

## 声優の演技力向上を目的としたセリフの「間」習得支援システム

大谷 紀子 研究室

2272029 笠原 吉平

### 1. はじめに

声優は俳優とは異なり、感情や場面を表情や体の動きで表現するのではなく、声の抑揚やテンポ、間の取り方で表現する。特に、間の取り方は、前後のセリフの印象を決定づけることに大きく影響する。しかし、声優が使う台本には、どこでどのくらいの間を取るのかなど具体的な指示が書かれていないことが多く、声優が自ら考えて演技をしなければならない。養成所で間の取り方の指導を受ける機会もあるが、「少し長めに間を取って」や「その場面ではそんなに間を取らない」など、抽象的な表現での指導に留まる。また、レッスン外で間の取り方を個人練習する場合、自ら間の長さを測る必要があるが、測ることに専念すると肝心の演技に身が入らなくなる。現状、声優志望の演技初心者が、一人で間の取り方を客観的に確認しながら演技に集中することは、極めて難しい。

本研究では、声優志望の演技初心者が一人で間の取り方を習得することを目的とし、ユーザが音声入力した特定のキャラクターのセリフにおいて適切な間の取り方を学ぶシステムを構築する。

### 2. システムの概要と評価基準

本研究では、声の演技を収録した音声データを発話区間と非発話区間に分け、ある発話区間の終端から後続する発話区間の始端までの時間を「間」と定義する。発話区間の検出には、音声区間検出アルゴリズム WebRTC VAD[1]を用いる。WebRTC VAD は、入力音声を 10~30ms のフレーム単位に分割し、各フレームにおける発話の有無を判定する。発話なしと認識された連続するフレームの時

間を間の長さとする。

本システムを利用するにあたり、ユーザはまず提示された 5 つのセリフから対象とするセリフを選択し、模範音声を再生する。各模範音声の間の長さは事前に算出済みである。続けて、自身の習熟度や録音環境に基づき、判定許容誤差  $E$ 、発話検出感度、および語尾補正範囲を設定する。判定許容誤差  $E$  は、模範音声とユーザ音声の間の差異を許容する範囲であり、0.1~1.0 秒の数値で指定する。発話検出感度は、WebRTC VAD が音声を検知するための閾値であり、周囲の騒音レベルに合わせて指定する。語尾補正範囲は、WebRTC VAD が判定した発話終端の位置を変更するための項目であり、語尾の呼吸音などを「間」とする音量幅を指定する。

パラメータ設定後、ユーザは模範音声を聴取し、PC のマイクを通じて自身の音声を録音する。入力された音声は 16kHz モノラル形式に変換され、設定した発話検出感度を用いて WebRTC VAD により発話区間を検出し、語尾補正範囲に基づいて発話終端の位置を変更して、間の長さを算出する。模範音声とユーザ音声に含まれる複数の間について、それぞれの出現順序に基づき、前方から順に対応付ける。

表 1：判定記号と判定条件

判定記号	判定条件
◎	$d \leq E \times 0.25$
○	$E \times 0.25 < d \leq E \times 0.50$
△	$E \times 0.50 < d \leq E$
-	$E < d$

表 2：第 1 回および第 7 回試行における時間差  $d$  の比較

セリフ内の間	試行	被験者 A	被験者 B	被験者 C	被験者 D
間 1	第 1 回	0.45	0.68	0.52	0.71
	第 7 回	0.12	0.42	0.57	0.15
間 2	第 1 回	0.55	0.72	0.64	0.65
	第 7 回	0.15	0.38	0.28	0.12
間 3	第 1 回	0.62	0.55	0.55	0.48
	第 7 回	0.08	0.45	0.14	0.10
間 4	第 1 回	0.62	0.80	0.45	0.70
	第 7 回	0.10	0.52	0.10	0.18

解析終了後、入力音声と模範音声における間の長さの差  $d$ 、入力音声における間の長さの適切性を表す判定記号、および入力音声の波形グラフが表示される。判定記号は◎、○、△、-の 4 種である。判定条件を表 1 に示す。入力音声の波形グラフでは、入力音声と模範音声の発話区間および間の区間が異なる網掛けで表示される。また、セリフ内に含まれる間のうち、判定記号が◎および○となった箇所の割合を一致率とする。

### 3. 評価実験

評価実験では、本システムの有効性を確認するため、提案システムによる間の可視化およびフィードバックが、声の演技の自習において技能向上をどの程度支援できるかどうかを評価する。被験者は、声優の演技経験 2 年未満の 4 名とする。被験者には、4 箇所の間を含む同一のセリフを課題として提示した。また、間の個数の不一致による誤照合を回避するため、模範音声と同一の箇所に間を設けるよう被験者に教示した。判定基準となる判定許容誤差  $E$  は、4 名一律で 0.5 秒に固定した。計 7 回の反復練習を行わせた後、利点・改善

点に関するアンケート調査を実施した。

表 2 に、第 1 回および第 7 回の試行における時間差  $d$  を示す。第 1 回では  $E$  を上回る数値が散見された。第 7 回では  $d$  の大部分が  $E$  を下回り、練習による精度の向上が数値から確認できる。

被験者 4 名の試行回数ごとの一致率の推移を図 1 に示す。第 3 回までは習得速度に個人差が見られたものの、第 4 回以降は 4 名の一致率に顕著な向上が確認された。

### 4. 考察

図 1 より、反復練習によって模範に近い間の取り方を習得できることが示されたため、本システムは技能向上に有用であったといえる。また、アンケートでは、本システムのよかった点として、「秒数で具体的な修正の方向が見えたので迷わず練習できた」との回答を得た。客観的な指標の提示は、演技初心者の抱く不安を解消し、効率的な反復練習を促進させる効果があるといえる。

一方で、改善点として、「自分の録音音声の即時再生機能」など機能面での要望や、「スマホ対応」など利便性に関する意見が挙げられた。また、本実験の被験者数は演技経験 2 年未満の 4 名に限定されているため、習熟度の異なるより広範な被験者層に対する検証や、抑揚・音圧等の指標追加による表現の多様性への対応を今後の課題とする。

### 参考文献

- [1] 石塚健太郎, 藤本雅清, 中谷智広, “音声区間検出技術の最近の研究動向”, 日本音響学会誌, 65 巻, 10 号, pp.538-541, 2009.

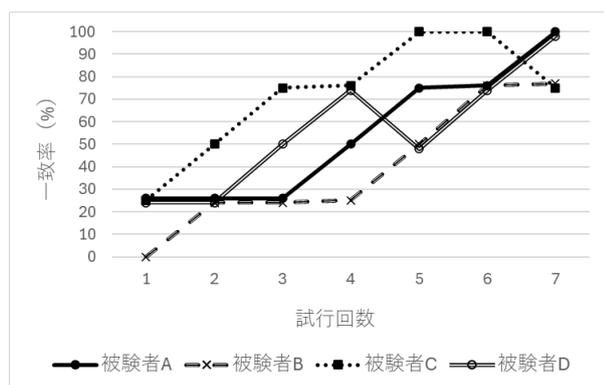


図 1 試行回数ごとの一致率の推移