

高音質ノイズサプレッサの研究開発

加藤 正徳 杉山 昭彦 野村 俊之 嶋田 修 芹沢 昌宏 宝珠山 治

NEC 共通基盤ソフトウェア研究所

「ドキュメンタリーのリアリティを作る上では、音声がとても大切だと、有吉さんは言う。極端な話を言えば、映像がなかったとしても、音声がいちちゃんと録れていれば、それでドキュメンタリーは成立するというのである。脳の領域のうちの、約3分の1は視覚にかかわっているとも言われる。視覚は、意識の中であれこれと把握しやすい。一方、音声は、背景に退きやすい。たとえば、ぼんやりと散歩をしている時に、環境から聞こえてくる音は、必ずしも意識されないだろう。しかし、そこに音があるということが、一つのリアリティを立ち上げる上で大切なことも事実である。[1]」

脳科学者・茂木健一郎

1. はじめに

全てのコミュニケーションの基本は、映像ではなく、音声である。これは、人と人のコミュニケーションだけでなく、人と機械のコミュニケーションにおいても変わらない。地理的に離れた地点におけるコミュニケーションに利用される電話、特に携帯電話は、最近15年間で10倍に増加した(図1)。しかも、驚くことに、その増加速度は衰えていない。直近数年は、iPhone 3Gに代表されるスマートフォン(電話と情報端末の兼用端末)が普及してきており、年率10%で市場が拡大している(図2)。2012年の国内市場は370万台、世界市場は23億台と見込まれている。携帯電話以外の例として、インタビューの録音や口述メモに用いられるICレコーダがある。図に示すように、2000年代に入ってから急速に市場が拡大し、年間120万台が出荷されている。世界市場は、年間600万台と言われている(図3)。これらの端末には、一台に最低一つのマイクロホンユニットが使用されている。世界市場は年間40億個であり、未開の地、老若男女を含めて全世界の二人に1人が、マイクロホンを利用していることになる(図4)。特に年率50%で市場が拡大しているMEMS(Micro Electro Mechanical System)と呼ばれる集積デバイスに基づくマイクロホンでこの傾向が強く、今後もその勢いが続くと見られる。

このように広く普及して、我々の生活の必需品となったマイクロホンにおける最も大きな問題は、目的とする信号に伴って混入するノイズや妨害信号である。これは、狭義のコミュニケーション端末だけでなく、カーナビゲーション、通訳端末、さらにロボットにも共通の課題である。マイクロホンにおけるノイズの問題は、その使用環境がこれまでよりも遥かに多様になっていることによる。例えば携帯電話は、街頭や車内、レストラン、駅構内などでも一般的に利用されている。

周囲騒音の混入による通話品質の低下は、特に携帯電話で高まっていた。この問題は第二世代携帯電話の時代から存在しており、通話音声に含まれる雑音成分を抑圧することで音声成分を強調するノイズサプレッサ（NS）が一部の機種に導入されていた。2000年代初頭には、第三世代携帯電話の規格を策定する3GPPにおいてNSの標準化が模索された[2]。しかし、定められた要求条件を全て満足する方式が存在しなかったため、標準方式の代わりに、第三世代携帯電話用NSが満足すべき要求条件とその評価方法を規格化した[3]。

雑音が混入した音声を原音声に近づける技術は、ICレコーダやデジタルビデオカメラからカーナビゲーションや通訳端末などのマイクロホンを備える全ての機器に共通の課題であり、民生機器のデジタル化進展に伴って、この技術に対する期待はますます顕著になっている。また、利用形態や抑圧対象である雑音の特性が多岐にわたるため、様々な機器や環境で利用できる、広範な適応力も商業化の観点からは重要となる。

我々は、上記課題の解決に向けて、優れた雑音抑圧性能と広範な適用性を兼ね備えた高音質ノイズサプレッサを開発した。開発成果の多くは、著名な国際学会、論文誌における発表やプレスリリースを通じて広く公表されており、携帯電話やICレコーダに広く搭載されている。また、中核技術である重み付き雑音推定に関する論文は、2002年電子情報通信学会論文賞を受賞し、その技術レベルは専門家も認めるものである。

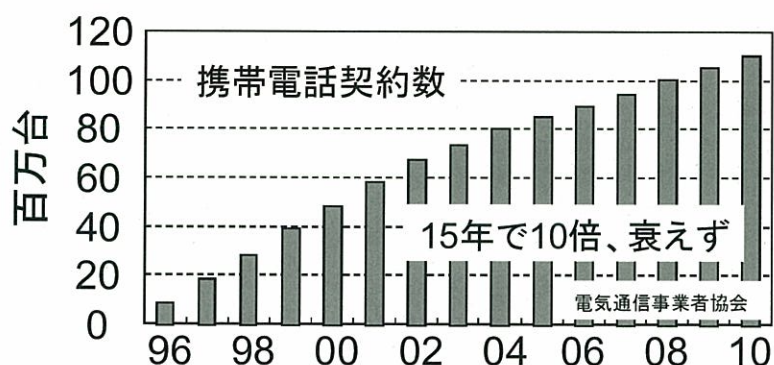


図 1 携帯電話契約者数

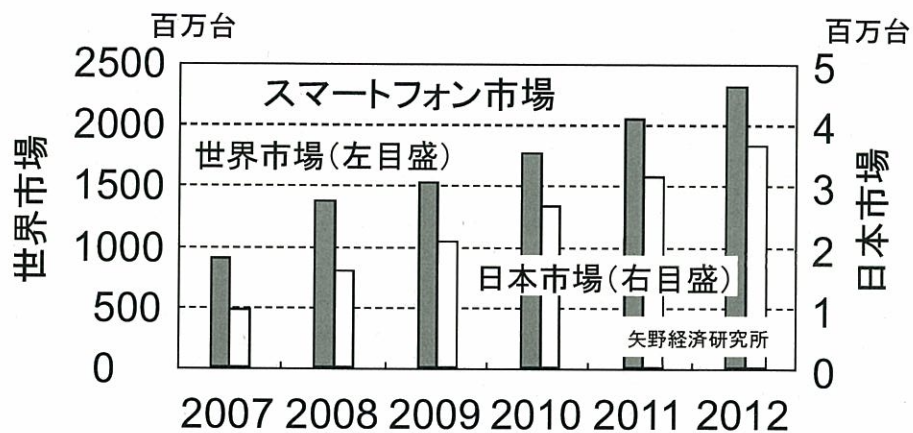


図 2 スマートフォン市場

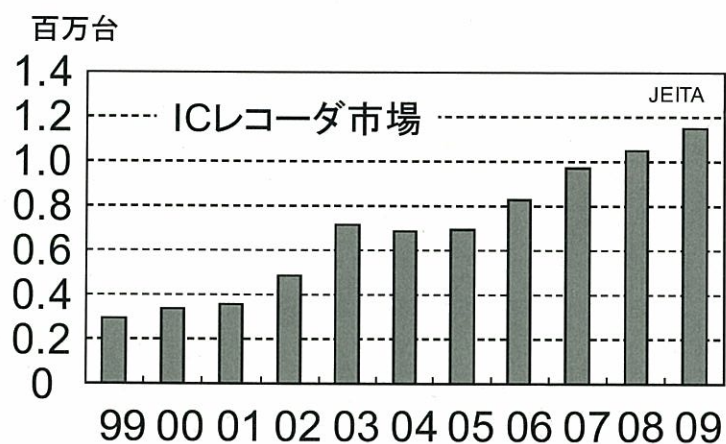


図 3 ICレコーダ市場

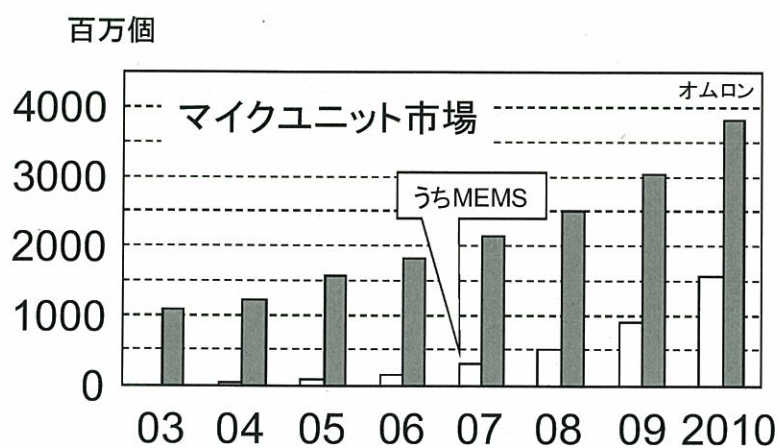


図 4 マイクユニット市場

2. 研究開発の背景

3 G P P の標準化動向にも表れているように、携帯電話向け技術の研究開発が最も急務であり、次の項目が主要な課題となった。1つめは、様々な実環境で生じる雑音に対して高音質（＝高い原音再現性）を達成することである。従来は、非音声区間で事前に求めた推定雑音を入力音声から差し引いて、雑音成分を抑圧する[4]。しかし、実環境における雑音の特性は非定常的なので、非音声区間で推定した雑音の特性が、その後の音声区間で不変とみなすことはできない。従って、誤って音声成分を雑音とみなすことなく、音声区間でも継続的に推定を行う方法が必要である。2つめは、省リソースである。特に、低消費電力型プロセッサを装備する携帯電話に搭載するためには、少ない計算量での動作は重要である。

携帯電話に限らず、様々な機器や環境で利用できるようにするためには、次の項目が主要な課題になった。1つめは、符号化復号後の信号に適用できる雑音抑圧処理の必要性である。従来のNSは、音声送信側または記録側にある音声符号化の直前に導入されており、符号化歪の存在しない状態で十分な雑音抑圧が可能であった。携帯電話の通話相手が雑音抑圧の不十分な端末を利用しているときでも良好な音質を確保するためには、音声受信・再生側の符号化復号信号に適用可能なNSが必要となる。

2つめは、聞き取り易いと感じる抑圧量と音声歪のバランスが、個々人に応じて異なる事実への対応である。一般に、抑圧量を大きくすれば残留雑音は少なくなるものの、歪が目立つ音声になる。我々独自の評価結果から、このバランスは万人に共通ではないことが明らかになった。利用者の更なる満足度向上のためにも、抑圧量と音声歪を利用者自身が好みに応じて制御できる機構が必須である。

3つめは、キーボードのタイプ音などの「衝撃雑音」の抑圧である。衝撃雑音は、マイクとの距離が口元よりも近いことが多いので、聞き取り易さを大きく損ねる要因となりやすい。衝撃雑音の波形形状や周波数特性は、周囲騒音とは異なった特徴を有するため、携帯向けNSでは抑圧することが難しい。その他にも、様々なサンプリングレート・チャンネル数への対応が課題となった。各種条件に対して少ない計算量で高音質を達成しつつ、しかも単体で実現できることが望ましい。

次節以降、上記課題を解決する上で鍵となった技術を中心に、開発したNSの詳細について説明する。

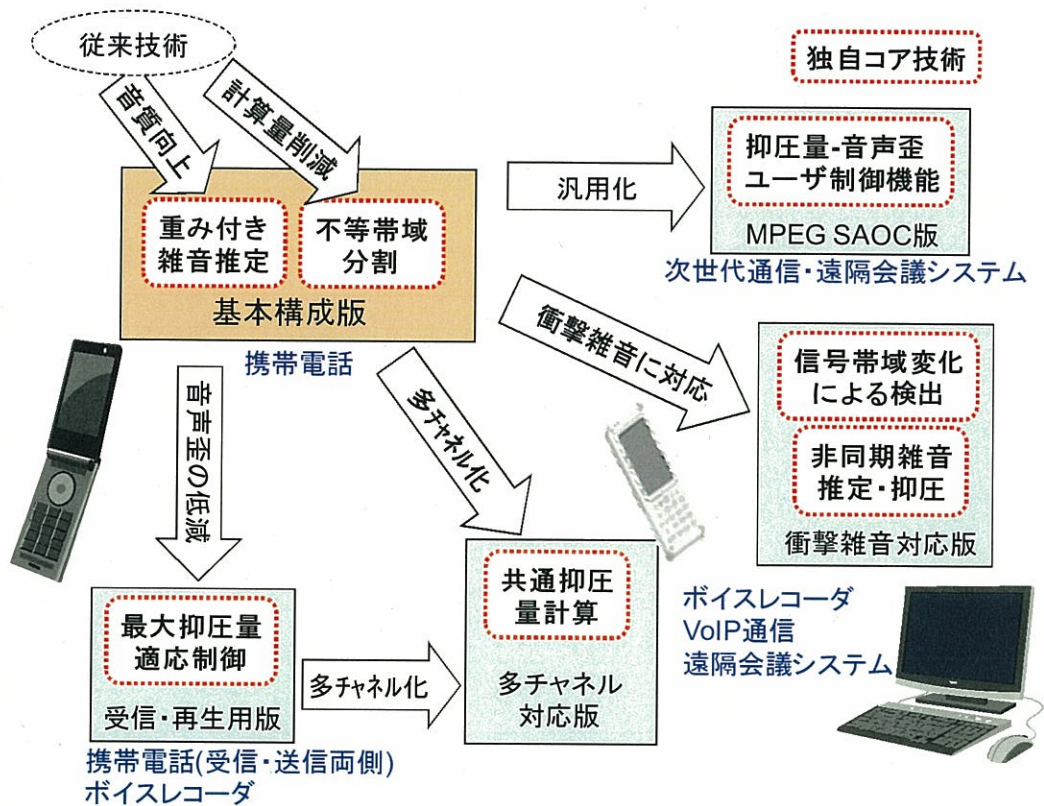


図 5 開発したNS群の全体像

表 1 開発したNS群の内容

分類	NSのバージョン名	主要な独自コア技術	特徴	文中との対応	主な応用先	備考
基本構成	携帯電話向け基本構成版	重み付き雑音推定	多様な実環境における雑音推定精度、及び音質を改善。	4. 2 節	携帯電話	3GPPの公式認証取得(2004年)
		不等帯域分割	計算量を削減。	4. 3 節		
派生版	受信・再生版	最大抑圧量制御	音声歪を低減。音声受信側・再生側への適用を可能にする。	5. 1 節	携帯電話(受信・送信両側)、ボイスレコーダ	3GPPの公式認証取得(2006年)
	MPEG SAOC版	抑圧量-音声歪ユーザ制御機能	抑圧量と音声歪をユーザが自分の好みに合わせて自由に制御できる。	5. 2 節	次世代通信・遠隔会議システム	MPEGのSAOC標準に採択
	衝撃雑音対応版	・信号帯域変化による検出 ・非同期雑音推定・抑圧	「衝撃雑音」の抑圧を可能にする。	5. 3 節	ボイスレコーダ、VoIP通信、遠隔会議システム	「衝撃雑音」：キーボードのタイプ音、ペンで叩く音など
	多チャンネル対応版	共通抑圧量計算	様々なサンプリングレート・チャンネル数の組合せに単体で対応。	5. 4 節		8kHz～48kHzサンプリングに対応

3. 開発したNS群の全体像

図5と表1は、開発したNS群の内容と関係を示している。基本構成は、携帯電話向けの基本版で、重みつき雑音推定と不等帯域分割を導入している。重みつき雑音推定の導入により音質の向上を達成し、計算量を削減する目的で不等帯域分割を開発した。

携帯電話向けに開発した基本構成から、四種類の派生版を開発した。一つ目は、抑圧に伴う音声歪を低減した「受信・再生用版」である。最大抑圧量の適応制御が導入されており、音声受信側・再生側への適用を可能にした。二つ目は、抑圧量と音声歪をユーザが制御できる機能を搭載した「MPEG SAOC 版」である。これによりユーザは各自の好みで抑圧量と音声歪を制御することが可能になった。

三つ目は、衝撃雑音へ対応した「衝撃雑音対応版」である。信号帯域変化による衝撃雑音の検出と、非同期雑音推定及び抑圧が導入されている。基本構成では困難だった衝撃雑音の抑圧を実現した。四つ目は、多チャネルに対応した「多チャネル対応版」である。各チャネルに共通の抑圧量を計算する手法を導入している。計算量を抑えつつ、様々なサンプリングレート・チャネル数に単体で対応できるようになった。なお、多チャネル対応版は、受信・再生用版を多チャネル化したものにも相当する。

4. 携帯電話向けノイズサプレッサの開発

4.1. 携帯電話向けNSの概要

図6は、携帯電話向けNSのブロック図である。入力された雑音混じりの音声信号をFFT（高速フーリエ変換）によって多数の周波数成分に分解し、各成分の大きさと位相を取り出す。出力音声、すなわち雑音抑圧信号の大きさは、入力音声の大きさと計算で求めたスペクトル利得（1より小さい値で、減衰率に相当）の積で与えられる。逆FFTでは、入力音声の位相と雑音抑圧信号の大きさを組み合わせて全周波数成分を合成し、出力音声を求める。つまりNSは、多数の周波数成分に分解された信号（特に、その大きさ）を適切に減衰させることで、雑音抑圧処理を行う。位相は、話の内容を理解するために与える影響が小さいため、入力されたものをそのまま出力計算に使用する。

スペクトル利得とは抑圧度を示す係数のことで、0と1の間の値を用いる。例えば、0.25のスペクトル利得は、入力信号の大きさを1/4にして出力することを表す。つまり、スペクトル利得は、入力音声に含まれる音声成分の割合の推定値であり、入力音声と推定雑音を用いて計算する。従って、雑音推定の精度が、抑圧後の音声品質に大きな影響を与える。開発したNSには、独自技術である重み付き雑音推定[5]が導入されており、高音質を達成している。なお、雑音の推定誤差で生じる過剰抑圧が音声歪を引き起こすことを防ぐために、スペクトル利得の下限値を制限するフロアリング処理を適用している。スペクト

ル利得の計算には、Ephraim らの開発した MMSE STSA 方式[6]を採用した。

雑音の推定とスペクトル利得の算出を、複数の周波数成分を統合した帯域ごとに行うことで、演算量を削減する。帯域の構成方法として、独自技術である不等帯域分割[7]を導入し、同一帯域内で共通のスペクトル利得を複数の周波数成分に適用している。周波数成分の統合をスペクトル利得の計算だけに適用し、入力信号との乗算は全周波数成分に対して行うことと分割パターンが独自技術であり、高音質化の鍵となっている。不等帯域分割とスペクトル利得の共用によって、高い音質を維持したまま NS 全体の演算量を 50%削減した。次節以降、重み付き雑音推定と不等帯域分割の詳細について詳しく説明する。

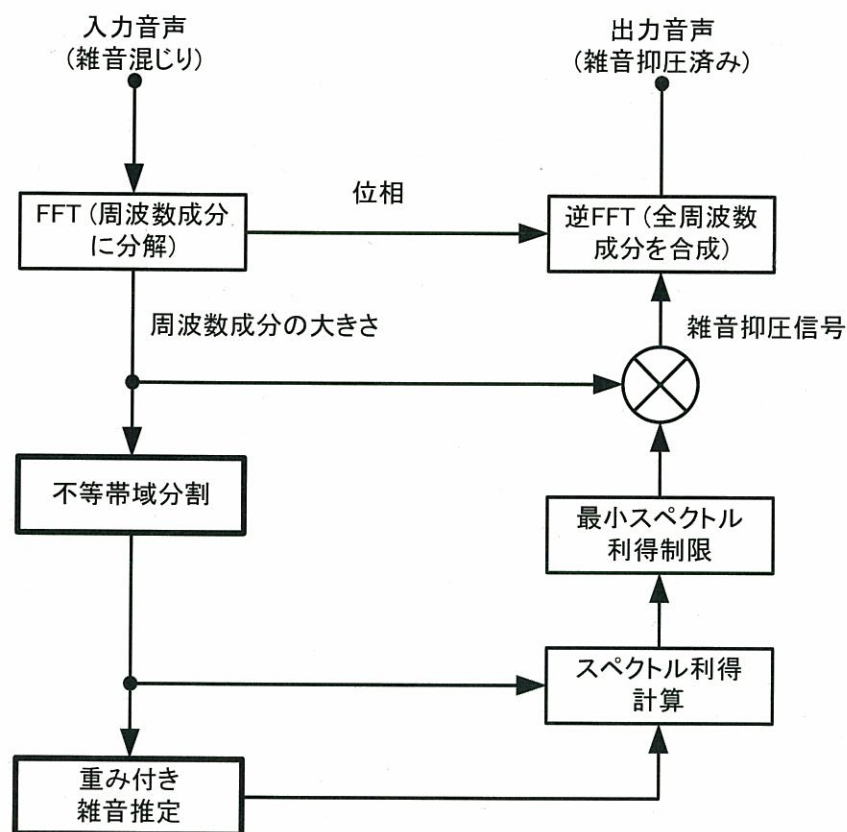


図 6 携帯電話向け NS のブロック図

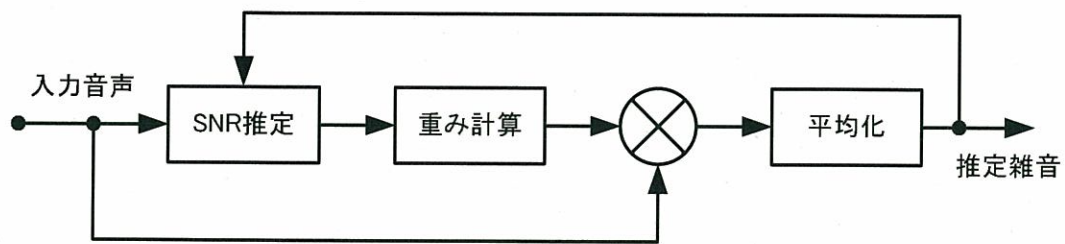


図 7 重みつき雑音推定のブロック図

4.2. 重み付き雑音推定による音質向上

重みつき雑音推定は、様々な非定常雑音源に対して、過大推定を発生させずに高い追従性を達成する。図 7 は、重み付き雑音推定[5]のブロック図である。音声と雑音の混合信号である入力を、SNR で重み付けして雑音を求める。SNR とは「音声対雑音比」(Speech-to-Noise Ratio) のことであり、入力音声と直前に求めた推定雑音の比で近似する。この推定 SNR とほぼ反比例するように設計された非線形関数 (図 8) を

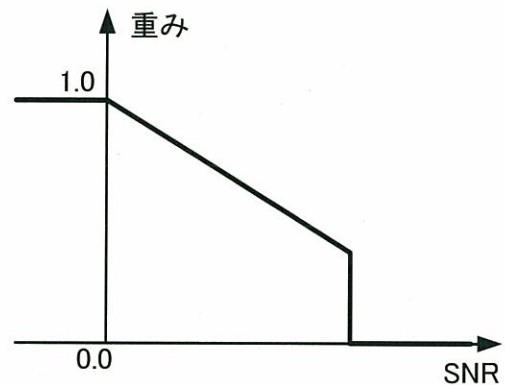


図 8 重み関数

を用いて重みを計算し、入力音声に乗じる。図 8 の効果で、図 9 に示すように音声成分が多く含まれた入力音声に対しては小さな重みづけがなされるため、音声成分の影響をほとんど受けずに雑音推定を継続的に行うことが可能になる。最後に、重みづけされた入力音声の平均値を計算する。図 10 に、レストラン雑音とクリーン音声を混合した信号を対象にした雑音推定の結果を示す。重み付き雑音推定が、他の方式よりも全般的に高い推定精度を達成している。

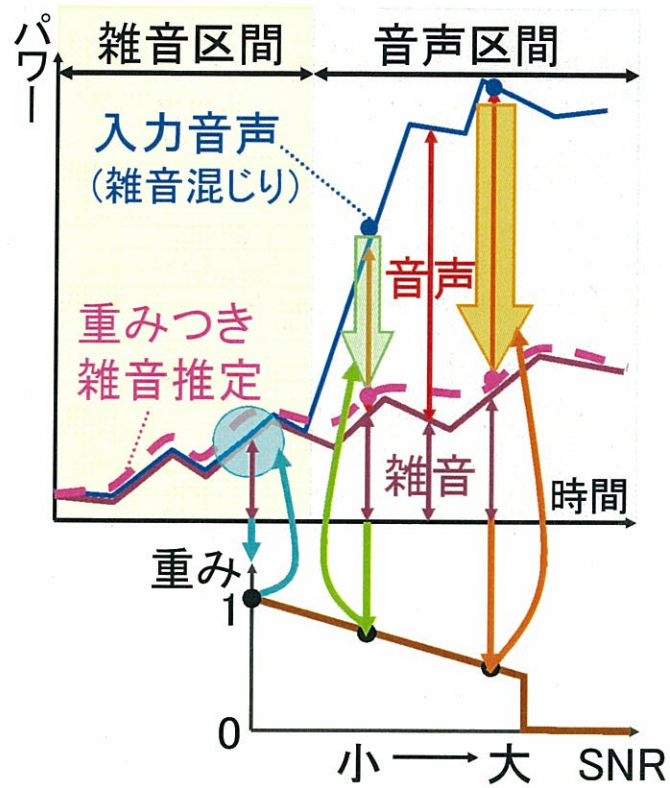


図 9 SNR に応じた入力音声の重みづけ

※推定対象雑音源: レストラン雑音

音声波形(雑音混入前)

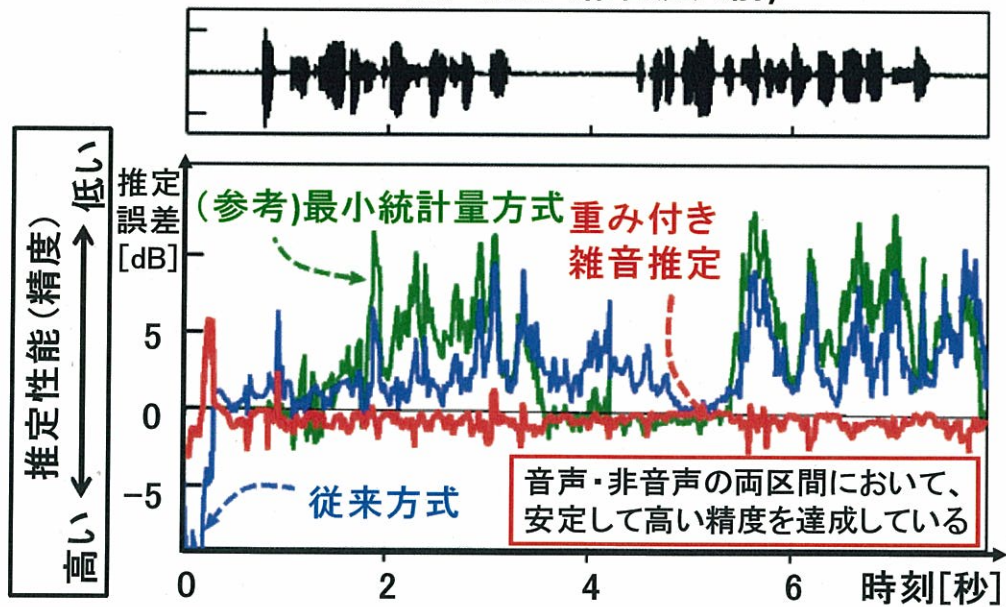


図 10 雑音推定精度の比較実験結果

4.3. 不等帯域分割による計算量削減

出来るだけ細かく帯域を分割して、高い分解能でスペクトル利得を計算したほうが、雑音抑圧後の音声の品質も高くなる。しかし、帯域数が増えるにつれて、演算量が増加する。この課題に対して、心理聴覚の知見に従って、高い周波数では帯域幅を広く、低い周波数では帯域幅を狭く設定することで、音質と演算量の双方を適度に満足することができる。しかし、最も広く知られている「臨界帯域」に従って帯域幅を設定すると、等分割方式よりも音質が劣ることが実験結果から判明した。原因は、低周波数帯域での分解能不足による、スペクトル利得の精度劣化である。そこで、図 11 に示す帯域分割方法を開発した[7]。低域は従来と同等の細かさで帯域分割し、高域では臨界帯域をそのまま採用した。中域では、臨界帯域よりも少し細かく分割して、低域から高域に滑らかに遷移する。この結果、総帯域数は従来の 128 から 32 に削減され、雑音推定とスペクトル利得の計算量を約 75%削減することに成功した。

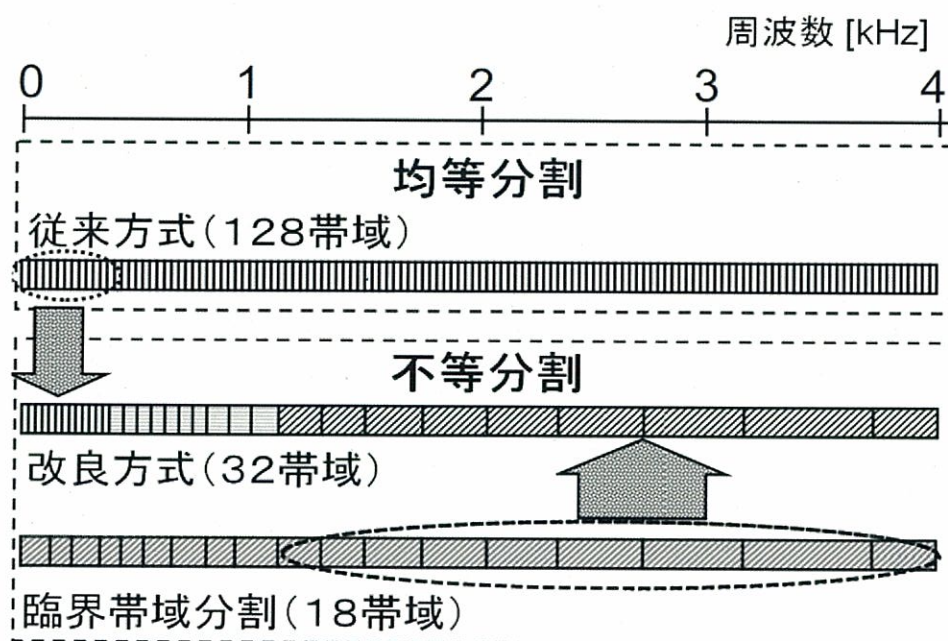


図 11 不等帯域分割

4.4. DSP 実現

DSP（デジタルシグナルプロセッサ）実現には、16 ビット固定小数点 DSP である μ PD77210（NEC 製）と TMS320VC5510（TI 製）を使用した[8]。 μ PD77210 に実装した NS の動作検証は、図 12 に示す評価ボードを用いて行った。8kHz サンプルング信号に対する計算量の測定結果を表 2 に示す。チップの総演算量に対して、1.1% という低演算量で動作することが明らかになった。

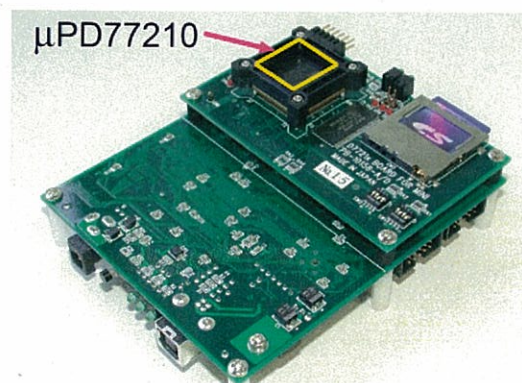


図 12 DSP 評価ボード

表 2 DSP に実装した NS の計算量

	μ PD77210	TMS320VC5510	
FFT サイズ [サンプル]	256	256	512
フレームサイズ [サンプル]	160	160	256
フレームシフト [サンプル]	40	40	256
計算量 [MIPS]	3.2	3.2	1.8

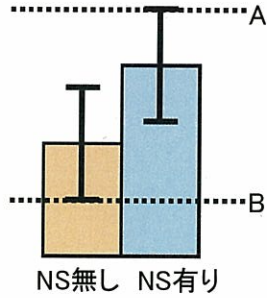
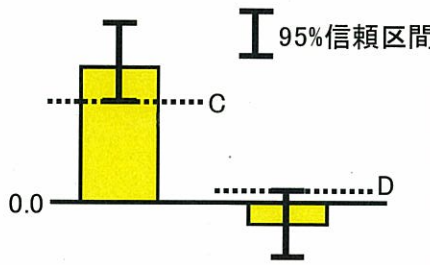
4.5. 3 G P P 規格に基づく評価

開発した NS が、3 G P P の要求条件を満足していることを確認するため、規格書[3]に記載されている全ての項目について評価を実施した[8]。第 3 世代携帯電話用 NS として、AMR 音声符号化[9]と組み合わせた評価が規定されている。主観評価の条件は表 3 に示す通りである。評価用音声「NS 無し」と「NS 有り」は、図 13 のフローに従って作成した。なお、評価項目数が膨大なことから、表 4 に掲載した内容に関する主観評価結果の一部のみを掲載する。

表 3 主観評価条件

言語	日本語、北米英語
雑音源	車内雑音（走行中の自動車内における雑音）
	街頭雑音（市街地の路上を想定）
	バブル雑音（公共の場所における雑音、人の話し声を多く含む）
AMR のビットレート	12.2kbps, 5.9kbps
被験者数	48 名（日本人: 24 名、米国人: 24 名）

表 4 主観評価の内容

評価項目	5段階絶対評	7段階相対評価
評価方式の説明	評価音声の音質の良し悪しを、1～5までの5段階で評価	評価音声を聴き比べ、どちらがどの程度優れているかを、-3～+3までの7段階で評価(0=同等)
評価目的	NSがもたらす音質劣化や、残留雑音に対する悪影響の評価	NS無しと比べて、NS有りが達成する音質改善度の評価
3GPPの要求条件	NS有り $\geq$ NS無し (全条件)	(1) NS有り $>$ NS無し (8/12条件) (2) NS有り $\geq$ NS無し (12/12条件)
評価結果に基づく条件満足の判断目安 (実際には検定を行って判定)	 NS無し NS有り ラインAがラインBを超えていれば条件を満たす。	 95%信頼区間 0.0 ラインCが0.0より大きい → (1)を満足 ラインDが0.0より大きい → (2)を満足

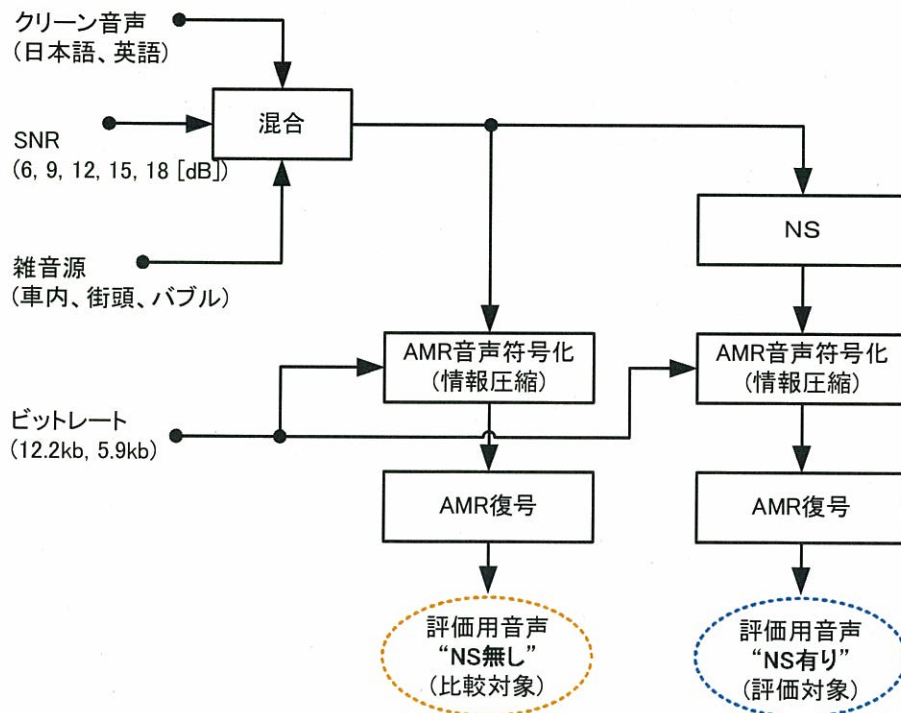
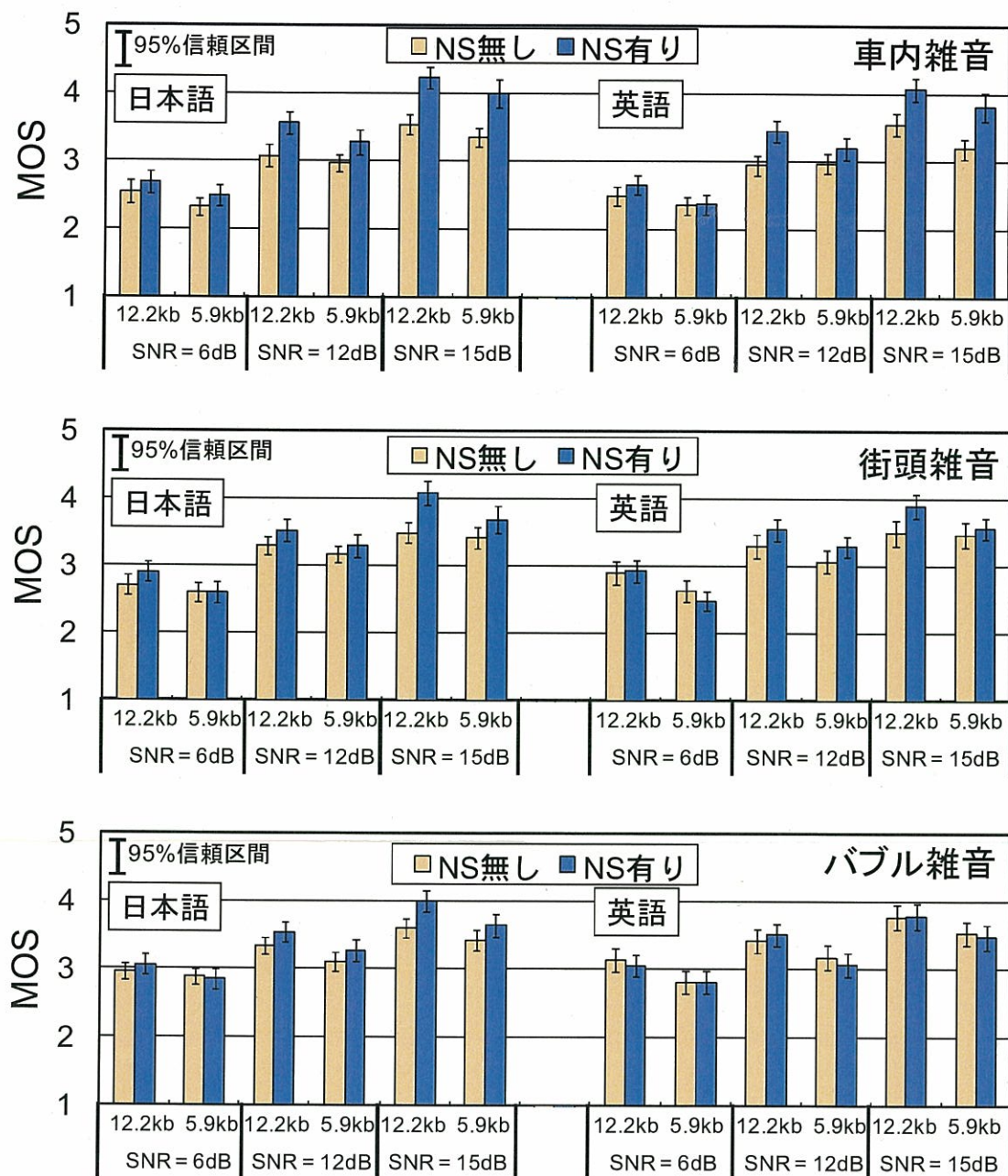


図 13 評価用音声の作成フロー

図 14 は5段階絶対評価、図 15 は7段階相対評価の結果である。表 4 の内容と図 14 の結果を参照すると、全ての条件において、NS有りはNS無しと比べて同等以上であり、要求条件を満足していると言える。また、表 4 の内容と図 15 を参照すると、一条件（車内雑音、SNR=6 dB）を除く全条件で、NS有りの音質がNS無しよりも優れている。また、残りの一条件も同等以上であり、要求条件を満足している。



MOS: 平均オピニオンスコア (Mean Opinion Score)

図 14 5段階絶対評価の結果

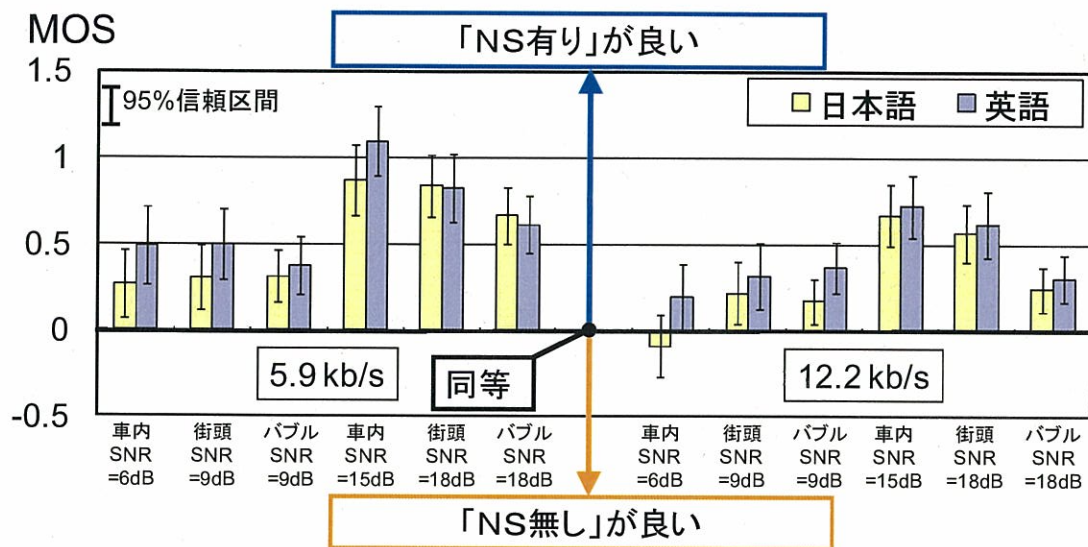


図 15 7段階相対評価の結果

5. 用途の拡大

5.1. 音声受信側への適用

携帯電話の場合、音声符号化復号(情報圧縮伸長)が強調音声に圧縮歪をもたらすため、NSは通常、音声符号化の前処理として導入される。搭載されているNSは端末製造業者毎に異なるため、相手側の雑音抑圧が不十分な場合もある。通話相手の端末に依存せずに、高い受信音声品質を達成するためには、図 16 のように受信側にNSを導入することが有効である。但し、受信側NSの入力には符号化復号による歪が存在するので、歪を考慮していない送信側NSの適用は望ましくない。

同様の議論が、ICレコーダなどの音声録音機器にもあてはまる。マイクの入力音声は、符号化後にメモリに直接録音されて、再生時に復号と雑音抑圧が行われる(図 17)。音声符号化復号後の音声(NS)に入力される状況は、受信側端末と同様である。音声に混入した符号化歪を考慮すると、雑音抑圧に伴う音声歪を最小限に留めつつ、非音声区間の残留雑音を更に抑圧することが課題となる。

そこで、スペクトル利得の最小値を音声区間検出結果と推定SNRに応じて適応的に制御することにより、音声受信側に適したNSを開発した[10](図 18)。このNSについても、3GPP規格に基づく評価によって全ての要求条件を満足することを確認し、公式認証を取得した。これは、同一のNSを携帯電話の送信側にも受信側にも適用可能なことを意味しており、装置設計・製造の簡略化とコスト削減に貢献する。受信側NSは、2006年から、携帯電話に搭載されている。

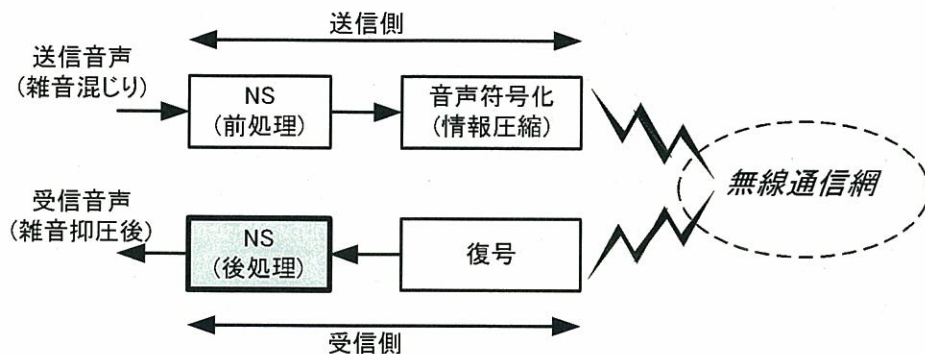


図 16 携帯電話の場合

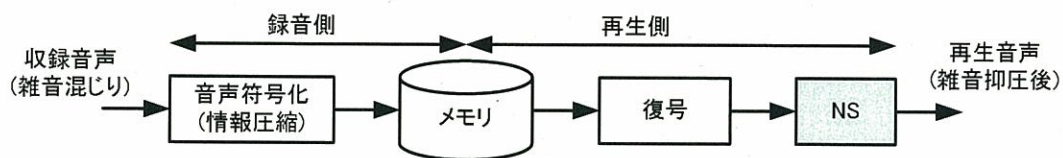


図 17 ICレコーダの場合

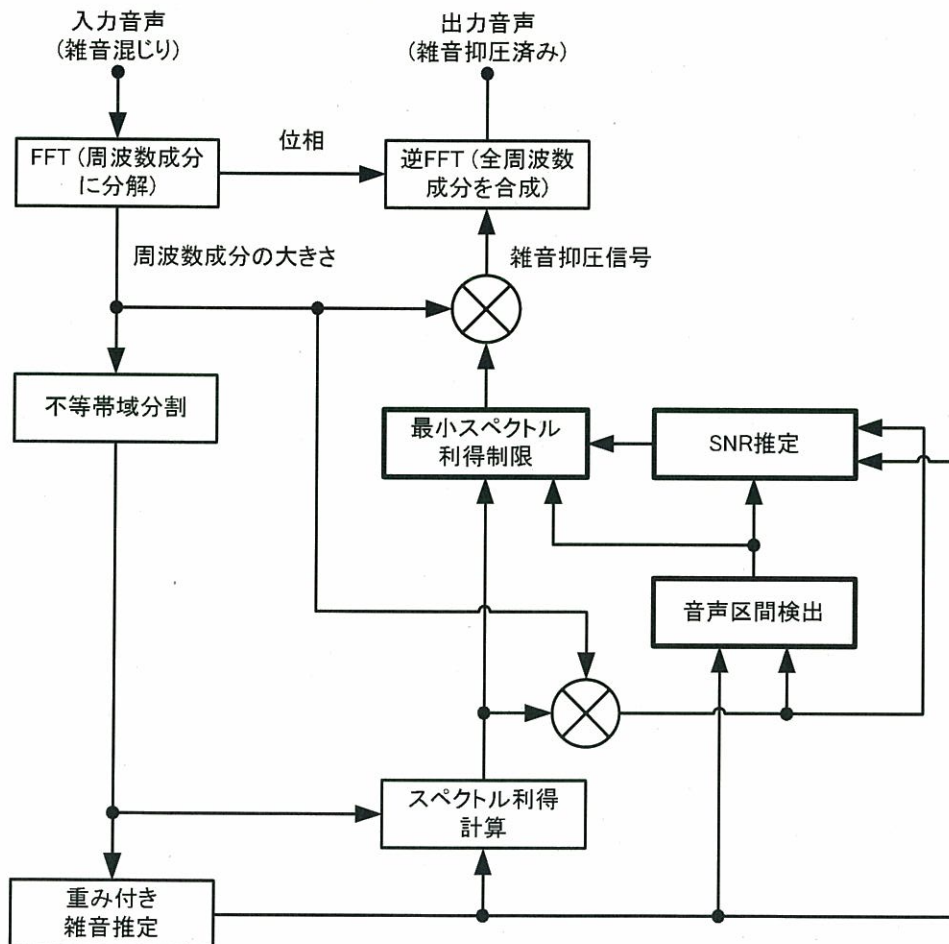


図 18 音声受信側 NS のブロック図

5.2. M P E Gオーディオ SAOC 標準への展開

受信側NSは、送信側NSとは異なり、雑音抑圧後の音声を聞くユーザの端末に装備されている。これは、所望の抑圧度をユーザが調整できることを示唆している。これまでは、好みの抑圧度がユーザに応じて異なるとは考えられてこなかったため、ユーザによる抑圧度制御については検討されなかった。一方、抑圧度の変化が、雑音抑圧処理に伴う音声歪と残留雑音量のバランスに影響を与えること自体は広く知られていた。現行の技術水準では、どのような入力音声に対しても、音声歪をもたらさずに雑音を十分小さくすることは不可能に等しい。一般に、残留雑音を十分少なくする目的で強く抑圧すれば、音声歪が目立つ不自然な音声になる。逆に、抑圧度を小さくして音声歪を低減すると、残留雑音が多くなり、音声の透明感が損なわれる。従って、聞き取りやすいと感じる音声歪と残留雑音のバランスが万人に共通ではないことが明らかになれば、ユーザによるNSの抑圧度調整が大きな価値を持つ。

我々は、M P E Gの最新オーディオ規格である SAOC (Spatial Audio Object Coding) [11]の標準化において、この課題を提起した。そして、音声歪と残留雑音のバランスをユーザが制御することの有効性を検証した[12]。32名の被験者による主観評価を実施し、残留雑音が少ない音声を好むグループと、歪が小さい音声を好むグループが明確に出現することを、世界で初めて明らかにした。この結果は、MPEG SAOCにおいてユーザが音声歪と残留雑音のバランスを制御できる機能の規格化につながり、Separation Meta Dataとして標準に明記された [11]。

5.3. 衝撃雑音の抑圧

P Cを用いた遠隔会議システムを利用する場面や、I Cレコーダを机の上に置いて会議をする際に、会議の様子や内容を音情報で収集・伝達する機会が増えている。そこで大きな問題となるのが、キーボードのタイプ音や、ペンなどの固形物で机を叩く音に代表される「衝撃雑音」の抑圧である。衝撃雑音の波形形状や周波数特性に関する特徴は、周囲騒音とは異なるため、通常のNSでは抑圧が難しい。

このような背景から、衝撃雑音の抑圧に高い効果を発揮するNSを開発した[13]。衝撃雑音対応NSには、信号帯域変化による検出と非同期雑音推定・抑圧が導入されている。衝撃雑音抑圧は、通常雑音抑圧の後処理として動作するために、他の様々なNSへ容易に導入することが可能である。

5.4. 多チャンネル化

高い臨場感を提供するために、様々な商品でステレオを含む多チャンネルシステムと広帯域システムが実用化されている。このような背景に鑑みて、ステレオマイク入力などの多チャンネル入力と多様な信号帯域に対応できるNSを開発した[14,15]。抑圧量の計算を共通化することで、増加したチャンネル数に対して少ない計算量で動作するだけでなく、音像の定位を乱さないという特徴がある。また、8kHz～48kHzまでの全てのサンプリングレートに対応する。8kHz、モノラルの信号に限定されず、様々なサンプリングレートとチャンネル数の組合せに単体で対応できるので、広範な用途に適用することが可能である。ステレオ版多サンプリングレート対応版は、ICレコーダや遠隔会議に搭載されている。

6. 3GPP認証と商用化実績

基本構成及びその一部機能を削除して演算量を低減した構成の双方に関して3GPP標準の定める音質評価を行い、全ての要求条件を満たすことで、2003年に二つの公式認証を取得した。これらのNSは、2004年に当社製第三世代携帯電話に搭載し、世界初の3GPP認証NSの搭載事例となったほか、他社製の携帯電話にも半導体の形で提供された。一方、不等帯域分割によって高音質・低演算量を達成した低演算量版に関しても、2004年に3GPPの公式認証を取得した。さらに、受信・再生側用の低歪版に関しても、2006年に3GPPの公式認証を取得している。低演算量版及び低歪版は、2005年から当社製の第三世代携帯電話に標準搭載されている。このように、開発したNSの4つの異なる方式で3GPPの標準に従った評価を行い、音質の要求条件を満たすことで公式認証を取得してきたことは、NSに関する技術レベルの高さを示すものである。

開発したNSは、当社製の携帯電話以外にも、2006年からオリンパス製ICレコーダの主要機種に搭載されている。JEITAの国内出荷統計と調査会社発表の市場シェアを基に推計すると、NSが搭載された携帯電話は過去5年間で約2500万台、ICレコーダは過去4年間で約200万台になる。前者の国内シェアは約15～20%、後者は38～50%で推移しており、これらの製品市場に大きく貢献している。この他にも、遠隔会議システムへの搭載実績がある。

7. おわりに

多様な騒音環境下における携帯電話の通話品質を向上するため、3GPP規格を満足する高音質ノイズサプレッサを開発した。実環境で生じる様々な雑音を高精度に推定する重み付き雑音推定を考案し、NSの高音質化に成功した。また、独自に開発した不等帯域分割に基づく複数周波数成分に対する共通なスペクトル利得の計算は、音質を損なわずに計算量を75%削減した。開発した4種類のNSは全て、3GPPの公式認証を取得した。現在、当社製の携帯電話に標準で搭載されており、過去5年間の累積出荷は約2500万台に達する。重み付き雑音推定に関する論文[5]は、2002年に電子情報通信学会の英文論文誌に掲載され、翌年には同学会の論文賞を受賞した。さらに、雑音が存在する環境における音声認識の前処理に適用して、70%の環境において音声認識率向上に有効であることを明らかにした[16]。これは、カーナビゲーションや音声通訳端末へ適用した際の有効性を証明するものである。

携帯電話向けに開発したNSは、様々な分野への展開を目指して改良が重ねられて、音声受信側NSや衝撃雑音対応NSなどの開発に繋がった。その一部は、ICレコーダや遠隔会議システムに応用されており、過去4年間に約200万台のICレコーダに搭載された。また、音声歪と残留雑音量をユーザが自分の好みに合わせて自由に制御できる方式を開発し、最新のMPEGオーディオ規格であるSAOC標準に採択された。

音声通信や音声録音の機能を有する機器は、今後も様々な形態で普及していくと考えられる。更なる高音質化を目指しつつ、応用展開を継続したい。そして、エコーキャンセラやノイズキャンセラなどの音響信号技術を統合し、人対人のみならず音声認識やロボット応用を含む人対機械を含めた音声コミュニケーションの快適性向上に貢献したい。

<関連特許> 11件

<参考文献>

1. <http://kenmogi.cocolog-nifty.com/professional/2010/02/post-82e8.html>
2. "Results of the AMR noise suppression selection phase," 3GPP TR 26.978 V4.0.0, Mar. 2001.
3. "Minimum performance requirements for noise suppresser application to the AMR speech encoder," 3GPP TS 06.77 V8.1.1, Apr. 2001.
4. S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-27, no. 2, pp.113-120, Apr. 1979.
5. M. Kato, A. Sugiyama, and M. Serizawa, "Noise Suppression with High Speech

- Quality Based on Weighted Noise Estimation and MMSE STSA," IEICE Trans. Fundamentals, vol. E85-A, No.7, pp.1710-1718, Jul. 2002.
6. Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-32, no. 6, pp.1109-1121, Dec. 1984.
 7. M. Kato and A. Sugiyama, "A Low-Complexity Noise Suppressor with Nonuniform Subbands and a Frequency-Domain Highpass Filter," Proc. ICASSP'06, vol. I, pp.473-476, May 2006.
 8. "Test Results of NEC's New Low Complexity AMR-NS Solution based on TS 26.077," 3GPP Tdoc S4-030601, Sep. 2003.
 9. "Digital cellular telecommunications system (Phase 2+); Adaptive Multi-Rate (AMR); speech processing functions; General description," 3GPP TS 06.71 Release 98.
 10. K. Yamato, A. Sugiyama, M. Kato, "Post-Processing Noise Suppressor with Adaptive Gain-Flooring for Cell-Phone Handsets and IC Recorders," Proc. ICCE'07, 6.1-4, Jan. 2007.
 11. ISO/IEC FDIS 23003-2, Information technology -- MPEG Audio Technologies -- Part 2: Spatial Audio Object Coding (SAOC), 2010.
 12. A. Sugiyama, O. Shimada, T. Nomura and T. Tanaka, "A Subjective Evaluation Method for Two Conflicting Criteria," Technical Report of IEICE, vol. 109, no. 435, SIP2009-126, pp. 17-22, Mar. 2010.
 13. A. Sugiyama, "Single-Channel Impact-Noise Suppression with No Auxiliary Information for Its Detection," Proc. IEEE Workshop on Appl. of Sig. Proc. to Audio and Acoustics (WASPAA), pp.127-130, Oct. 2007.
 14. K. Yamato, A. Sugiyama, M. Kato, "Implementation of a Multipurpose Noise Suppressor Based on a Novel Scalable Framework," Proc. ICASSP'07, pp.337-340, Apr. 2007.
 15. O. Shimada, A. Sugiyama, T. Nomura, "A Low Complexity Noise Suppressor with Hybrid Filterbanks and Adaptive Time-Frequency Tiling," IEICE Trans. Fundamentals, vol. E93-A, No.1, pp.254-260 Jan. 2010.
 16. M. Tsujikawa, M. Kato, A. Sugiyama, M. Serizawa, "Evaluation of a Noise Suppressor Based on Weighted Noise Estimation Combined with a Speech Recognition System," Proc. IEICE DSP Symp., pp. 99-104, Nov. 2001.