

ニューラルネットワークを用いた生活音識別 Living sound identification using Neural Network

○佐々木駿¹, 藤岡紘展² 関弘翔³, 細野裕行³

*Shun Sasaki¹, Hironobu Fujioka², Hiroto Seki³, Hiroyuki Hosono³

Abstract: In this report, we study improvement of recognition accuracy for living sound identification. Feature extraction is performed with Fully Connected Layer, and identification is performed with Softmax Layer. As the result, recognition accuracy has been improved to 99%. In addition, it was confirmed that it was resistant to white noise up to 10% with an accuracy of 95%.

1. まえがき

聴覚障害者の数は厚生労働省が行った平成 18 年度の調査結果^[1]では、18 歳以上で 34 万人、18 歳未満で約 1 万 6 千人であり、およそ 1000 人に 3 人が聴覚障害者という結果であった。また 2018 年に WHO が、世界的には聴覚障害者の数は年々増加傾向にあり、2050 年には現在の推計約 4 億 7 千万人から 9 億人に達する可能性がある^[2]と発表している。

生活音識別の既存研究^[3]では、16 種類の生活音で識別を行っていた。信号持続曲線を用いて特徴抽出をし、マハラノビス距離を用いた識別により、全体の識別率は約 95%であったが、中には 90%に満たないものもあった。

そこで本研究では入力から全結合層で特徴を抽出し、Softmax 層で識別を行うニューラルネットワークを用いた手法を検討する。また、ニューラルネットワークによる識別のノイズ耐性についても検討する。

2. データセット

今回識別する生活音は普段の生活に必要なものから発信される音と、気づけないと危険性のある音を対象にした。対象とする音を Table 1 に示す。今回作成したデータセットには、No. 1~8 までの音が各 200 サンプル、その他の音が約 100 サンプルずつある。サンプリング周波数は 16kHz で、それぞれデータの始まりから 3 秒間のみ使用する(要素数 48000)。これらの音源データを各種類 4:1 の割合で Train データと Test データに分け、学習と検証を行う。

3. 生活音識別モデルの構築

ニューラルネットワークに入力するための前処理として、それぞれの音を[-1,1]で正規化し、フーリエ変換を行う。

本報告では、全結合層 2 層で特徴を抽出し、Softmax 層でクラス識別を行うニューラルネットワークのモデルを作成した。Fig. 1 にネットワークモデルを示す。

Table 1 . Target sounds

1.Intercom	10.Water flowing(1)
2.LINE phone	11.toaster
3.Fire alarm	12.Sliding door open
4.Microwave(1)	13.Sliding door close
5.Opening door Open	14.Water flowing(2)
6.Opening door close	15.Microwave(2)
7.Refrigerator open	16.Black phone
8.Refrigerator close	17.Phone(1)
9.Gas stove	18.Phone(2)

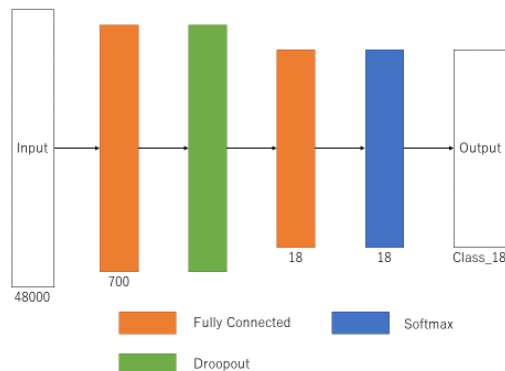


Figure 1. Network model

Table 2 . Dropout Result

Dropout rate	0%	10%	30%	50%	70%
Accuracy	99.42%	99.81%	99.62%	97.88%	98.65%

特徴抽出を行う全結合層のニューロン数は、GPU のスペック(メモリ 8GB)と精度の兼ね合いにより試験的に求めた値である 700 に設定し学習を行った。また、過学習抑制のために全結合層間に Dropout 層を挿入した。Dropout 層におけるドロップアウト率を複数試した結果を Table 2 に示す。試行結果より 10%の時の精度が良かったため、本報告では 10%を採用する。

4. 結果

学習した識別モデルによる識別精度を、Test データを用いて検証した。また、ノイズをのせた Test データ

1 : 日大理工・学部・情報 , 2 : 日大理工・院 (前)・情報, 3 : 日大理工・教員・情報

を用い、同一の識別モデルによる識別精度を確認し、ノイズ耐性に関する検証を行なった。

4. 1. 識別精度の検証

識別精度の検証結果を Table 3 に示す。開き戸を閉める音 (No.6) を開ける音 (No.5) へと 2.5%誤識別しているが、その他は概ね良好に識別できている。Fig.2 に閉める音と開く音のFFT後の波形を示す。この図より、識別誤りの原因は、2つの音が似た特性を持っていることである。

4. 2. ノイズ耐性の検証

Test データにホワイトノイズを加えたデータで検証を行った。ホワイトノイズは正規化した音源の 3%と 10%の 2 種類を用意した。Table 4 に 3%での検証結果を、Table 5 に 10%での検証結果を示す。Table 4 を見ると、ホワイトノイズ 3%時は耐性があり、全体の精度は 99%になった。Table 5 では開き戸の開ける音と閉める音、冷蔵庫の開ける音 (No.7) と閉める音 (No.8) の誤りが多くみられ、全体の精度は 95%であった。Fig.3 に冷蔵庫の閉める音と開く音の波形を示す。この図より、識別誤りの原因は、細かな特性がノイズに埋もれたことから、これらの音が似た特性となるためである。

5. まとめ

本研究では、ニューラルネットワークを用いた生活音識別とノイズ耐性について検討した。識別精度は、検証データ 18 クラス中 17 クラスが 100%で、誤識別したクラスでも 97.5%であった。ノイズ耐性の検討から、3%程度のノイズに対して耐性があることが明らかになった。今後はノイズ耐性強化のため、Train データにノイズを付加したものを増やした状態で検討を行い、さらに実際の生活環境下で録音された音源も増やしていきたい。

Table 3. Confusion Matrix

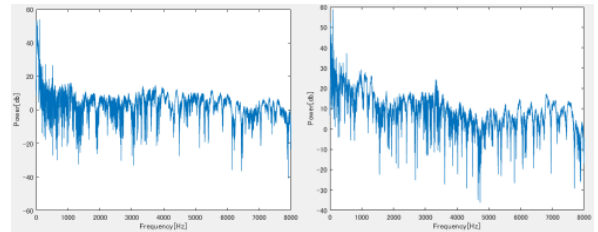
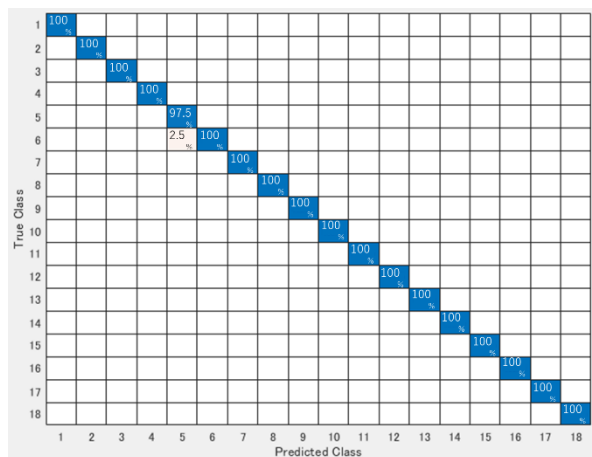


Figure 2. FFT wave (Left : No.6, Right : No.5)

Table 4 . Confusion Matrix(white noise 3%)

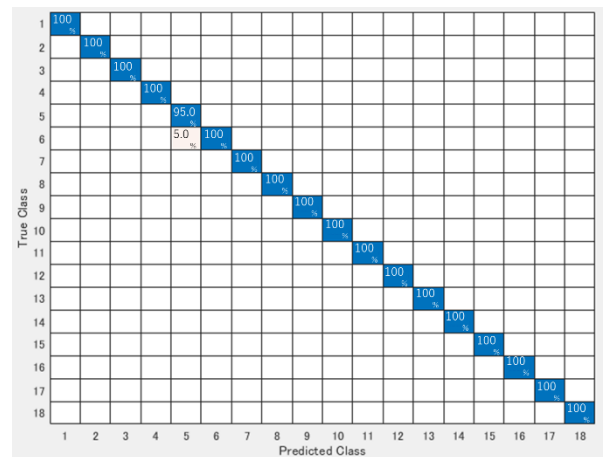


Table 5 . Confusion Matrix(white noise 10%)

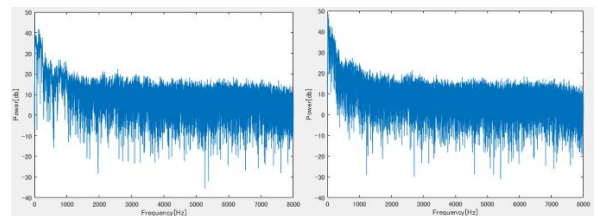
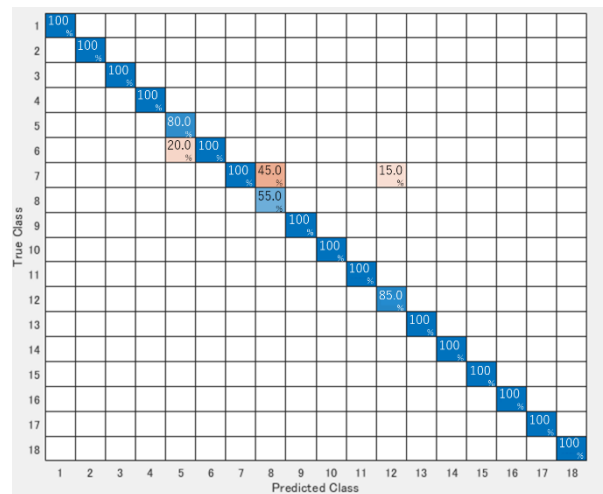


Figure 3. FFT wave2 (Left : No.8 Right : No.7)

6. 参考文献

[1] 厚生労働省社会・援護局障害保健福祉部企画課：「平成 18 年身体障害児・者実態調査結果」(2008)
 [2] World Health Organization: 「Deafness」, <https://www.who.int/news-room/facts-in-pictures/detail/deafness> (2019-09)
 [3] 稲井孝行: 「生活音の識別における信号持続曲線特徴量の比較」, 電子情報通信学会総合大会講演論文集 (2014)