラウドソーシングプラットフォームの発達により，一部の分野においては大規模実験をクラウドソーシングで行うことが可能になってきた[2-4]．ただし，クラウドソーシングの利用は，当然各家庭にある機器で実行可能な実験に限られ，さらに機器の特性による影響を比較的受けにくい分野に限定されてしまっている．

音像定位実験を行う場合，音源を多数配置するスピーカアレイを用いる方法と，ヘッドホンによるバイノーラル再生を用いる方法がある．スピーカアレイを各家庭で揃えることは困難であるが，バイノーラル再生はヘッドホンで受聴可能な環境があれば実行自体は可能になる．ただし，バイノーラル再生による受聴の場合，ヘッドホン等の特性によって評価結果が変動してしまうため[1]，必ずしも適切な結果を得られないことが予想される．このような理想環境よりも精度が低くなってしまう問題は，クラウドソーシング評価において共通し，ヒューマンコンピュテーション分野等では，各参加者の結果をどの程度信用できるかを示す信頼度指標が提案されている[5]．

†1 富山県立大学
Toyama Prefectural University
†2 東京大学
The University of Tokyo

そこで本報告では，クラウドソーシングによるバイノーラル再生での音像定位実験を実際に行い，どの程度の精度が得られるかを調査した結果を示すとともに，参加者信頼度の推定による信頼性の向上について議論する．

2. 実験方法

2.1 刺激音

刺激音には，持続時間を 3 s の白色雑音に頭部伝達関数(Head-related transfer function: HRTF)を畳み込んで作成した合成バイノーラル音を用いた．畳み込んだ HRTF は，ヘッドアンドトルソシミュレータ（4128-C, Brüel & Kjær）を半径 1.5 m，30°間隔で計測した物である．また，刺激音の最初と最後には 30 ms の線形テーパをかけた．白色雑音と HRTF のサンプリング周波数は 48 kHz とし，WAV 形式で量子化精度 16 bit で保存した．

2.2 システム

刺激音の再生には，JavaScript の Audio オブジェクトを用い，Web ブラウザから各参加者の再生デバイスを駆動した．したがって，受聴環境は任意である．図 1 に Web ブラウザに表示される実験用 GUI (Graphical User Interface)を示す．GUI は頭部のイラスト，試験開始・終了ボタン，回答ボタンで構成されている．参加者は「start」 ボタンを押して実験を開始し，刺激音の受聴後に知覚した音像の方向と，頭外定位であるか頭内定位であるかを該当するボタンを押すことで回答し，すべての回答終了後に「submit」ボタンを押して実験を終了した．なお，次の刺激音の呈示タイミングは回答ボタンが押された後とし，回答に制限時間は設けていない．また，聴き直しと回答の修正は認めないものとした．

Vol.2018-MUS-119 No.34
Vol.2018-SLP-122 No.34
2018/6/16

2.3 実験条件

練習試行として，頭部中心，0, 90, 180, 270° の刺激音を呈示した．ただし，練習中は頭部中心を除き呈示方向は参加者に示していない．なお，頭部中心の刺激音は，HRTF の畳み込みを行っていない同じ白色雑音を，左右の耳から Diotic 再生したものである．

本実験では，30° 間隔 12 方向の刺激音を 2 回ずつ，計 24 回の刺激音を参加者ごとにランダムな順で呈示した．

2.4 参加者

参加者は，クラウドソーシングサービス「ランサーズ」（http://www.lancers.jp）上で募集した．

3. 信頼度指標

信頼度を推定する手法として，真値が既知のデータ(gold standard data) を試験に含ませ，その正答率を信頼度とする手法や別実験結果から判断する手法[6]，参加者本人に作業に関する自信を判断させる手法[7]などがある．本報告では，これらの手法を参考とし，知覚方位角による信頼度(CoP :Creditability of Perceived azimuth)，イヤホンの信頼度(CoE: Creditability of Earphones)，参加者自身の確信度による信頼度(CoC :Creditability of Confidence)，感覚レベルの信頼度(CoS: Creditability of Sensation level)を用いた．

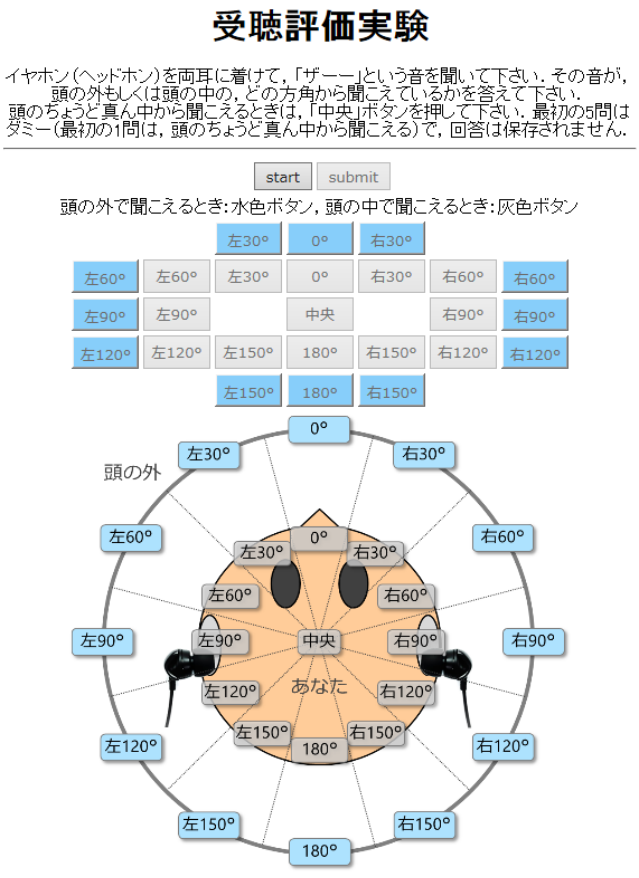

図 1 実験用 GUI

Figure 1 Experimental GUI.

3.1 知覚方位角による信頼度 CoP

音像定位実験の場合，知覚した方位に真値は存在しない．そこで本報告では，明らかに誤りと考えられる回答数を利用数する．水平面音源のバイノーラル再生においては，再現精度が低い場合であっても，左右を逆に判断することは少ない．したがって，90° または 270° の音源を，対側の 240～300° または 30～120° と回答した割合を 1 から引いたものを CoP とした．

3.2 イヤホンの信頼度 CoE

参加者が使用したイヤホンの品質及び伝達関数は，定位結果を大きく変動させる．そこで本報告では，定位実験と別に使用イヤホンに関する記述式の設問を用意した．それに対し，"Sony MDR-1000"などのアルファベット・記号・数字のみを用いた回答の CoE を 1.0，"iPhone4s についていたもの"などの，日本語を用いた回答の CoE を 0.0 とした．ただし，日本語表記のメーカ名はアルファベットに変換し，助詞を削除したものを使用した．

3.3 確信度による信頼度 CoC

確信度は本報告においても正しく方向を答えられた自信があるかで判断することが可能である．そこで本報告では，定位実験後に確信度に関する問を設け，参加者から正しく方向を答えられた自信があるかを 7 段階評価で得，これを CoC とした．

3.4 感覚レベルによる信頼度 CoS

呈示音圧は，定位結果を変動させると予想されるが，クラウドソーシングでの実験では，参加者に呈示する音圧の統制を取ることは困難である．どの振幅の刺激音が受聴できたかを把握すれば，おおよその感覚レベルを知ることはできる．そこで本報告では，定位実験の前に 0.5 s の白色雑音を，音圧を 6 dB ずつ上げて 7 回，0.5 s 間隔で呈示し，参加者から聴こえた回数の回答を 1～7 の 7 段階で得，これを CoS とした．最大音圧の白色雑音は，音源が正面にある刺激音と同じ音圧レベルである．

4. 実験 1:知覚方位角とイヤホンによる評価

CoP と CoE を評価した実験について述べる[8]. 本実験の参加者は 100 名で，各参加者には 85 円を支払っている．

実験終了後，参加者から「あなたの使っているイヤホン・ヘッドホンの製品名を教えて下さい.」という設問に対する自由形式の回答を得た．

4.1 実験結果

図 2 に 100 名から得た総計 2400 の定位結果をまとめたものを示す．横軸は刺激音の呈示角度，つまり，畳み込んだ HRTF の計測角度で，縦軸が参加者の回答角度もしくは頭部中央を示している．なお，角度は正面を 0° とした時計回りである．円の面積は回答回数に比例し，赤が音像を頭外に知覚した結果，青が音像を頭内に知覚した結果を示している．右上がりの対角線は刺激音の呈示角度と参加者

Vol.2018-MUS-119 No.34
Vol.2018-SLP-122 No.34
2018/6/16

の回答角度が一致した場合を示し，右下がりの線は前後を誤って知覚した場合を示している．

呈示角度と回答角度が一致した割合（定位正答率）は 22.6%で，呈示角度と回答角度が一致し，頭外に定位した割合（頭外定位正答率）は 13.5%であった．また，±30°を正答と許容した場合ではそれぞれ 52.8%, 32.6%であった．また，頭外に定位した割合は 55.8%であり頭内定位が多くみられた．これらの結果は，これまでに理想環境において監督下で行われた実験に比べ精度が低い．

4.2 知覚方位角による信頼度 CoP

CoP が 1.0 の参加者は 90 名，0.75，0.5，0.25 の参加者はそれぞれ 2 名，0.0 の参加者は 4 名であった．CoP が 1 の参加者の正答率の平均は 24.5%，±30°を許容した場合では 57.4%で，全体の平均より高くなった．一方，CoP が 0.75 以下の参加者の正答率は最大でもそれぞれ 12.5%, 29.2%で，±30°を許容した場合でも正答がない参加者が 3 名含まれていた．CoP が 0.75 以下の参加者はイヤホンを左右逆に装着している可能性があり，このような参加者を除外するのに CoP は効果的と考えらえる．

4.3 イヤホンの信頼度 CoE

CoE が 1.0 の参加者は 75 名，0.0 の参加者は 25 名で，定位正答率はそれぞれ 22.5%と 22.9%で，定位正答率に大きな変化は見られなかった．

5. 実験 2:確信度と感覚レベルによる評価

CoC と CoS を評価した実験について述べる．本実験の参加者は 145 名で，各参加者には 95 円を支払っている．

実験開始前に，CoS 確認用の刺激音を呈示し，参加者から聴こえた回数の回答を得た．

また，実験終了後，「正しい方角を答えられた自信度を"1 (全く自信がない)" から "7 (かなり自信ある)" で答えて下さい．」という設問に対する 7 段階評価の回答を得た．

なお，実験が難解にならないよう，頭内定位と頭外定位の区別をなくし，頭部中央の回答を削除して実験を行った．

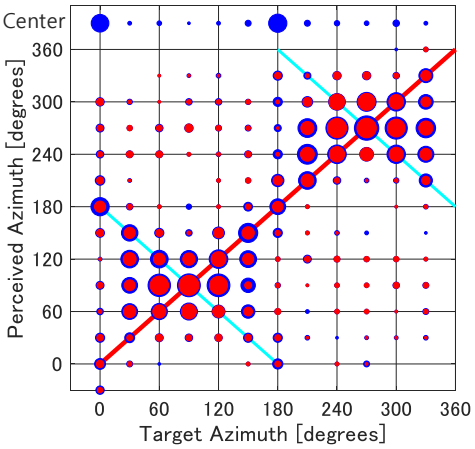

図 2 実験 1 の音像定位結果

Figure 2 Sound localization result of experiment 1.

5.1 実験結果

図 3 に 145 名から得た総計 3480 の定位結果をまとめたものを示す．定位正答率は 28.3%で，±30°を正答と許容した場合は 61.2%であった．

5.2 確信度による信頼度 CoC

CoC が 1～7 の参加者はそれぞれ 4, 36, 49, 22, 25, 7, 2 名であった．図 4 に CoC ごとの定位正答率の平均値と標準偏差を示す．青が定位正答率，赤が±30°を正答と許容した定位正答率である．

定位正答率は CoC が 5 で最も高く，6 では定位正答率が下がった．一方，±30°を許容した場合，あまり CoC に差はなかった．この結果から，定位正答率を測る場合には CoC が 3～5 の参加者を抽出した方が安定した結果を得やすいことが予想される．±30°を許容する場合，CoC はあまり効果が得られないが，CoC が 1 の参加者は標準偏差が大きいことから，除外した方が良い可能性もある．

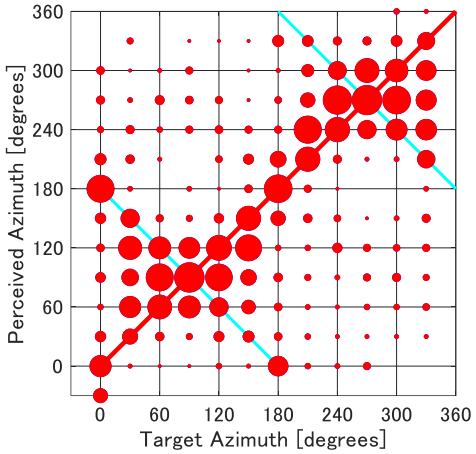

図 3 実験 2 の音像定位結果

Figure 3 Sound localization result of experiment 2.

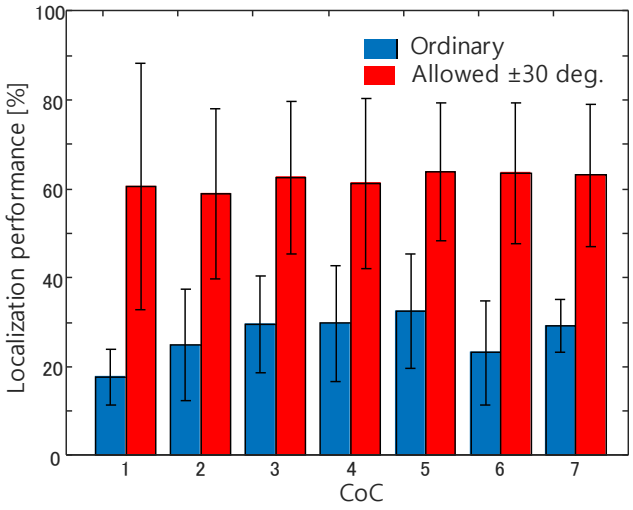

図 4 確信度による定位正答率

Figure 4 localization performance of each CoC.

Vol.2018-MUS-119 No.34
Vol.2018-SLP-122 No.34
2018/6/16

5.3 感覚レベルによる信頼度 CoS

聴こえた回数が 7 回の参加者が 127 名，6 回が 14 名，4 回が 3 名，3 回が 1 名であった．一番大きい刺激音と一番小さい刺激音で 42 dB 異なることから，ほとんどの参加者が感覚レベル 42 dB 以上の条件で実験を行っていたことがわかる．聴こえた回数が 3 回の参加者の定位正答率は 29.2%で，±30° を正答と許容した場合は 62.5%，4 回の参加者の平均はそれぞれ 34.7%と 75%，6 回の参加者の平均はそれぞれ 22.9%と 51.5%で，いずれも全体の平均とほぼ同じであった．

6. 防音室における実験との比較

クラウドソーシングによる実験の結果と，理想環境において監督下で行った実験の結果を比較するために，防音室において実験を行った．

6.1 システムの変更点

PC（Windows 8）上で読み込んだ刺激音を D/A 変換器 (RME, Fireface UCX) から出力し，ヘッドホンアンプ (audio-technica, AT-HA21) を通してヘッドホン(Sennheiser, HDA-200) を駆動した．刺激音の音圧レベルは，音源が正面にある場合に 70 dB とした．また，実験を行った防音室の暗騒音レベルは 21 dB 以下である．

刺激音の呈示間隔は 3 s とし，参加者はこの 3 s の間に回答用紙に刺激音の呈示番号を記入した．

6.2 実験条件の変更点

30° 間隔 12 方向の刺激音を 5 回ずつ，60 回を 1 セッションとし，4 セッションで計 240 回の刺激音を参加者ごとにランダムな順序で呈示した．ただし，クラウドソーシングの結果との比較のため，各参加者の 1 セッション目の最初の 24 回には，それぞれの方位角から 2 回ずつが選ばれる．

6.3 参加者

参加者は，20～30 歳台の健聴な受聴者 4 名である．

6.4 結果

図 5 に 6 名から得た総計 960 の定位結果をまとめたものを示す．定位正答率は 76.0%，頭外定位正答率は 65.8%で，±30° を許容した場合にはそれぞれ 93.8%と 82.9%であった．また，頭外に定位した割合は 88.4%であり頭内定位は音源方向が 0° の場合以外ほとんどみられなかった．1 セッション目の 24 サンプルだけの結果では，定位正答率は 79.2%，頭外定位正答率は 65.6%で，±30° を許容した場合にはそれぞれ 90.6%と 77.1%であった．また，真横の音源が対側に知覚されることや，合成バイノーラル音が頭部中央に知覚されることはなかった．

6.5 結果の比較

クラウドソーシングでの結果と比較すると，クラウドソーシングでの結果が防音室での結果を大きく下回り，提案した信頼度指標だけでは不十分であることがわかる．ここで，それぞれの実験において正答率の高い順に参加者を並べた場合の各参加者の定位正答率を図 6 に示す．青が実験 1，赤が実験 2，黒が防音室の結果で，破線は±30° を許容した場合の結果を示している．実験 2 に比べて実験 1 の正答率が全体的に低くなっているのは，実験 1 では頭部中央に判断していたものが，実験 2 ではいずれかの方向と回答されたためと考えられる．また，どちらのクラウドソーシングの実験であっても，正答率の高い参加者の結果は，防音室での結果に近接することが確認できる．したがって，信頼度指標によって，上位数名を選出できれば，バイノーラルでの音像定位実験はクラウドソーシングでも理想環境と同程度の精度が得られると考えられる．ただし，効果のみられた CoP と CoC であっても理想環境での精度が得られる参加者の抽出は困難である．現状で理想環境と同程度の精度でバイノーラルの定位実験をクラウドソーシングで行うためには，今回行った実験を予備テストとして行い，この結果を信頼度指標として用いる必要がある．

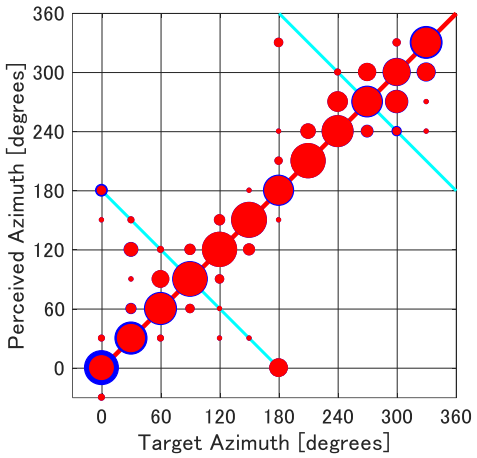

図 5　防音室での音像定位結果

Figure 5　Sound localization result in soundproof room.

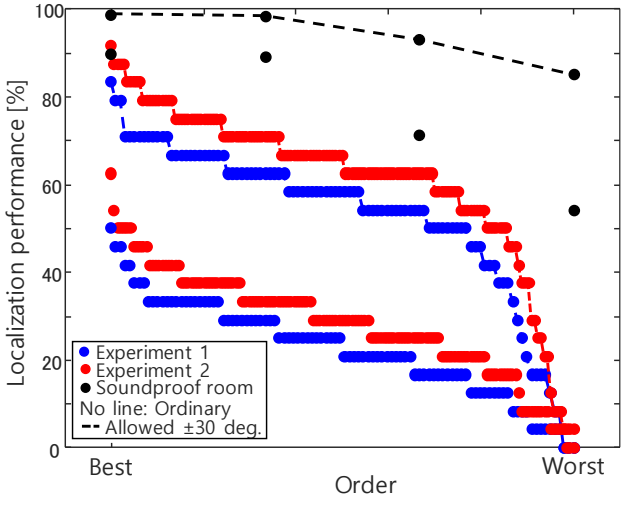

図 6　各参加者の音像定位結果

Figure 6　localization performance of each listener.

Vol.2018-MUS-119 No.34
Vol.2018-SLP-122 No.34
2018/6/16

7. まとめ

本研究では，バイノーラル再生の音像定位実験をクラウドソーシングで行い，評価結果の精度向上のために参加者信頼度指標を導入し，その有効性を確認した． その結果，知覚方位角が明らかに誤りと考えられる回答の割合と，参加者自身の回答に対する自信が信頼度の向上に効果があることがわかった．しかし，この指標だけでは十分な精度は得られなかった．また，使用したイヤホン，実験時の受聴音圧レベルは信頼度の向上に影響しなかった．

謝辞 本研究の一部は，セコム科学技術支援財団の助成を受け実施した．

参考文献

[1] 飯田一博, 頭部伝達関数の基礎と3次元音響システムへの応用, コロナ社, 2017.
[2] Jeanne Parson, Daniela Braga, Michael Tjalve, Jieun Oh, Evaluating Voice Quality and Speech Synthesis Using Crowdsourcing, *Proc. TSD*, pp. 233-240, Pilsen, Czech Republic, Sep. 2013.
[3] Cyrus Rashtchian, Peter Young, Micah Hodosh, Julia Hockenmaier, Collecting Image Annotations Using Amazon's Mechanical Turk, *Proc. NAACL HLTT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, pp. 139-147, California, U.S.A., Jun. 2010.
[4] Tomoyuki Kajiwara, Kazuhide Yamamoto, Evaluation Dataset and System for Japanese Lexical Simplification, *Proc. the ACL-IJCNLP 2015 Student Research Workshop*, pp. 35-40, Beijing, China, Jul. 2015.
[5] 鹿島久嗣，馬場雪乃，ヒューマンコンピュテーション概説，人工知能学会誌，vol. 29, no. 1, pp. 4-11, 2014.
[6] Gabriella Kazai, Jaap Kamps, Marijn Koolen, Natasa Milic-Frayling, Crowdsourcing for Book Search Evaluation: Impact of HIT Design on Comparative System Ranking, *Proc. ISGIR*, pp. 205-214, Beijing, China, Jul. 2011.
[7] Satoshi Oyama, Yukino Baba, Yuko Sakurai, Hisahi Kashima, Accurate Integration of Crowdsourced Labels Using Workers' Self-Reported Confidence Scores, *Proc. International Joint Conference on Artificial Intelligence*, pp. 2554-2560, Beijing, China, Aug. 2013.
[8] 高道慎之介，森川大輔，クラウドソーシング参加者の信頼度と音像定位精度の関係，日本音響学会秋季研究発表会 講演論文集，pp. 549-550, 2018.