

## ◆タイトル◆

近年の通信ネットワークの急速な発達により、テレビだけでなくウェブ上でニュースが配信されるなど、膨大な量の中から必要な情報のみを取捨選択する必要がある。これは、通信・放送業界では地上デジタル放送の開始や、新たな高速通信規格の誕生などに伴い、誰もがテレビやパソコンだけでなくスマートフォン・タブレットなど様々なデバイスを通して好きな時に好きな場所で情報を得ることが容易な時代となったためである。ニュース番組にはスポーツ、経済、社会、天気予報など様々なジャンルのトピックで構成されており、視聴者は時間の都合や興味の違いに応じて必要なトピックのみを早送りなどをして視聴している。

そこで、容易な情報の取捨選択のために、情報にインデクスを付与するインデクシングという技術がある [1]。インデクスが付与されていれば、所望のトピックのみを視聴することができる。また、情報を管理する側も、データベースの構築が容易になるというメリットがある。しかし世の中には膨大な量の情報が存在しているため、それら全てに人手でインデクスを付与することは事実上不可能であるため、自動的にインデクシングすることが望まれる。

インデクシングを行う上での重要な構成要素として「発話区間」「発話者」「発話内容」があり、これらを推定する技術をダイアライゼーションと呼ぶ。

発話区間の検出するための技術を VAD(Voice Activity Detection) 技術といい、先行研究 [山口さん] では音源識別によって音声データから「音声」「背景雑音」「音楽」「無音」の 4 種類の区間を検出することで発話区間の検出をおこなった。音源識別を行うためのスペクトルの傾きや変化を含めた 7 つのパラメータを用いており、音声区間検出の F 値は 0.948 であった。

発話内容の検出には音声データから音声認識を行う必要がある。近年の音声認識システムには深層学習を用いた音声認識によって成果が確認されている [音声認識技術の変遷と最先端]。また、発話者の声の特徴に着目した先行研究 [吉村さん] では会議音声認識を対象とした話者適応を行っており、音声認識の Acc(Word Accuracy) が 74.56% から 75.23% に向上した。

発話者の特定には先行研究として、i-vector を用いたニュースアンカー (アナウンサー) の検出手法が提案されている。ニュースアンカーの検出精度が % から % に向上した。しかし、短い発話からは i-vector が話者の特徴を十分に抽出することができず、ニュースアンカーとして検出されない発話の原因として述べられている。

本研究の目的は、先行研究でアンカーとして特定できなかった短い発話に着目してアンカーの発話検出精度の向上することである。

目的を達成させるためのキーアイデアとして同一話者の発話区間の結合がある。

検証の結果、F 値は 0.70 から 0.772 になった。

発話区間を結合することでアンカーの短い発話の検出精度が向上した。