

# 音声の母音区間に対する簡易的自動検出法

—残響環境下における音声明瞭度の改善を目的として—\*

☆辻美咲, 荒井隆行 (上智大・理工), 程島奈緒 (東海大・情報通信)

## 1 はじめに

室内などの残響環境下では, 音声の聴き取りが困難になる. この原因のひとつとして, overlap-masking[1]という, ある音素に付加された残響の尾が後続する音素にマスクする現象がある. この影響を低減するため, 荒井らによって定常部抑圧処理[2]が提案された. この手法は, 単音節明瞭度試験において効果が得られた. しかし定常部抑圧処理は複雑であるため, 実用化に際し, より簡易な処理を用いる方が, 実時間処理への適用が容易であり, 利便性が高い.

本稿は, 実時間処理を将来的な目標としている. 定常部は母音に多く存在するため, 母音区間を簡易的に自動検出する手法を提案し, この手法がどれほどの精度で母音区間を検出出来るか評価した.

## 2 母音区間検出

本稿で提案する母音区間検出法は, 三好らによって提案された, 特徴量 $\alpha$ を用いた音声区間検出法[3]を参考にしている. 三好らは音声を25msのフレームごとに分け, 各フレームにおいて線形予測分析を用いてスペクトル包絡上に現れるフォルマントを推定し, 50~3000Hzの周波数帯域内に現れるフォルマントのピークを算出する. そして算出された全てのピークのパワーを足し合わせた値を特徴量 $\alpha$ としている.

我々が提案する手法は, 音声の内, 母音を検出するため, 第1, 第2フォルマントを推定する. よって上記のピークの内, パワーの大きなものから2つのピークを抽出し, それらのパワーの和をそのフレームの特徴量とする.

本稿では母音区間を決定する際, 各音声の特徴量の中央値から閾値を設定した. これよりも大きな特徴量を有するフレームを母音フレームとした.

## 3 評価手法

母音区間検出法の精度の評価に使用した刺激

は, アルカディアの音声合成ソフトウェア「SPeeCAN SFT5」を用いて作成した. 話者は女性であった. 子音(C)から母音(V)への遷移部と, VからCへの遷移部を含ませるため,  $C_1V_1C_2V_2$ の2モーラ語を使用した. これに提案する母音区間検出法を施し,  $V_1$ がどれほどの精度で検出されるか検討した. 使用したCとVの種類は,  $C_1$ と $C_2$ は/s/, /t/, /ch/, /n/の4種類,  $V_1$ は/a/, /i/, /u/の3種類を用い,  $V_2$ は/a/のみとした. 以上の内, /ti/は日本語の音節には存在しないため, 評価には用いなかった. そのため音声の総数は,  $C_1(4種類) \times V_1(3種類) \times C_2(4種類)$ の48通りから, /ti/+ $C_2V_2$ である4通りの音声を除いた44通りとした.

処理を施す前の刺激のスペクトログラムやフォルマント等を参考に目視で母音区間( $V_1$ )を判定し, 母音区間の開始点と終了点を定め, これを正解ラベルとした. 開始点と終了点に関して, 判定箇所と正解ラベルの差を誤差とし, 評価を行った. 結果は $V_1$ ,  $C_1$ ,  $C_2$ の種類において分散分析を行い, Sidakによる多重比較を行った.

## 4 結果および考察

$V_1$ ,  $C_1$ ,  $C_2$ の種類ごとにまとめた結果を, Fig. 1~3に示す. 正解ラベルから判定箇所が遅れていれば正, 早ければ負として表した. 子音部の情報は, 音声の聴き取りの際に重要であるため, 今後は子音部が母音区間と判定されることを, 誤判定であるとする. 判定開始点の値が負, または判定終了点の値が正である時,  $V_1$ の前後の子音部分を誤判定していることを意味している.

Fig. 1~3より, 全44通り中32通りの音声が, 開始点が0~35ms, 28通りの音声が, 終了点が-35~0msという範囲の結果が得られた. 使用した音声の $V_1$ の長さは, 平均が95msであった. よって, この手法は使用した音声の内, 約6割の音声に対し, 母音区間のみを正しく検出することが出来たと言える. また, 全44通り中40通りの音声において, 50ms以内の誤差で母音

\* Simple automatic vowel detection method for improving speech intelligibility in reverberant environments, by TSUJI, Misaki, ARAI, Takayuki (Sophia University) and HODOSHIMA, Nao (Tokai University)

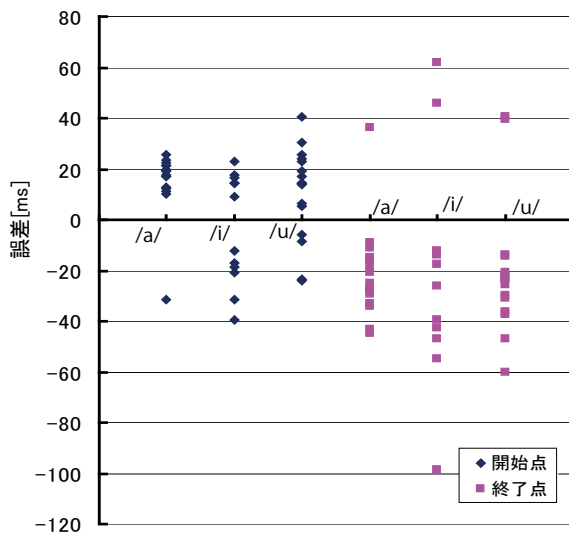


Fig. 1  $V_1$  の種類ごとにまとめた結果

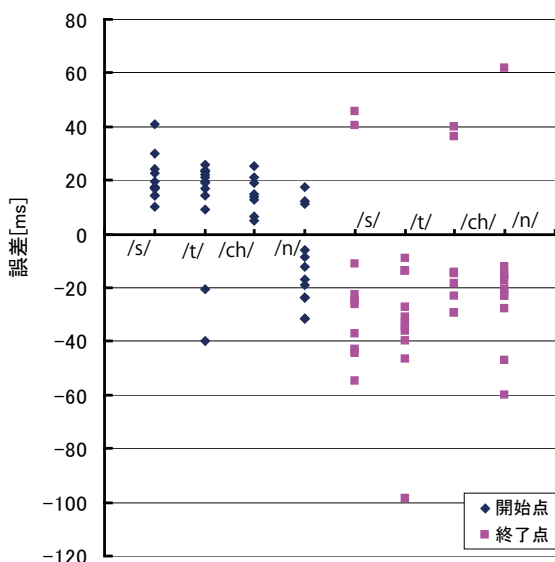


Fig. 2  $C_1$  の種類ごとにまとめた結果

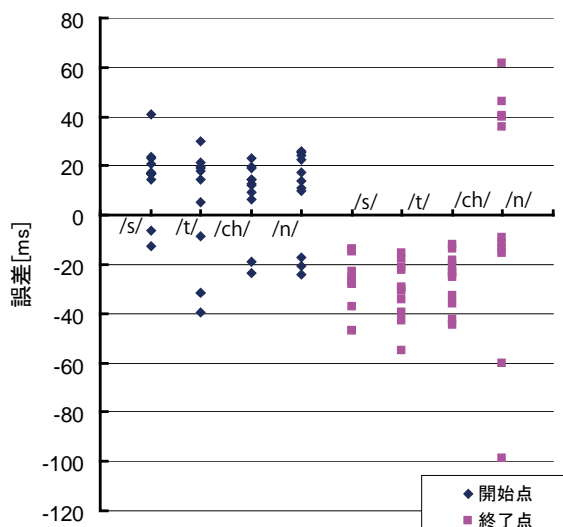


Fig. 3  $C_2$  の種類ごとにまとめた結果

区間検出を行うことが出来た。

Fig. 1 より、 $V_1$  が /a/ である 16 通り中 15 通りで開始点が正、終了点が負の値をとった。一方  $V_1$  が /i/ である 12 通り中 6 通りの開始点が負となった。開始点について、/a/ の平均値は +15 ms、/i/ の平均値は -4 ms であり、/i/ の方が子音区間を誤判定していることが多く、これらの間には有意差が見られた ( $p < 0.05$ )。よってこの手法は、/i/ の直前の子音区間を誤判定しやすいことが分かった。これは /i/ のパワーが /a/ のパワーよりも小さかったため、子音と母音との差が少なく、判別が困難であったからであると考えられる。

Fig. 2 より、開始点に関して、/s/, /t/, /ch/ の平均値は正であったのに対し、/n/ の平均値は負となった。さらに、/n/ とその他の全  $C_1$  の間に、有意差が見られた ( $p < 0.05$ )。これは /n/ に後続する母音が鼻音化し、子音と母音の判別が困難になったからであると考えられる。

Fig. 3 より、終了点に関して、正の値をとったものは  $C_2$  が /n/ のものだけだと分かった。これは  $C_1$  ごとの結果と同様に、母音の後に鼻音が続く、母音を鼻音化したためにこのような結果が得られたと考えられる。

## 5 まとめ

今回提案した母音区間検出法は、評価に使用した 44 通りの音声中、6 割の音声に対し、母音区間のみを正しく判定することが出来た。しかし鼻音を含む音声では、子音区間を母音区間と誤って検出しやすいことが分かった。

今後は母音区間に誤って判定した子音を保持する手法を検討し、さらにこの手法を用いて、残響環境下での音声明瞭度の改善法を検討していきたい。

## 謝辞

本研究は、上智大学オープン・リサーチ・センター「人間情報科学プロジェクト」の支援を受けて行われた。

## 参考文献

- [1] A. K. Nabelek *et al.*, J. Acoust. Soc. Am., 86 (4), 1259-1265, 1989.
- [2] T. Arai *et al.*, Acoust. Sci. Tech., 23, 229-232, 2002.
- [3] 三好徹, 上智大学大学院理工学研究科電気・電子工学専攻 2004 年度修士論文.