

## 通信会議における遠方音声収録のための残響抑圧方式の研究

古家, 賢一

<https://doi.org/10.15017/459055>

---

出版情報 : Kyushu University, 2005, 博士 (芸術工学), 論文博士  
バージョン :  
権利関係 :

# 第1章

## 序論

### 1.1 研究の背景

21世紀に入り、電話・インターネットなどに代表される情報通信は、メディアのデジタル化、ネットワークのブロードバンド化により急激な進歩を遂げ、現在も大きく発展し続けている。すでに各家庭には100Mbpsの大容量光ネットワークを引くことが可能となり、より魅力的で豊かなサービスを受けることが誰にでもできるようになってきている。このようなブロードバンド時代における双方向通信サービスの夢の一つとして、遠く離れた相手とその場で自然に会話しているような臨場感のある通信の実現がある [1][2]。臨場感のある通信を実現するためには、五感で得られるすべての情報を相手側に送り再現する必要があるが、その中でも音（音声）は意思や感情などを伝えるためには最も重要かつ不可欠な情報である [3][4]。

例えば、これからの少子高齢化、核家族社会、においては、様々な理由により家族が必ずしも同じ場所に住めるとは限らないが、遠く離れた両親といつでも距離を越えてネットワークで繋がり、図 1.1 示すような、まさにその場で会って会話しているような臨場感のある自然な通信が実現できれば、人の生活はより豊かになると考えられる。遠く離れた両親にとっては、従来の電話などの用件を伝えるだけの通信に比べ、雰囲気など伝えることができる臨場感のある自

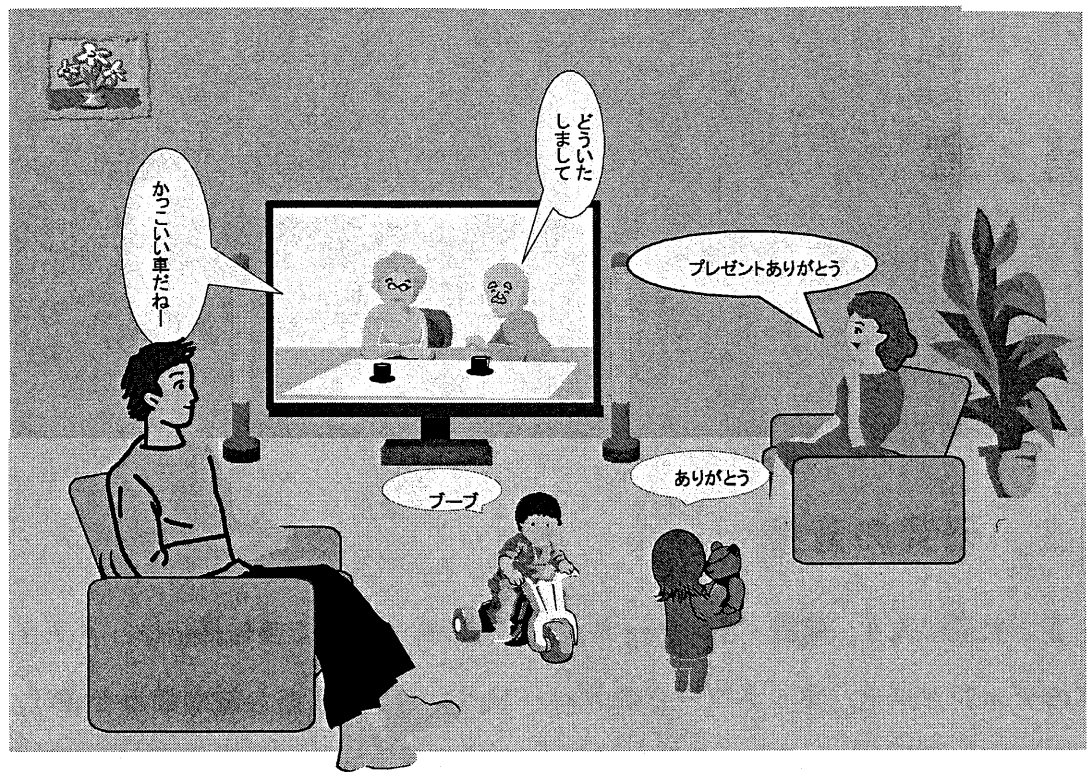


図 1.1: ブロードバンド時代における家庭での臨場感ある自然な通信

然な通信では、いつでも子供や孫に自宅に居ながら会うことができ、子供にとっても両親の元気な様子を感じる事が可能となる。

また、近年、ビジネスにおいても出張などの旅費の節約や移動時間の節約といった経済的側面、テロやSARSなどによる緊迫した社会情勢による海外および国内出張の禁止や自粛などの社会的側面、及び出張時の乗り物による二酸化炭素排出規制など地球温暖化に代表される環境的側面を背景として、人が乗り物によって移動する出張を減らさざるを得ない要因が増え、遠隔地をネットワークで結んで行うTV会議、音声会議、パソコンを用いたデスクトップ会議、遠隔教育、遠隔医療などの共同作業が盛んになって来ている。

従来から通信に用いられている音声入出力装置として、電話機のハンドセットやヘッドセットがあるが、これらは単に話の内容のみ伝えるには十分である

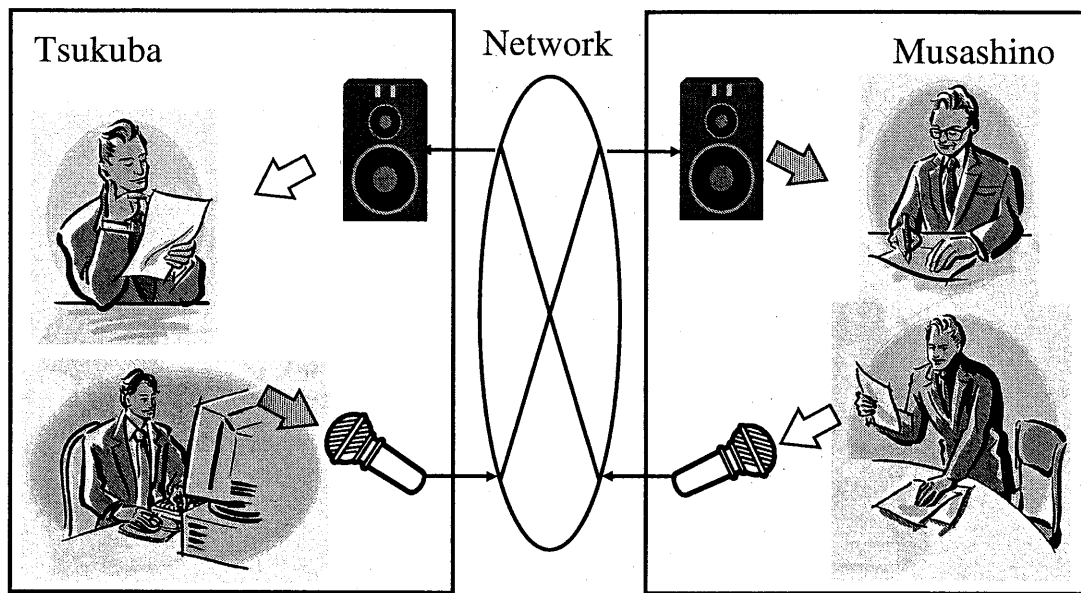


図 1.2: ハンズフリー拡声通話

が、片手が塞がる制約や装着の煩わしさがああり、自然な会話とは程遠く、臨場感ある通信の実現を妨げる。これらの不便さを解消したものがハンズフリー拡声通信である。ハンズフリー拡声通信は、図 1.2 に示すように相手の声をスピーカで再生（拡声）し、自分の声を卓上においたマイクロホンで收音して会話を行う。ハンズフリーで会話を行うことにより、より自然に近い通信が可能となる。また、スピーカによる拡声を行うことにより、他の何かをしながら、複数人同士が同時に会話することが可能となる。例えば、遠隔地の両親 2 人と洗い物をするお母さん、ソファで新聞を読むお父さん、リビングで遊ぶ孫 2 人が自然に 6 人同時に会話することがハンズフリー拡声通信で実現できる。また、ハンズフリーによってビジネスにおいては議事録の筆記やパソコン操作など作業しながら、大人数の会議が実現できる。このように、会話の自然性や臨場感、通信サービスの品質向上のために、より高品質なハンズフリー拡声通信の実現が強く望まれている。

また、ハンズフリー拡声通信の応用として、人間同士だけでなく機械と人間

### 音声認識合成システム

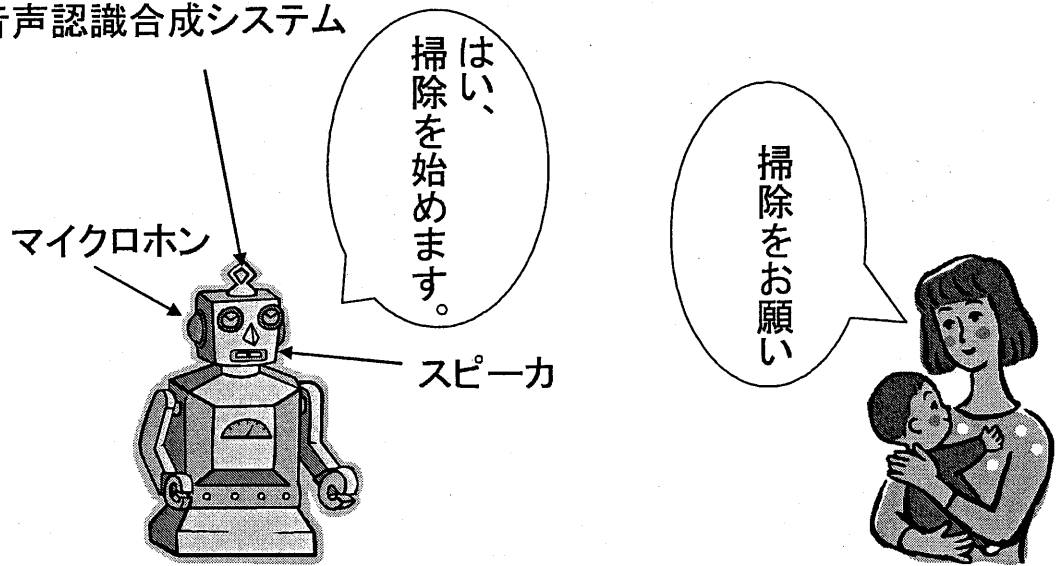


図 1.3: ハンズフリー拡声通信の音声認識合成システムへの応用例

の通信も考えられる。人間にとって最も自然で使いやすいインタフェースは音声を用いた情報の伝達である。いわゆる人間と機械の通信を、より人間同士で会話するように自然な形で行うためには、機械にマイクロホンを埋め込み、機械からの情報をスピーカから拡声するような普段、人間同士で会話しているような形で行うハンズフリー音声認識合成システムの実現が必要である。例えば、図 1.3 に示すように、ロボットに掃除を実行させたい場合、「掃除をして」と一言いっただけで、ロボットが人間の意図を理解し掃除を始める。この「アトム」のような音声認識合成システムが実現すれば、複雑化の一途を辿る機械への情報伝達にも音声を用いることによって、コンピュータに親しみのない子供や高齢者でも容易に機械を操作することが可能となる。

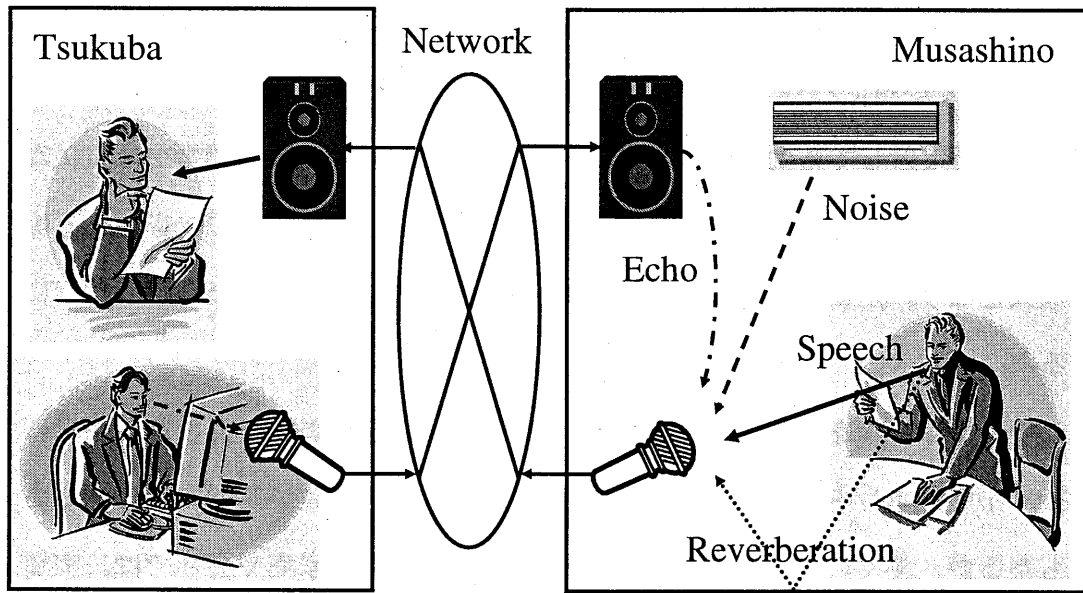


図 1.4: ハンズフリー拡声通話における音響的な諸問題

## 1.2 課題と従来技術

高品質なハンズフリー拡声通話を実現するために、現在、研究開発が進められている收音に関する課題と従来技術について説明する [5][6]。図 1.4 は、「Musashino」の人の話した声がマイクロホンに收音され、ネットワークを介して「Tsukuba」の人に拡声される状況を示している。ハンズフリー拡声通話では、マイクロホンは机の上に置かれており、ハンドセットなどに比べて口元から離れた位置で使用される。この場合、マイクロホンと口元の距離は数十センチメートルから数メートルになり、ハンドセットの数センチメートル口元につけて使用する場合より距離にして数十倍から数百倍離れるため、音声の大きさ（エネルギー）が約 20~40dB（デシベル）小さくなってしまふ。そのため、マイクロホン感度を上げる必要が生じ、相手に送りたい話者の声以外の不要音も大きな音で拾ってしまう問題が発生する。図 1.4 に示すとおり、そのような不要音は大別して 3 つあり、エアコンや OA 機器などが発する (1) 周囲雑音、スピーカから拡声された

相手の声が自分側のスピーカ・マイクロホンで「やまびこ」のように折り返し「キーン」といったハウリングをの原因にもなる(2)音響エコー, 自分の話し声が壁などで反射して発生する(3)残響音である。これらの音は, 通常では気にならない程度であっても, 先に述べた状況のためにマイクロホンで收音される際に過大となり, 送話音声を妨害して音声品質を劣化させ, 快適な会話の妨げになる[7].

このような問題を解決する次善策として放送局の番組制作のように各話者に近接マイクロホンを取り付け専用のミキシングエンジニアが個別に操作する方法もあるが, 多くの費用が掛かり, 日常使用するような家庭や会社では, 実質的に不可能であり, 広く一般に普及させることは難しい。ここでは, 家庭でも普及させられる方式について考え, 現在の電話機に代わる端末をテレビ, テーブルなどの上に設置するだけで, 高品質なハンズフリー拡声通信が実現できるような技術を検討する。

#### (1) 従来の雑音低減技術

従来の雑音低減技術としては, 目的音と雑音の信号の時間周波数的な性質を利用した方法と目的音と雑音の空間的な性質を利用した方法に大別できる。前者としては, 雑音の統計的な定常性に着目し, 1つのマイクロホンで受信した信号を信号処理することで, 雑音成分のみを低減し目的音を強調する短時間スペクトル振幅推定を用いた雑音抑圧技術がある。例えば, 最も代表的なスペクトル・サブトラクション (Spectral Subtraction) の手法[8]を始め, ウィーナー・フィルタ (Wiener Filter) を用いた手法[9], 最尤 (Maximum Likelihood) 推定を用いる手法[10], MMS E (Minimum Mean-Square Error) を用いた手法[11]などがある。これらの技術は, 雑音が重畳したマイクロホン入力信号を周波数領域などに変換し, 音声と周囲雑音の統計的な性質を利用して, 雑音成分だけを差し引くことにより周囲雑音を抑圧する。従来, 音声成分の欠落によるミュー

ジカルノイズの発生などによる処理品質の劣化が問題であったが、雑音レベル推定の精度向上 [12]-[16], 周波数領域での抑圧手法の改善 [17]-[21], マスキング効果を用いた後処理 [22][23] によって, 実用レベルに十分な方式がいくつか実現されている。後者として, 目的音と雑音の空間的な性質つまり音の到来方向に着目し, 複数のマイクロホンで受信した信号を信号処理することで, 目的音を空間的に選択收音するマイクロホンアレー技術がある。この技術には, 事前情報のみを用い, 目的音源に鋭い指向性を形成する固定ビームフォーマ [24]-[27] と, 目的音源に対する感度を拘束し, 雑音が最小となるように指向性を形成する適応型アレーがある [28]-[32]。両者を比較した場合, 適応型アレーのほうが雑音抑圧性能が高く, アレーサイズも小さくすることが可能である。

## (2) 従来の音響エコー低減技術

従来の音響エコー低減技術 [33] には, 音声スイッチに代表される挿入損制御と適応フィルタを用いたエコーキャンセラの大きく2つの技術がある。挿入損制御 [34]-[40] は, 受話端または送話端のどちらか一方に損失 (ロス) を挿入しエコーの返りを防止し, ハウリングを引き起こす一巡ループを遮断する。演算量が小さくアナログ回路でも実現できる利点がある反面, 通話を行っている2つの部屋の両方の話者が同時に話した場合 (ダブルトーク状態) に, どちらかの声が途切れ主観的な通話品質を劣化させる欠点がある。適応フィルタを用いたエコーキャンセラ [41]-[43] は, スピーカとマイクロホン間のエコー経路を適応フィルタを用いて推定し, 推定したエコー経路と受話信号から擬似エコーを模擬し, マイクロホン入力信号から差し引くことによってエコーを消去する。これにより, ダブルトークが可能となるが, エコー経路の推定精度 (=エコー消去量) や収束速度に限界がある。

## (3) 従来の残響抑圧技術

従来の残響抑圧技術については, 第2章で詳しく述べるので, ここでは, 簡単



に説明する。音源から受信点までの室内インパルス応答を測定し、その逆フィルタをマイクロホンで収録された音に畳み込む MINT 法 [44],[45] があるが、これらの方法は日常の家庭や会議など室内インパルス応答が測定できない場合への適用が難しい。室内インパルス応答を測定せず観測信号の統計的性質を利用して逆フィルタを構成する Blind Deconvolution により残響抑圧を行なう方法として、音源の長時間平均スペクトル情報を用いて最小位相逆フィルタを構成する方法 [46] などがあるが、周波数振幅特性のみ補正しており位相特性の補正を行っていないため音源波形の回復は困難であり残響抑圧量も小さい。位相特性の補正も含めた波形回復を目的とした Blind Deconvolution 法 [47], [48] が提案されているが、モデルのインパルス応答長を適切に決定しないと誤差が大きく不安定であるという問題があった。マイクロホンアレーを用いて、目的音方向に指向性を向けそれ以外の方向からの残響成分を抑圧する遅延和アレー [24] がある。この方法は、最も簡単で容易に実現できるが抑圧量を大きくするためには多数のマイクロホンが必要となる問題点があった。

### 1.3 研究の目的と本論文の構成

第 1.2 節で、高品質なハンズフリー拡声通信を実現するための課題と従来技術について述べた。そこで、本研究では、その課題のうち残響による音声品質の劣化の問題を解決すべく、家庭でも安く簡単に現在の電話機に代わる端末をテレビなどの上に設置するだけで、マイクロホンから話者までの距離が離れている場合でも、部屋の残響を十分に抑圧して音声を収録し、快適なハンズフリー拡声通話を提供できる残響抑圧技術の実現を目的とする。さらに、家庭など世の中に広く普及させるためには安価なパソコンや DSP などに残響抑圧の処理プログラムを載せる必要があるが、そのためにはアルゴリズムの高速化と演算精

度の検討が不可欠である。まとめると、快適なハンズフリー拡声通話を提供できる残響抑圧技術の実現を目的として、その実現に必要な

- (i) 残響特性の事前測定が不要な残響抑圧手法の確立
- (ii) 一般的なパソコンでリアルタイム残響抑圧処理可能な高速化アルゴリズムの確立
- (iii) パソコンやDSPなど様々なハードウェアに搭載したときの演算誤差の推定手法の確立

を本研究の目標とする。

各章の構成は以下のとおりである。

第2章では、本論文で用いられている手法に関する基礎的な予備知識および代表的な従来の残響抑圧技術について概説する。

第3章及び第4章において、残響抑圧手法の確立のため第1.2節で述べた残響特性を事前に測定しなければならなかった従来技術の問題点を解決するため、事前に測定しなくてもよい方法について検討を行う。

第3章では、残響抑圧の本来の目的である音源の波形回復に立ち返り、残響特性の事前測定の代わりに観測可能な複数マイクロホンで収録した入力相関行列と音源の波形が回復ができたときに最小となる評価関数を用いることにより残響を抑圧する二入力残響抑圧法を提案する。この方法では、まず入力相関行列の最小固有ベクトルをいろいろなモデル化次数に対して計算し、原音声が回復できたときに最小となる評価関数を用いて次数を決定する。次にその次数で推定された伝達関数の逆フィルタを計算し波形回復をおこなう。この手法は、短いインパルス応答に対してほぼ完全に波形回復が可能であることをシュミレーション実験により確認する。

第4章では、二入力残響抑圧法においては室内インパルス応答の長さが数千次におよぶと計算誤差、雑音などにより性能が劣化する問題点を解決するため、二入力残響抑圧法と同様に複数マイクロホンで収録した入力相関行列の性質に着目するが、二入力残響抑圧法のように入力相関行列の最小固有ベクトルを用いインパルス応答を推定し逆フィルタを計算するのではなく、直接、入力相関行列から方程式を立て、MINT法[44]の多チャンネル逆フィルタを計算するSemi-blind残響抑圧法を提案する。Semi-blind残響抑圧法では計算誤差、雑音に対してロバストであり、実際の室内インパルス応答に対しても効果があることを実験により確認する。

第5章において、高速化アルゴリズムの確立のため、行列演算における入力相関行列の性質に着目しFFT（高速フーリエ変換）を用いることによって、飛躍的な高速化を試みる。第4章のSemi-blind残響抑圧法における逆フィルタ計算の高速化手法を提案する。高速化のため、相関関数行列の性質である畳み込み行列との類似性に着目し、共役勾配法にFFTによる畳み込み高速演算を利用する。さらに、提案手法を組み込みパソコン上でのリアルタイム残響抑圧システムを構築し、室内で実験を行い、高速化しても残響抑圧効果が劣化しないことを確認する。

第6章では、演算誤差の推定手法の確立のため、音響信号などの実数データのFFT計算に対して実用上利用される高速ハートレー変換（FHT）および実数FFTの場合の計算誤差の白色雑音モデルを用いて演算誤差を解析し、どの程度の演算精度があれば残響抑圧が実現可能か推定する。

第7章では、本論文の結論として、研究の成果のまとめと今後の課題を述べる。