社団法人　日本音響学会

<http://www.asj.gr.jp/qanda/answer/50.html>

CiNii　論文引用

　子供の音声は，声道長や基本周波数が成人音声と異なるだけでなく，発声スタイルが自発的で年齢や個人による声の変動も大きいため，通常の成人用音韻モデルでは認識精度が著しく劣化する．また，子供が正確に文章を読み上げることは多大な労力が必要であり，大規模な整った音声データベースの作成が難しい．本研究では，子供の自発的な発話の高精度な認識を目指して，音声情報案内システムによる子供の実音声の大規模収集，年齢層別子供用音韻モデルの構築と評価，および教師なし話者適応の検討を行った．大語彙連続音声認識実験より，実環境で収集した子供音声を用いることで，単語認識精度が71.1%と既存の読み上げ音声モデルに比べて絶対値で23.9%の改善が得られた．また，年齢層別の傾向では，特に幼児の音声において年齢層依存モデルによる大幅な精度改善が見られた．次に，自動収集した話者ラベルなしの大量データに対する，自動話者クラスタリングを用いた十分統計量に基づく教師なし話者適応を提案した．提案法により59 966個の発話データをクラスタリングし，近傍話者クラスタを用いて音韻モデルを適応することで，クラスタ数200の条件において，年齢層依存モデルに対してさらに幼児で2.2%，低学年子供で1.7%，高学年子供で0.5%の認識性能の改善が得られた

（1回目）月読アイの認識テストを終えて

<結果>

　・動物英単語の最も認識率の良かったパラメータは”速さ½”の43％で、英単語は”cow”の76％であった。

　・飲食物英単語は全体的に認識が悪く、最も良いパラメータは”音程変化”の20％で、英単語は”hamburger”の48％であった。

　・動物英単語と飲食物英単語の認識率に大きな差が生じた。環境、使用音源はすべて同じにした。

<原因>

　・iPodtouchに音声を入れその音声をマイク入力でjuliusに認識させたから？(余計なノイズが入っていた？)

　・単語辞書の中に似た音をもつ単語がはいっており、誤認識がおこったのではないか

　・マイクデバイスそもそもの集音効果が悪かったのではないか

　・月読アイの音声波形または周波数がjuliusが認識しにくいものだったのではないか

<改善>

　・日本語英語の入力ができる証明→つまり、自分の声である程度の認識率を確保すればjuliusの認識は正しい。そして、認識率の悪さは小学生の音声周波数に合わせた音響モデルを作成することで改善できる。

　・音が被らないように単語辞書を調整する。

　・マイクデバイスを良いものに変える。

＜単語別の誤認識の例＞

・salmon

hamburger,cow

・fish

　語尾を伸ばしているので、単語辞書の修正で改善できるかも

・

(2回目)

<認識率５０％未満の英単語>

　・動物英単語で最も良かったパラメータは最適化と速さ1/2の”43％”だった。最も良かった英単語は”cow”の96　　　％であった。

　・飲食物英単語で最も良かったパラメータは音程変化の”26％”だった。最も良かった英単語は"salmon"の60％　　であった。

　・1回目と比べてみると、動物英単語のパラメータのパーセンテージに変化はないが、英単語の認識率に大き　　な向上が見られた。そして1回目結果が芳しくなかった飲食物英単語は、同じく音程変化がパラメータ中最も　　よく、6%ほど改善がみられた。

動物英単語の読みを読みと一音目の子音と、使用されている母音を抜き出した。(ん、小文字、伸ばしは考えない）結果をいかに示す。

単語名　　　　認識率　　子音、母音　　　　　　誤認識回数と％

・cat（ｷｬｯﾄ）　　28　　　　C　　　a,o　　　　　　　17回　cow（29％）、wolf（41％）、water（5％）、lion（17％）

・cow（ｶｳ） 　　96　　　　C　　　a,u　　　　　　　1回

・giraffe(ｼﾞｭﾗﾌ） 60 G　　　u,a,　　　　　　10回　wolf（20％）、horse（50％）、lion（20％）

・bear（ﾍﾞｱ）　　　8　　　　 B　　　e,a　　　　　　　23回　wolf（52％）、horse（21％）、hamburger（4％）、　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　 認識せず（21％）

・monkey（ﾏﾝｷｰ）16　　　M　　　a,i　　　　　　　22回　horse（90％）、salmon（13％）、cow（4％）、　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　 giraffe（4％）、candy（4％）

・wolf（ｳﾙﾌ）　　　24　　　W　　　u,u,u　 　　　　19回　horse（63％）、cow（10％）、tiger（26％）

・tiger（ﾀｲｶﾞｰ）　　0　　　　T　　　a,i,a　　 　　　　25回　wolf（20％）、horse（88％）、salmon（4％）

・lion（ﾗｲｵﾝ）　　　16　　　L　　　a,i,o　　 　　　　21回　wolf（95％）、horse（14％）、hambuger（4％）

この結果を見ると、認識率に影響しているのは母音の”u”ということがわかる。また、”a,u”の組み合わせはこのシステムにとって非常に認識しやすい音になっているようだ。反対にした”u,a”でも認識率60％を確保している。

"a,i”や"e,a”の組み合わせは、人間では聞きやすくてもシステムにとっては認識しにくい音になっていた。

子音については確定的ではないが、”C”から始まる英単語は認識しやすいということが読み取れた。反対に”T"や”B"は認識率が特別悪かった。

　次に誤認識について。総合的にみると"wolf"と"horse”がパーセンテージが高いものとなった。”horse”に関しては今回の認識テストにおいて必要ないので単語辞書から削除しておくべきである。必要である単語以外を登録すると思わぬところで障害があることがわかった。

飲食物英単語の読みを読みと一音目の子音と、使用されている母音を抜き出した。(ん、小文字、伸ばしは考えない）結果をいかに示す。

単語名　　　　　　　　　　　認識率　　子音、母音　　　　　　誤認識回数と％

・hamburger(ﾊﾝﾊﾞｶﾞｰ)　　 52　　　　　H　　a,a,a

・water(ｳｫｰﾀｰ)　　　　　　 20　　　　　W　　o,a

・salmon(ｻｰﾓﾝ)　　 　　　 60　　　　　S　　 a,o

・meat(ﾐｰﾄ)　　　　　　　　 4　　　　　 M 　 e,o

・candy(ｷｬﾝﾃﾞｨ)　　　　　　 4　　　　 　C 　　a,i

・fish(ﾌｨｯｼｭ) 4　　　　 　F 　　i,u

・yogurt(ﾖｰｶﾞｰﾄ) 　0　　　　 　Y 　　o,a,o