音声入力の高品質化技術とタブレットへの適用

High-Quality Voice Capture Technologies and Application to Tablet

井阪　岳彦　　須藤　　隆　　天田　　皇
■ISAKA Takehiko　■SUDO Takashi　■AMADA Tadashi

近年，音声入力を応用した機能として，ビデオチャットや音声認識が注目されている。こうした用途では，話者の声をいかにクリアに集音できるかが重要であり，高品質な音声入力が求められている。

東芝は，音声入力の高品質化に対するニーズに応えて，スピーカからマイクへ回り込むエコーを防ぐエコーキャンセラ，方向性ノイズを抑圧するビームフォーミング，及び様々な方向から来る拡散性ノイズを抑圧するノイズキャンセラを開発した。これらの音声入力の高品質化技術をレグザタブレットAT703/AT503に搭載し，快適な音声入力を実現した。

Demand has been increasing for voice input applications including video chat systems and speech recognition systems. To improve the usability of these applications, it is essential to capture voices as clearly as possible.

In order to minimize factors that degrade quality in voice input applications, Toshiba has developed the following high-quality voice capture technologies: (1) an echo canceller to suppress sounds from a speaker being picked up by a microphone, (2) beamforming to suppress directional noise, and (3) a noise canceller to suppress diffuse noises entering a microphone from various directions. These technologies have been implemented in the REGZA Tablet AT703/AT503 models, which feature a smooth voice capturing function.

1　まえがき

音声は，人と人，人と機械のコミュニケーションで重要な役割を果たしている。パソコンやタブレットをはじめとする情報機器で音声入力は，録音や，ビデオ撮影，電話，ビデオチャット，音声認識，対話システムなど，様々な場面で利用されるようになってきた。このような音声入力を快適に使えるようにするためには，話者の声を高品質に集音することが重要である。

こうした背景を踏まえ，東芝は音声入力の高品質化に取り組んでいる。ここでは，これらの技術の概要とタブレットへの適用事例について述べる。

2　音声入力の品質の劣化要因

話者の声がマイクに取り込まれるまでには，次のような音質の劣化要因がある（**図1**）。

(1)　情報機器のスピーカから出た音がマイクに回り込むエコー
(2)　ドアの開閉音のような単一の方向から来る方向性ノイズ
(3)　繁華街の雑踏のように様々な方向から来る拡散性ノイズ

音声入力の形態には，マイクが話者の口元にある“ハンドセット”と，口元から離れた“ハンズフリー”があり（**図2**），ハンドセットでは口とマイクの距離が近いため，話者の声はこれらの劣化要因の影響を受けにくい。しかし，マイクを口元に保持する必要があり，ヘッドセットマイクを装着したり，端末を手で持ったりしなければならないなど，使用上の制約がある。

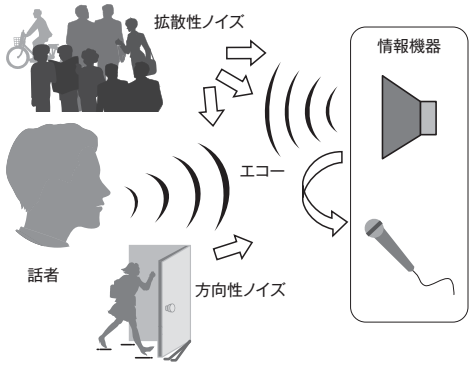

図1．音声入力の品質の劣化要因 ― 音声入力の品質には，スピーカから出た音がマイクに回り込むエコー，方向性ノイズ，及び拡散性ノイズといった劣化要因がある。
Factors degrading voice capture quality

一方，話者と端末の距離をとれるハンズフリーには，手が自由になる，大音量で再生して複数人で同時に利用できる，などの利点があるが，マイク及びスピーカのゲインを大きく設定しなければならないため，劣化要因の影響を強く受けて音質が劣化しやすい。

3　音声入力の高品質化技術

当社は，2章で述べた劣化要因の影響による音質の劣化を防ぐために，音声入力の高品質化技術の開発に取り組んでい

図2．音声入力の形態 ― マイクが話者の口元にあるハンドセットと，口元から離れたハンズフリーがある。
Scenes of voice input

る。この技術では，まずエコーキャンセラでエコーを抑圧し，次にビームフォーミングで方向性ノイズを抑圧し，最後にノイズキャンセラで拡散性ノイズを抑圧して，話者の声だけを高品質に抽出する（**図3**）。それぞれの技術の詳細について，次に述べる。

ハンズフリー通話では，スピーカから出る通話音がマイクに回り込むと，時間遅れを伴って通話相手にエコーとして戻っていき，通話しづらくなる。また，相手側の通話音量も大きいと，ハウリングが起きて通話が困難になる場合がある。エコーキャンセラは，スピーカから出る前の音声を参照してエコーを推定し，マイク入力に含まれるエコーを抑圧し，ハウリングを防ぐ技術である。マイク入力には，エコー以外にも一般に周囲ノイズや自分の声も混ざる。エコーを精度よく推定して抑圧しないと，相手の声だけでなく，自分の声も抑圧してしまい，同時通話性の低下を招く。そのため，マイク入力からエコーだけを精度よく推定して抑圧する独自のエコーキャンセラ技術を開発した。

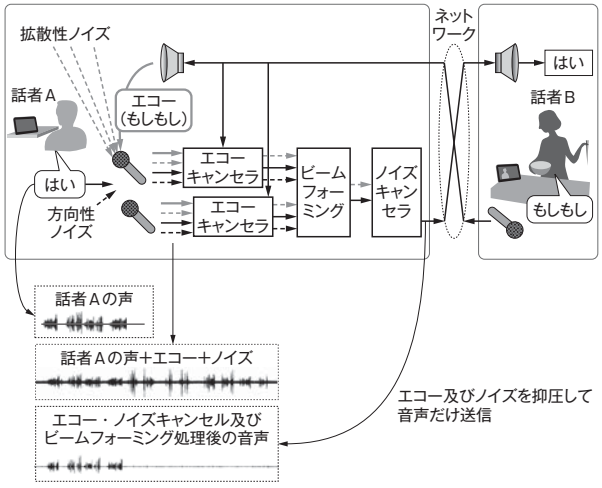

図3．音声入力の高品質化の原理 ― スピーカからマイクに回り込むエコー，方向性ノイズ，及び拡散性ノイズを抑圧することで，話者Aの声だけを送信する。
Principle of high-quality voice capture

ビームフォーミングは，複数のマイクとこれらに到達した時間差（位相差）を利用して，特定の方向から来る方向性ノイズを抑圧する技術である。話者が，左右に等距離で二つ並んだマイクに向かって話すと，二つのマイクにはほぼ同時に声が入る。このとき，二つのマイクに同時に入ってきた音を話者の声とみなし，時間差のある音を話者の声以外の音，すなわち方向性ノイズとみなして抑圧する。

一方，繁華街の騒音のように，様々な方向から来る拡散性ノイズをビームフォーミングで抑圧することは一般に難しい。拡散性ノイズは，二つのマイク間で音声の相関が小さく，位相差を算出することが難しいためである。ノイズキャンセラは，平均的な周波数特性からノイズ成分を推定して抑圧する技術である。特に拡散性ノイズは，様々なノイズが混ざって周波数特性が定常状態になりやすく，平均的な周波数特性で安定するため，ノイズキャンセラによる抑圧が効果的である。

これらの技術のうち，エコーキャンセラとノイズキャンセラは既に携帯電話用音声信号処理コーデックLSI TC94B24WBGに適用し，高いエコー・ノイズ抑圧量を1マイクで実現している[(1)]。今回，ビームフォーミングも含めて音声入力の高品質化技術をタブレットへ適用したので，その事例の詳細を次に述べる。

4　タブレットへの適用

ネットワーク環境の普及と通信コストの低下によって，誰でも手軽にコミュニケーションできるようになり，タブレットは遠隔地とのビデオチャットに活用されている。また，ビジネスや教育の場でも，講義や会議の録音やビデオ撮影に活用されている。こうした利用シーンではハンズフリーで音声入力されることが多く，タブレットでの音声入力の高音質化のニーズが高まっている。

4.1　タブレットの概要

当社は，マルチメディアコンテンツを快適に楽しめるレグザタブレット　ATシリーズを開発してきた。今回開発したAT703には，省電力に優れたモバイルプロセッサNVIDIA® Tegra™（注1）4-PLUS-1と二つのマイクを搭載した（**図4**，**表1**）。

今回，音声入力の高品質化をソフトウェアで実現したことで，従来必要であった専用LSIを削減し，ハードウェアコストを抑えた。また，音声入力の高品質化を，アプリケーションに依存せずに実現し，例えば，ビデオチャットで会話しやすくした。更に，録音やビデオ撮影では，再生時に内容が聞き取りやすく長時間聞いても疲れにくい，快適な音声入力を実現し

（注1）　NVIDIA，Tegraは，米国及びその他の国におけるNVIDIA Corporationの商標又は登録商標。

図4. レグザタブレットAT703 — NVIDIA® Tegra™ 4-PLUS-1及び二つのマイクを搭載したAndroid™(注2)タブレットである。
REGZA Tablet AT703

表1. レグザタブレットAT703の主な仕様
Main specifications of REGZA Tablet AT703

項　目	仕　様
プラットフォーム	Android 4.2
プロセッサ	NVIDIA®Tegra™4-PLUS-1 モバイル用クアッドコア 動作周波数 最大1.8 GHz
ディスプレイ	10.1型ワイドWQXGA TFTカラー PixelPure LED液晶 （広視野角，及び省電力LEDバックライト）
メモリ容量	2Gバイト
ストレージ容量	32Gバイト
Webカメラ	本体前面（有効画素数 約120万画素）×1 本体背面（有効画素数 約800万画素）×1
マイク	ステレオマイク
外形寸法（突起部含まず）	約 261（幅）×179（奥行き）×10.5（厚さ）mm
質量	約 671 g

WQXGA：2,560×1,600画素　　TFT：薄膜トランジスタ

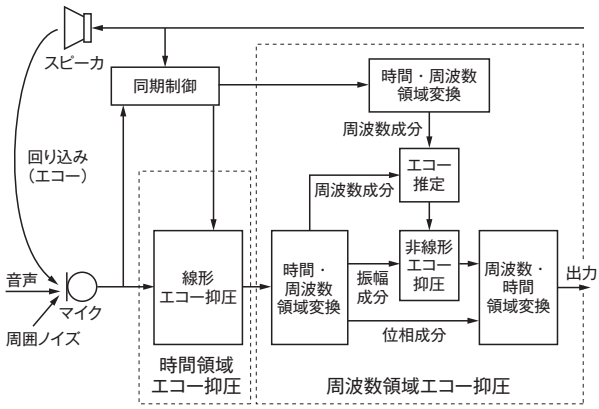

図5. 線形・非線形エコーを抑圧するエコーキャンセラ — 時間領域と周波数領域の処理を組み合わせることで高いエコー抑圧量と同時通話性を両立させた。
Echo canceller to suppress linear and nonlinear echoes

た。タブレットに適用した当社独自の方式による音声入力の高音質化技術について，次に述べる。この技術は，AT703の廉価モデルであるAT503にも適用している。

4.2　線形・非線形エコーを抑圧するエコーキャンセラ

タブレットでは，スピーカとマイクが一つの筐体（きょうたい）に搭載され，スピーカから大きな音量で再生されることが多い。この場合，スピーカのひずみや筐体を伝わる振動によって非線形にひずんだエコーがマイクに回り込みやすい。また，ソフトウェア処理の場合，スピーカ出力とマイク入力を独立に処理するため，スピーカ出力とマイク入力の信号間の時間間隔が一定とならない。これらによって，正しいエコー抑圧が難しくなり，エコー抑圧量が劣化しやすい。

そこで当社は，独自のエコーキャンセラ技術[2]を開発し，タブレットに搭載した。この技術では，スピーカ信号とマイク信号を同期制御したうえで，時間領域のエコーキャンセラでマイク入力から線形エコー成分を除去してから，その信号を周波数領域に変換して周波数領域で非線形エコー成分を推定してエコー除去した後，時間領域に戻して出力する（**図5**）。これによって，非線形にひずんだエコー，自分の声，及び周囲ノイズの混じった音からエコーだけを精度よく推定して抑圧できる。また，同期制御を採用することで，従来は専用LSIを用いていた処理をソフトウェアで行えるようにして，かつ従来と同等以上のエコー抑圧量と同時通話性の両立を実現した。

4.3　方向性ノイズを抑圧するビームフォーミング

タブレットでビームフォーミングを行う場合，従来は二つのマイクをディスプレイ側の前面とその背面に配置する構造が多く採用されてきた。しかし，この構造では，前面にマイク穴を開けて前後に時間差を作り出す必要がある。例えば，サンプリング周波数が16 kHzである場合，1サンプルの時間差を作るためには，マイク間の距離を20 mm以上とる必要がある。薄型化が進んでいるタブレット市場では，厚さが10 mm以下の機種が主流となっており，ディスプレイの前後に20 mmの距離を空けてマイクを配置することは難しい。

そこで当社は，二つのマイクを前後に配置するのではなく，タブレットの側面に距離を離して配置する方式を開発した。この方式では横方向から来るノイズに対して時間差を付けて抑圧できる。これにより，ディスプレイ側の前面にマイク穴を開けずに，額縁の狭い美しいデザインを実現できる。更に，タブレットの厚さに関係なくマイク間の距離を横方向にとれることで，高いノイズ抑圧量と自然な音質を両立できる。一方，話者がタブレットの正面から外れ，話者の声が二つのマイクへ到達する時間がずれると，声を方向性ノイズとみなしやすいという課題があった。

この課題を解決するため，独自のビームフォーミング技術[3], [4]を開発した。この技術では，主副のビームフォーマを設け，副ビームフォーマで話者の声の方向を推定してから，この話者方向の情報を利用して主ビームフォーマで方向性ノイズを推定し

（注2）　Androidは，Google Inc.の商標又は登録商標。

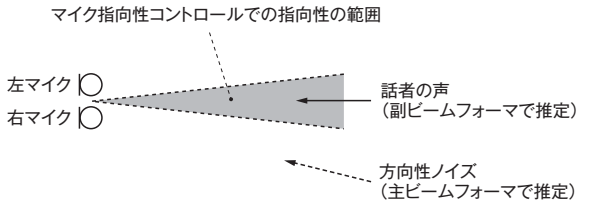

図6. 方向性ノイズを抑圧するビームフォーミング ― 二つのビームフォーマを用いることで，話者の声の方向が変わっても正確にノイズ抑圧できるようにした。

Beamforming to suppress directional noise

て抑圧する（**図6**）。主ビームフォーマでは，副ビームフォーマで推定された話者の方向以外をノイズの方向とみなすことで，話者の声を誤って抑圧することを防ぐ。これによって，指向性範囲内での話者の声を残しつつ，方向性ノイズの正確な抑圧を実現した。このビームフォーミング技術は，“マイク指向性コントロール”の名称でタブレットに搭載した。

4.4 拡散性ノイズを抑圧するノイズキャンセラ

タブレットでのビデオチャットでは，マイクと話者の口との距離が離れ，拡散性ノイズに対して声が小さくなりやすい。このような状態では，ノイズと声を区別することが難しく，ノイズ抑圧を強くするとノイズだけでなく声も抑圧して，抑圧後の音質が不自然になりやすかった。また，周波数帯域が広いと，特に周波数が高い領域で声の成分がノイズに埋もれやすく，処理後に音質がこもりやすかった。

そこで当社は，独自のノイズキャンセラ技術[5]を開発した。この技術では，マイク入力を周波数領域に変換してから声及びノイズ成分を推定し，両者から周波数帯域ごとに抑圧係数を算出して拡散性ノイズを抑圧した後，時間領域に戻して出力する（**図7**）。このノイズ抑圧係数は，声と拡散性ノイズをそれぞれ異なる確率モデルで近似してからその計算式を解析的に最適化している。このノイズキャンセラにより，ハンズフリー環境でも高いノイズ抑圧量と自然な音質を両立させた。

また，ノイズに埋もれやすい高域は中低域の情報も加味して処理することで，明瞭な音質を実現した。このノイズキャンセラは“ノイズ抑圧”の名称で，タブレットに搭載した。

前述のビームフォーミング技術とノイズキャンセラ技術を組み合わせることで，拡散性・方向性ノイズを抑え，従来に比べて格段に高いノイズ抑圧量（当社比で約20 dBの改善）と自然な音質を実現した。

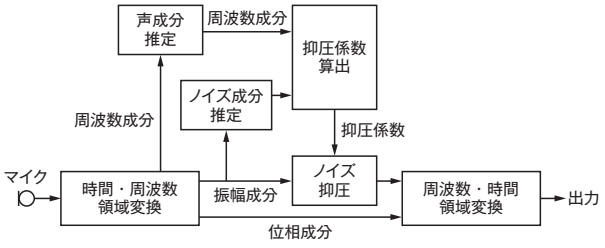

図7. 拡散性ノイズを抑圧するノイズキャンセラ ― 声と拡散性ノイズをそれぞれ異なる確率モデルで近似することで，高いノイズ抑圧量と自然な音質を両立させた。

Noise canceller to suppress diffuse noises

5 あとがき

当社は，音声入力の品質を劣化させる課題を独自技術で克服し，快適な音声入力を実現した。

今後も音声入力を高品質化することで，いっそうの差異化を図り，適用を拡大するとともに，顧客満足度の向上に努めていく。

文 献

(1) 紀伊雅之 他. 携帯機器の高音質化を実現するスピーカアンプLSI TC94B-23WBGと音声信号処理コーデックLSI TC94B24WBG. 東芝レビュー. **67**, 10, 2012, p.21-24.
(2) 須藤 隆 他. “線形エコー抑圧量を考慮したスペクトル選択に基づく非線形エコー抑圧処理（SS-ES）の改善”. 日本音響学会秋季研究発表会講演論文集. 甲府, 2007-09, 日本音響学会. 2007, p.615-618.
(3) 天田 皇 他. 音声認識のためのマイクロホンアレー技術. 東芝レビュー. **59**, 9, 2004, p.42-44.
(4) 永田仁史. 話者追随型2チャネル マイクロホン アレー. 東芝レビュー. **52**, 10, 1997, p.47-50.
(5) 井阪岳彦 他. Laplace分布型確率密度関数と非線形SNR補正に基づく改良型MMSEノイズサプレッサ. 電子情報通信学会技術研究報告. **104**, 631, 2005, p.7-12.

井阪 岳彦 ISAKA Takehiko
デジタルプロダクツ＆サービス社 プラットフォーム＆ソリューション開発センター エンベデッドソフトウェア技術開発部主務。音響信号処理技術の開発に従事。電子情報通信学会，日本音響学会会員。
Platform & Solution Development Center

須藤 隆 SUDO Takashi
デジタルプロダクツ＆サービス社 プラットフォーム＆ソリューション開発センター エンベデッドソフトウェア技術開発部主務。音響信号処理技術の開発に従事。電子情報通信学会，日本音響学会会員。
Platform & Solution Development Center

天田 皇 AMADA Tadashi
デジタルプロダクツ＆サービス社 プラットフォーム＆ソリューション開発センター エンベデッドソフトウェア技術開発部主務。音響信号処理技術の開発に従事。日本音響学会，IEEE会員。
Platform & Solution Development Center