N-34　遠近感のある混合音声を生成するリング型音声混合方式の評価

谷井　一人†　清水　敬司†　高原　厚†
Kazuhito Tanii　Takashi Shimizu　Atsushi Takahara

1．はじめに

Interactive Distance Learning（IDL）では、他の参加者と時間を共有している事による臨場感が学習効率の向上に貢献すると考えられている。このような臨場感を実現するために、筆者らは遠近感のある混合音声を生成するリング型音声混合方式を提案している[1]。

本稿では、予め用意した音声データファイルを用いて行ったシミュレーション実験とその結果について述べる。そして実験結果から提案方式が遠近感のある混合音声を生成する事、および音声がリングをまわりこむ事によるエコーを適切に制御可能である事を示す。

2．本研究における臨場感

臨場感は、例えば、学校の教室等を考えた場合、座席の近い人の音声は比較的明瞭に聞こえ、座席の遠い人の音声は不明瞭ながらも聞こえるという遠近感のある音声を聞く事により生まれると考える。つまり、座席の近い人の音声は比較的明瞭なため、講義内容についての相談等を行うことができ、また、座席の遠い人の音声は不明瞭ながらも聞こえるため、講義内容に対する参加者の反応や、全体的な雰囲気を捉えることができるような状況である。このような遠近感のある混合音声をネットワークを介して擬似的に生成するシステムとして、リング型音声混合方式を提案している。

3．リング型音声混合方式

3．1　システム構成

リング型音声混合方式では1参加者に対応するミキシング機能（以後、ミキシングユニットと呼ぶ）を1つ以上持った装置をネットワーク上に配置する。次に、ネットワーク上に配置したミキシングユニット間をリング状に接続するチャネルを設ける。このチャネルは、複数の発話者の音声データを混合した音声データを転送するために使用する。音声データの送信は双方向であるため、ミキシングユニット間にそれぞれ2チャネル用意する。これに加えて、発話者の音声をk段先のミキシングユニットまで混合せずに直接転送するチャネルを用意する。つまり、このチャネルにより、発話者から双方向にk段以内のミキシングユニットには直接音声が届くことになる。一方、それより先のミキシングユニットへは混合音声チャネルを通って他の音声と混合された音声として届く。

3．2　ミキシングユニット

ミキシングユニットは、各発話者の音声データ、混合音声データを入力音声データとし、両隣のミキシングユニットへの出力音声データと参加者端末への出力音声データの作成という2種の混合処理機能を持つ。ミキシングユニットにおける入力音声と出力音声の関係を図1に示す。ここで、

- 全参加者数をN、各参加者をn_i(i=1, 2, ・・・・, N)
- 参加者n_iから送信される音声データをd_i
- 参加者n_iに対応したミキシングユニットをM_i
- d_iが混合されずに到達するミキシングユニットの数をk（k=2, ・・・・, N）

とする。

また、各音声データを便宜的に以下のように表記する。

- M_iにおける時計回り方向のミキシングユニットへの出力音声データをSr_i
- M_iにおける反時計回り方向のミキシングユニットへの出力音声データをSl_i
- 参加者n_iが聞く音声データをZ_i

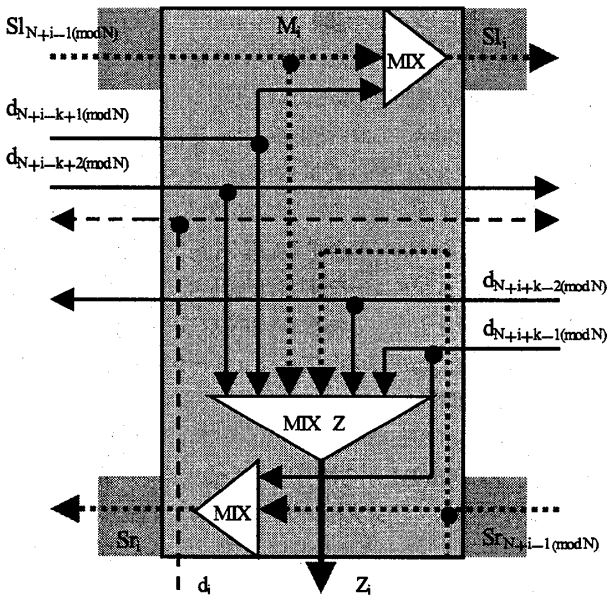

図1　ミキシングユニットにおける音声処理

ミキシングユニットへの出力音声混合処理

例としてk＝3とした場合のミキシングユニット間における音声データの流れについて述べる。参加者n_iの音声はミキシングユニットM_{i+2}、M_{i-2}までは他の音声と混合されずに参加者へ届けられ、そこで、他の音声と混合される。それより先のミキシングユニットに接続されている参加者には、他の音声と混合された状態で届く。ここで、

- 次段のミキシングユニットへの出力音声データ作成時における直接転送音声の混合割合をα

とし、M_iにおける次段のミキシングユニットへの出力音声データSr_i、Sl_iを次式に従って処理を行う。Sr_iは時計回り方向への出力であり、Sl_iは反時計回り方向への出力である。

$$Sl_i = (1-\alpha)Sl_{N+i-1(\mathrm{mod}\ N)} + \alpha d_{N+i-k+1}$$
$$Sr_i = (1-\alpha)Sr_{N+i+1(\mathrm{mod}\ N)} + \alpha d_{N+i+k-1}$$

ただし、＋記号は音声混合処理を表す。

本システムでは、参加者の音声は双方向に対称に伝播していくため、リング上で一番離れた参加者に対して音声が伝播すれば十分である。そこで、一番遠く離れた参加者まで伝播する音声成分をω（$0 \leqq \omega \leqq 1$）とし、次式を満たすようなαを設定する。

† 日本電信電話株式会社　未来ねっと研究所

$$\alpha(1-\alpha)^{N/2-k} > \omega$$

参加者端末への出力音声混合処理

参加者が聞く音声は、双方向に対してk－1段以内にいる他の参加者から直接転送されてきた音声データと、双方向分の混合音声データを全て混合した音声になる。よって、M_iにおける参加者端末への出力音声データ、すなわち参加者n_iが聞く音声データZ_iは次式で表せる。

$$Z_i = \beta_r Sr_{N+i+1(\mathrm{mod}\ N)} + \beta_l Sl_{N+i-1(\mathrm{mod}\ N)} + \sum_{a=1}^{k-1} \beta_{ra} d_{N+i-k+a(\mathrm{mod}\ N)} + \sum_{a=1}^{k-1} \beta_{la} d_{N+i+k-a(\mathrm{mod}\ N)}$$

ただし、$1 \leqq a < k$

$$\beta_r + \beta_l + \sum_{a=1}^{k-1} \beta_{ra} + \sum_{a=1}^{k-1} \beta_{la} = 1$$

である。

4．シミュレーション実験

リング型音声混合方式では、混合された音声はリング上を伝播していくため、ある参加者の発した音声がリングを1周し、その参加者に戻ってきた場合、エコーと認識され、その参加者の音声通信を阻害することになる。このエコー対策として、ある参加者の発した音声がリングを1周伝播することがないように、直接転送音声と混合音声を適当な割合αで混合する処理を行っている。そこで、リング型音声混合方式がエコーを適切に制御している事、および遠近感のある混合音声を生成している事を確認するため、フリーソフトウェアのSoundEngine[2]を用いてシミュレーション実験を行った。

4．1　実験概要

実験には、①400Hzから19200Hzまでの30種のsin波（s1～s30）と、②CD-ROMから抽出した9種の音声データを用いた。いずれのデータも、サンプリング周波数44.1kHz、16bit量子化PCM形式である。これらのサンプルデータを用いて、以下のようなシミュレーション実験を行った。

[実験1] 図2に示すように、①のデータを混合割合α＝0.6として混合処理していき、各段階で生成されたデータに対してFFTを行い、周波数特性を評価した。次に、②のデータを用いて同様の処理を行い、生成された音声データを聞き主観的評価を行った。

[実験2] 図1 MIX Zの混合処理をシミュレーションするため、双方向から直接転送されてくる音声として4種のsin波、双方向からの混合音声2種を用意し、全てのデータの混合割合が等しくなるように混合処理し、生成された混合音声に対してFFTを行い周波数特性を評価した。次に②のデータを用いて同様の処理を行い、生成された音声データを聞き主観的評価を行った。

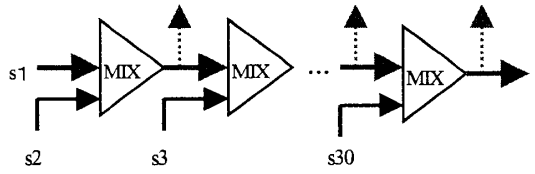

図2　実験1におけるデータ混合処理

4．2　結果

[実験1] ①のデータを用いた実験結果として、13段の混合処理により生成したデータにFFTを行った結果を図3に示す。人間の耳に聞こえる最小の音を0dBとすると、普通の会話音声は約60dBと言われている。今回実験に用いたFFTソフトウェアでは、最大値が0dBで表現されるため、-60dBが聞こえる音の限界と考えられる。図3では6段の混合処理を経ると-60dBを下回り、13段では約-90dBまで減衰している。すなわち、13段以上の混合処理段数を経て発話者に音声が戻って来ても、聞こえる音量ではないためエコーと認識される事はないと考えられる。

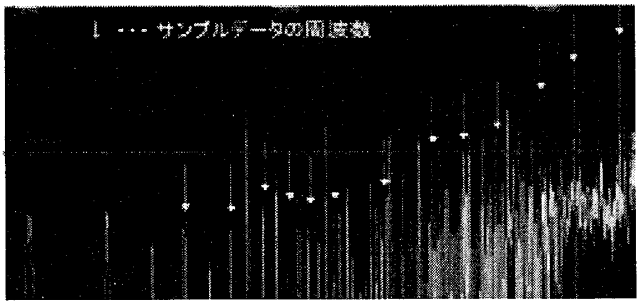

図3　実験1におけるFFTの結果

②のデータを用いた主観的評価では、混合段数7段を超えると元の音声が聞こえない事が確認できた。これらの事から、参加者数Nが15、k＝2の場合、α=0.6に設定することで、適切にエコーを制御する事ができる。

[実験2] 混合処理により生成した混合音声に対してFFTを行った結果を図4に示す。

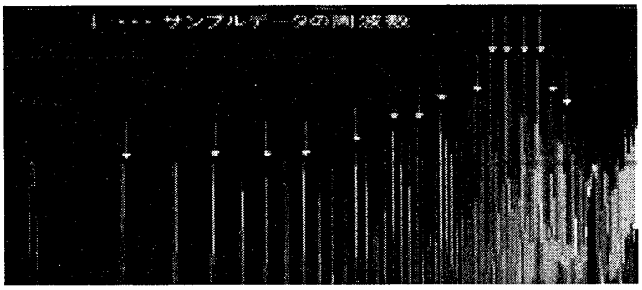

図4　実験2におけるFFTの結果

直接転送音声と見立てた4種のsin波の周波数成分は約-20dBと強く残っている事、及び対象参加者からの距離が遠くなるに従い、周波数成分が弱まっている事が確認できた。②のデータを用いた主観的評価では、直接転送音声が比較的明瞭に聞こえ、混合音声は不明瞭ながらも聞こえる状態が確認できた。

5．まとめ

本稿では、ネットワーク上で遠近感のある混合音声を生成するリング型音声混合方式の評価として、音声データファイルを用いて行ったシミュレーション実験について述べた。実験の結果から提案方式が遠近感のある混合音声を生成する事、および音声がリングをまわりこむ事によるエコーを適切に制御可能である事を示した。

参考文献

[1] 谷井、清水、高原 “遠隔教育における臨場感実現のための音声ミキシング手法”, 電子情報通信学会全国大会, D-15-29, 2002

[2] Cycle of 5th WebPage (SoundEngine)
http://www.cyckeof5th.com/