音響分析の概念と実際

細川 清人^{1, 2)}・岩橋 利彦²⁾・小川 真²⁾
加藤 智絵里²⁾・猪原 秀典²⁾

The Principles and Practice of Acoustic Analyses

Kiyohito Hosokawa^{1, 2)}, Toshihiko Iwahashi²⁾, Makoto Ogawa²⁾
Chieri Kato²⁾ and Hidenori Inohara²⁾

In general, the evaluation of the degree of abnormal voice quality has a crucial role in the clinical assessment of voice-disordered patients. As the gold standard for the assessment of voice quality, auditory-perceptual methods have been utilized worldwide. However, the subjective nature of these methods can strongly influence the reliability and accuracy of the evaluation. Due to such limitations, acoustic measurements have been developed to increase the objectivity in the evaluation of voices.

In this article, we first review the traditional acoustic measures to improve our understanding of the principles of acoustic analyses. Subsequently, for clinical practice, we introduce several available instruments, such as analysis software programs and microphones, that are suitable for acoustic analyses. Lastly, we describe the differences between two software programs and two microphones, and discuss their respective utility.

Key words : auditory-perceptual judgement, acoustic analysis, microphone, Multi-Dimensional Voice Program, Praat

はじめに

音声障害患者の病状評価において、患者の音声の質の良悪は最も重要視される評価の一つと言える。音声の質の良悪について認識するのは人間の感覚である以上¹⁾、音声の質の評価法のゴールドスタンダードがGRBAS尺度^{2~4)}に代表される聴覚心理的評価であるということに異論は無い。しかしながら、聴覚心理的評価の評価値は評価者の感覚に依存することから、科学において最も重要視される客観性を確保することに限界があることも確かである。そこで、人間の声の波形、すなわち音声波形の形状の解析により音声の性質を客観的な指標として評価するため、様々な音響分析的手法が開発されてきた^{5~12)}。

本稿ではまず、ゴールドスタンダードである聴覚心理的評価について概説した後に、客観的指標である音響分析の理解に必要な背景、すなわち音声波形の性質について確認し、音響分析パラメータが示す事象について解説する。その後、音響分析を実際に施行するために必要なマイクロフォンなどの機材や録音環境および分析ソフトウェアについて紹介し、最後にマイクロフォンと分析ソフトウェアの相違による音響分析パラメータの実値および診断精度の相違について提示する。

聴覚心理的評価

聴覚心理的評価とは、患者の音声を評価者が聴取しその性質について何らかの判断を与える評価法である⁴⁾。当評価法の確立において一色ら^{2, 3)}は、音声の質の聴覚心理的評価における主観的および非普遍的な性質を極力排除するために、意味微分法による因子分析を用いて嗄声の質を4つの因子に分類した。そして、それをもとにGRBAS尺度が作成され⁴⁾、現在全世界的に使用されている^{13, 14)}。

GRBAS尺度は、患者の音声の性質について、各性質(Grade: 嗄声度, Roughness: 粗糙性, Breathiness: 氣息性, Asthenicity: 無力性, Strain: 努力性)を0(正常)から3(重度)までの4段階で評価する方法であり⁴⁾、評価において一定の信頼性が認められている¹⁴⁾。しかしながら、GRBAS尺度においても、異なる評価者が評価した場合のみならず、同一の評価者が複数回評価した場合にも評価値が一致しないことが稀ではない^{15, 16)}。聴覚心理的評価の信頼性には評価者の経験年数や専門性など様々な要因が影響すると報告されるが^{17~19)}、実際に信頼性の高い評価を行うためには、一定の基準に沿った評価を可能とするための訓練を行う必要がある^{20~22)}。訓練には市販されているDVDソフト「動画で見る音声障害」や、同梱のCD「嗄声

1) 大阪警察病院耳鼻咽喉科

2) 大阪大学大学院医学系研究科耳鼻咽喉科・頭頸部外科学教室

1) Osaka Police Hospital, Department of Otorhinolaryngology

2) Osaka University Graduate School of Medicine, Department of Otorhinolaryngology-Head and Neck Surgery

図1 当施設において聴覚心理的評価に用いられているツールであるGRBAS checkerのインターフェイス。

のサンプルテープ」を用いることができる。また、さらに評価の信頼性を高めるため、評価の際に参照用音声を聴取したり^{20~22)}、詳細な評価基準を明文化する²²⁾などの試みがなされている。我々の施設では、Microsoft Excel Visual Basic for Applicationsで作成されたユーザーフォーム (GRBAS checker²³⁾) を用いて、リアルタイムに参照用音声を聴取しながら評価するシステムを採用している (図1)。

音響分析の概念

前項では、主観的評価である聴覚心理的評価をより客観的に記述するための工夫について述べた。本項では、聴覚心理的評価以外の方法で音声の質の評価を客観的に行う方法、すなわち音響分析について述べる。

音響分析の概念を理解するためには、音の波形の性質について知る必要がある。音とは物体の振動が媒体 (通常は

空気) の分子密度に粗密を生じさせ、それが媒体を伝播する現象を捉えたものであり、媒体分子の粗密を波形として表示したものが音の波形である。最も単純な音 (純音) の波形は同じ形状の正弦曲線の繰り返しパターン (ここでは便宜上「サイクル」と呼ぶ) からなる (図2)。この波形を規定する特徴量には、1サイクルの波形の時間長である「周期」および波形の大きさ (粗密差) である「振幅」が挙げられるが、それぞれ音の高さおよび大きさに直接的に関連している。なお、「周期」の逆数を「周波数」と呼び、1秒間に含まれるサイクル数を示す。人間は周波数が大きい音ほど高い音と感じ、振幅が大きい音ほど大きい音とを感じる。

人間の声である音声も上記の性質を持つが、その波形はより複雑な形状をしている (図3)。この複雑な波形は、様々な周波数や振幅の正弦曲線の総和 (フーリエ級数) によって表現できることが判明している。図3に正常例と音声障害例の音声波形をそれぞれ示す。まず、両波形を観察すると、いずれも純音ほど均一ではないものの準周期的なサイクルを持っていることがわかる。この1サイクルの時間間隔を「基本周期」その逆数を「基本周波数 (F0)」と呼び、声の高さについての重要な指標となるとともに、音声障害の程度を評価する後述のパラメータの算出に必須となる。

次に正常例と音声障害例の波形を比較すると、音声障害例では波形のサイクル間の相同性の乱れが大きいこと、および形状がスムーズではないことがわかる (図3)。これらを音声波形の特徴量である周期および振幅の乱れとして表現した計測値がそれぞれ jitter および shimmer と呼ばれるパラメータである。これらの計算方法について jitter を例に挙げて解説すると (図4)、jitter とは隣り合うサイクルの周期との差を分析範囲内で平均し、全体の周期の平均値で除算したものとイえる。また、周期の代わりに振幅を用いて同様の計算を行うことにより shimmer が算出される。これらの計算値は、短時間での周期・振幅の変動を数値化

図2 上段：黒色実線は200Hzの純音の波形。太線は1サイクルの波形を示す。
下段：灰色実線は600Hzの純音、灰色点線は200Hzで0.4倍の振幅の純音。

したパラメータであるが、長時間での周期・振幅の変動（トレンド）の影響を少なからず受けることから、これらのパラメータからトレンドの影響を減弱させる計算式を追加したperiod perturbation quotient (PPQ) およびamplitude perturbation quotient (APQ) がKoikeらにより開発され^{6,7)}、現在広く使用されている。また、波形のスムーズさについては、音声波形中の雑音成分のエネルギーを計測するharmonics-to-noise ratio (HNR) やnormalized noise energy (NNE) がそれぞれYumoto^{8,9)} およびKasuya¹⁰⁾ により開発され広く使用されている。これらの音響分析パラメータと聴覚心理的評価の評価値には相関関係が認められており^{7,24~28)}、音声障害の程度や治療効果を表現するのに用いられるほか^{29~31)}、音声障害の有無についての診

断にも用いられている^{32~34)}。

音響分析の実際

音響分析パラメータの算出は、音声波形の電子データを記録したコンピュータを用いて行われる。しかしながら、患者の音声はコンピュータ内にデータとして存在するためには、様々な機器を媒介する必要がある、正確な分析を行うためには極力正確な音声の記録を必要とする。以下に、音響分析を実際に行うにあたり必要となる環境および機器について解説する。

まず、人間の聴覚認識能力とは異なり、コンピュータは周囲の雑音と人間の声を区別することができない。よって、録音に際しては極力雑音が無い環境で行う必要がある

正常例

音声障害例

図3 上段：正常音声の音声波形。周期・振幅に乱れが少なく、波形が平滑である。
下段：音声障害例の音声波形。周期・振幅に乱れがあり、波形が粗造である。

$$\begin{aligned} \text{jitter} &= \frac{\frac{1}{N-1} \sum_{k=1}^{N-1} |p_k - p_{k+1}|}{\frac{1}{N} \sum_{k=1}^N p_k} = \frac{(\text{隣り合うサイクルの周期の差}) = \Delta p_k}{(p \text{ の平均})} \\ &= \frac{\frac{1}{N-1} \sum_{k=1}^{N-1} \Delta p_k}{\frac{1}{N} \sum_{k=1}^N p_k} = \frac{(\Delta p \text{ の平均})}{(p \text{ の平均})} \end{aligned}$$

図4 jitterの計算式。分子は隣り合うサイクル間の周期の差の平均値を示し、分母は周期全体の平均値を示している。

ため、耳鼻咽喉科外来で施行する場合は聴力検査室での録音が望ましい。その際には、室内の空調や検査機器による雑音に注意を払う必要がある。また、記録された波形からの観点では、信頼性の高いF0の算出には波形の信号雑音比が15dB以上、jitterやshimmerの算出には30dB以上が必要とされており³⁵⁾、それを満たす環境であれば静かな室内(周辺雑音<50dB(C)^{36,37)}での録音も許容されうる。

次に、音声を記録する機器について解説する。まず音声を波形として記録するにはマイクロフォンが必要である。現在、様々なメーカーから様々な用途のマイクロフォンが市販されているが、音響分析を行う際には、元の音声の音圧の変化を歪みなく忠実に記録できる機能が最も重要である³⁸⁾。適切なマイクロフォンを選択するには、マイクロフォンの特性についての知識が必要となるが、ここでは特に周波数特性および指向性について言及する。前述のように音声波形は、様々な周波数や振幅の純音波形の総和によって構成されることから、音響分析を行う際には、音声波形を構成する各純音波形が適切に録音される必要がある。マイクロフォンの周波数特性とは低音から高音までの周波数を一定の音圧でマイクロフォンに入力した場合に出力される音圧が、周波数によってどの程度異なるかについての特性のことであり、一般に横軸を周波数、縦軸を振幅とするグラフにより表現される(図5)。歪みなく忠実に音声を記録するためには周波数特性が平坦なマイクロフォンの使用が望ましい。また、マイクロフォンの信号変換方式の違いによりダイナミック型またはコンデンサ型(またはその他の型)という分類がある。一般にコンデンサ型マイクロフォンの方が周波数特性が平坦であり、音響分析を目的とする録音にはより適切とされる³⁸⁾。

次に、マイクロフォンの指向性について述べる。マイク

ロフォンにおける指向性とは、音源(ここでは患者の口唇)とマイクロフォンとの角度により音声信号の変換強度がどの程度異なるかを表現する指標である(図5)。無指向性マイクロフォンではどの角度から入力された音圧も均等な強度で変換される。一方、単一指向性(カーディオイド)に代表される指向性マイクロフォンでは、マイクロフォンの前方からの強度が大きく他方向からの強度が小さくなるよう調整されている。単一指向性の方が、周囲の雑音の入力を低減させることができるため高い信号雑音比が得られ音響分析には有利となるが、音源とマイクロフォンとの距離の長短により低音域の周波数の変換強度が変化する(近接効果)³⁸⁾ため³⁸⁾、周波数特性の平坦さにおいては無指向性マイクロフォンの方が有利ともいえる。音響分析に必要なとされるマイクロフォンの機能についてはŠvec and Granqvistの総説³⁸⁾に詳述されているが、検査の目的や録音環境を考慮し選択すべきと考える。

マイクロフォンに入力された音圧は電気信号へと変換され、録音装置へ入力される。過去にはアナログカセットテープやDAT(Digital Audio Tape)が用いられたが、音響分析はコンピュータを用いて行われるためコンピュータで可読なフォーマットを用いて音声ファイルとして記録するほうが簡便である。また、アナログカセットテープの使用は雑音の混入の問題からも推奨されない。音声ファイルのフォーマットにはリニアPCM方式などの無圧縮の音声ファイル(.wavファイルなど)を用いる。オーディオプレーヤーで頻用されるmp3などの非可逆圧縮方式のフォーマットを用いた場合、高音域の周波数が除去されてしまうため分析結果に影響を及ぼすことから推奨されない^{39,40)}。コンピュータに直接録音する場合は、コンピュータ付属のサウンドカードの特性およびコンピュータのファ

図5 マイクロフォンの主な仕様。無指向性のマイクロフォンでは周波数特性が均等であり、単一指向性などの指向性を持つマイクロフォンでは近接効果が認められるが、指向性を持つ方が、周囲の雑音の混入を低減させることができる。

ンによる雑音に注意を払う必要がある。なお、音響分析ソフトウェアが同梱された音響分析システム (CSL4300: KayPentax, USA) により適切に音声波形を録音することが可能であるが高価である。現在では、リニアPCMレコーダーが比較的安価に市販されており、取り扱いも容易である。いずれの機器においてもデジタル録音が行われるが、サンプリング周波数や量子化ビット数については通常の音楽CDと同等の品質 (サンプリング周波数 44100Hz, 量子化ビット数 16bit) で差し支えない³⁶⁾。

次に、実際に分析を行う対象となる音声サンプルに用いられるタスクについて述べる。分析の対象となる発声タスクについては、検査の目的により様々なタスクが想定されるが、本稿で紹介した音響分析パラメータの計算過程において、正常例の持続母音では波形の形状の均一性が高く、音声障害例では均一性が低いという性質を利用しているため、通常は持続母音発声タスクを分析対象とする。一方、文章音読や会話音声の波形は様々な種類の音素やイントネーションの変化を内包するため、正常者においても波形の均一性が低いことから、当該の音響分析パラメータによる音声障害の程度の推定が困難となる。なお、持続母音サンプルにおいても、発声開始時と終了時には波形の周期や振幅が乱れるため、分析においては起始部と終止部を除外して、波形が安定した中間の部分を用いる。その分析範囲には少なくとも100サイクル以上の周期が含まれていることが望ましいとされる³⁶⁾。すなわち、100Hzの男性の声なら1秒以上の分析範囲が必要となる。また、実際の分析の際には音声波形およびF0の推移曲線を目視し、分析の適切性を確認する必要がある。特にF0の推移において、突然にF0が半減または倍増する場合には、ソフトウェアが複数周期分の波形を1周期として誤認していることが多いため、ソフトウェアのF0抽出範囲の設定項目を調整すべきである⁴¹⁾。なお、これらのパラメータの算出はF0の正確な抽出に基づくため、重度の音声障害患者の音声波形においてはF0の抽出が困難となりパラメータの信頼性が低下する³⁶⁾。それゆえ、重度の音声障害におけるパラメータの数値の評価には注意が必要である。

文章音読サンプルに対する音声障害の程度の評価については、1994年および1996年にHillenbrandらがcepstral peak prominence (CPP) およびsmoothed CPP (CPPS) について報告した^{11, 12)}。CPPおよびCPPSの算出はF0の正確な抽出に依存しないため、文章音読サンプルにおける音声障害の程度を比較的高精度に評価することができることが報告されている^{11, 12)}。また、CPPSを含めた複数の音響分析パラメータの回帰式により、文章音読サンプルに対する音声障害の定量的評価が試みられており、Awanらはcepstral-spectral index of dysphonia⁴²⁾の有用性を、MarynおよびBarstiesらはacoustic voice quality index^{23, 43, 44)}の有用性を報告している。持続母音とは異なり、文章音読については使用言語の違いにより評価の信頼性が変化する可

①持続母音発声

いー———
えー———
あー———
おー———
うー———

②文章音読

「北風と太陽」
ある日、北風と太陽が、
力くらべをしました。
旅人の外套を脱がせたほうが
勝ちということに決めて、
まず、風から始めました。

*自然な大きさ、高さの声で発声する。

図6 当施設で用いている音声録音用の発声タスク。

能性があり、使用言語における評価の信頼性を確認する必要がある²³⁾。

上記のことから、音響分析における音声録音においては持続母音サンプルの録音が必須となるが、文章音読サンプルのほうが患者の日常の音声使用状況により近いサンプルであることから^{23, 37, 43, 44)}、文章音読サンプルの録音を併施する方が望ましい。図6に著者の施設で用いている発声タスクを示す。

録音された音声波形を解析する音響分析ソフトウェアについては、本邦ではMulti-Dimensional Voice Program (MDVP, KayPentax, USA) が有償のソフトウェアながら頻繁に用いられている。一方、近年ではアムステルダム大学のBoersmaらが開発したPraat⁴⁵⁾であれば無償で利用することができるほか、ウィスコンシン大学のMilenkovicらが開発したTF32⁴⁶⁾は機能制限版であれば無償で利用することができる。本邦では音響分析ソフトウェアVA⁴⁷⁾についての報告がある。

前項で述べたように、音響分析を行う際にはまず音声波形のF0を抽出する必要があるが、ソフトウェアにより異なった方法が採用されている。簡単に説明すると、MDVPでは波形が1サイクル中にピークに達する時点を算出してピーク間の時間間隔を元に抽出する (極大値検出法)⁴⁸⁾。一方、PraatやTF32では1~2サイクルに相当する波形を時間軸に平行移動させて次のサイクルと最も一致する時点を元に抽出する (波形整合法)⁴⁹⁾。また、VAでは1サイクル中に波形が時間軸と交差する時点を補間を用いて推定する方法 (零交差法) により抽出する⁴⁷⁾。このような違いから、同一名称のパラメータであったとしても使用するソフトウェアにより数値が異なる。どの方法がより適切かについては議論がある^{28, 50~56)}。

マイクログフォンおよびソフトウェアの違いによる音響分析パラメータ数値の変動について

マイクログフォンには様々な種類があり用途に応じて選択する必要があること、また音響分析ソフトウェアの種類によりパラメータの数値が異なることを述べた。音声波形の録音について、Winholtz and Titze⁵⁷⁾はヘッドセット型の

コンデンサ型マイクロフォンの使用を推奨しているが、ダイナミック型マイクロフォンを用いた報告も散見される^{58~60)}。また本邦ではMDVPを用いて音響分析を行った報告は多数存在するが^{59, 61~64)}、Praatを用いた報告は少数である^{65~67)}。そこで、本項では複数のマイクロフォンおよび音響分析ソフトウェアを用いた際のパラメータの数値および診断精度の違いについて検証した結果を示す。

対象と方法

音声障害症例95例（男性28例・女性67例、平均年齢 57 ± 14 歳）および正常例42例（男性12例・女性30例、平均年齢 55 ± 12 歳）の計137例において音声録音を行い、得られた158サンプル（音声障害例治療後の21サンプルを含む）を分析対象とした。

自然な声の高さおよび大きさでの持続母音発声 /i:/, /e:/, /a:/, /o:/, /u:/ および「北風と太陽」の第一および第二文の文章音読（図6）を録音した。録音にはヘッドセット型エレクトレットコンデンサ型マイクロフォン（SE50, Samson, USA）およびボーカル用ダイナミック型マイクロフォン（PRO-300, Audio-Technica, Japan）を使用した（図7）。前者を左口角から約3cm外側の位置に固定し、後者を口角から約20cm・体軸と約45度の位置に検者が用手固定した。両マイクロフォンに入力した音声信号をリニアPCMレコーダ（H4n, Zoom, Japan）にサンプリング周波数44100Hzおよび量子化ビット数16bitでステレオ同時録音した。

音響分析パラメータの算出には、持続母音発声 /a:/ の中間の3秒間を用いた。音響分析ソフトウェアMDVPおよびPraatを用いてF0, PPQ, APQ, およびHNRを算出した。なお、MDVPにおいては $HNR = 10 \times \log_{10} (1/NHR)$

の計算式によりNHRからHNRへと換算した。また分析の際には、検者が波形の形状とF0の推移曲線を目視し、必要時に抽出周波数範囲の調整を行った上でパラメータの算出を行った。上記4種のパラメータのいずれかが算出不能であった4例を除外し、計154サンプルを用いて以下の検討を行った。

なお、聴覚心理的評価には過去の研究³⁴⁾における、5人の評価者によるGRBAS尺度のGスコアの平均値を用いた。今回の対象サンプルにおける評価者内信頼性はCohen's $\kappa = 0.662$ 、評価者間信頼性はFleiss's $\kappa = 0.352$ であり、以降の分析を行うことが可能と判断された⁶⁸⁾。

両マイクロフォンにおける各パラメータのソフトウェア間の一致性および相関性について級内相関係数 (ICC: intra-class correlation) およびスピアマンの順位相関係数 (r_s) を用いて評価した。また、各パラメータと平均Gスコアとの相関性を r_s を用いて評価した。さらに音声障害の有無についての診断精度を、平均Gスコア < 0.5 を正常音声と定義した場合における receiver operating characteristic 解析の曲線下面積 (AUC値: area under the curve) を用いて評価した。

結 果

まず、ソフトウェア間における両マイクロフォンのパラメータの差異および一致性について検討した（表1）。F0については、両マイクロフォンにおいてほぼ差異を認めず、ほぼ完全な一致性（共にICC = 1.00）を認めた。PPQについては、PPQ_{MDVP}がPPQ_{Praat}と比較して2倍弱の値となった。ソフトウェア間の一致性はエレクトレットコンデンサ型マイクロフォンではICC = 0.70、ダイナミック型マイクロフォンではICC = 0.59であり、いずれも中等度

図7 検討に用いたマイクロフォンの仕様。(PRO-300の指向性は非公開であるが、メーカーより上記の指向性と同等と回答あり。)

表1 マイクロフォンおよびソフトウェア別の、各音響分析パラメータの平均値および標準偏差

		Praat (平均 ± 標準偏差)			MDVP (平均 ± 標準偏差)			ICC
F0 (Hz)	コンデンサ	200.667 ± 55.307			201.639 ± 53.695			1.00
	ダイナミック	201.344 ± 54.899			201.698 ± 54.020			1.00
平均Gスコア		G (< 0.5)	G (0.5-1.5)	G (1.5-3.0)	G (< 0.5)	G (0.5-1.5)	G (1.5-3.0)	
PPQ (%)	コンデンサ	0.20 ± 0.13	0.65 ± 1.15	1.36 ± 1.11	0.37 ± 0.27	1.06 ± 0.96	2.48 ± 1.26	0.70
	ダイナミック	0.19 ± 0.12	0.59 ± 0.91	1.13 ± 0.74	0.45 ± 0.30	1.41 ± 1.84	2.60 ± 1.39	0.59
APQ (%)	コンデンサ	1.82 ± 0.82	4.13 ± 2.90	8.56 ± 4.54	2.00 ± 0.86	4.22 ± 2.80	8.34 ± 4.33	0.95
	ダイナミック	2.16 ± 0.96	5.05 ± 2.92	9.93 ± 5.62	2.60 ± 1.14	6.04 ± 2.68	10.39 ± 5.77	0.92
HNR (dB)	コンデンサ	25.04 ± 3.50	19.12 ± 4.48	13.12 ± 7.73	9.16 ± 0.63	8.02 ± 1.58	7.01 ± 3.63	-0.60
	ダイナミック	21.74 ± 4.00	15.14 ± 4.72	10.01 ± 7.70	8.89 ± 0.64	7.14 ± 2.07	5.99 ± 3.85	-0.38

ICC : 級内相関係数 (intra-class correlation)

の一致性を認めた。APQについては、 APQ_{Praat} の値と比較し APQ_{MDVP} の値がやや高値であったが、いずれのマイクロフォンにおいても高い一致性($ICC=0.95$ および $ICC=0.92$)を認めた。しかしながら、HNRにおいては、 HNR_{Praat} の値が HNR_{MDVP} の値と比較して2.5倍程度となり、いずれのマイクロフォンにおいても全く一致性を認めなかった($ICC=-0.60$ および $ICC=-0.38$)。なお、相関性については、いずれのパラメータにおいてもソフトウェア間において高い相関性を認めた(図8)。

また、マイクロフォン間におけるパラメータの値の比較では、いずれのソフトウェアにおいても、ダイナミック型マイクロフォンにおけるPPQおよびAPQが高値を認め、HNRは低値を認めた(表1)。

次に、平均Gスコアと各パラメータとの相関性について検討した。PPQおよびAPQについては、いずれのソフトウェアおよびマイクロフォンにおいても、中等度から高度の相関性を認めた($r_s = 0.66 \sim 0.74$)。一方、HNRにおいては、 HNR_{Praat} はいずれのマイクロフォンにおいてもPPQおよびAPQにおける相関と同等の相関性であったが($r_s = -0.68 \sim -0.70$)、 HNR_{MDVP} においては HNR_{Praat} における相関よりも低い相関性であった($r_s = -0.54 \sim -0.63$)。

音声障害の診断精度については、相関性の検討と同様にPPQおよびAPQにおいては、いずれのソフトウェアおよびマイクロフォンにおいても高い診断精度($AUC = 0.853 \sim 0.901$)を認めたが、HNRにおいては、いずれのマイクロフォンにおいても HNR_{Praat} と比較し HNR_{MDVP} の診断精度が劣っていた(表2)。

考 察

上記の結果から、ソフトウェアやマイクロフォンが異なった場合においてもF0の抽出範囲を各波形に対して適切に調整すれば、F0についてはほぼ同一の結果が得られることが確認された。一方、音声障害の指標となるパラメータについて、PPQおよびHNRにおいては特にソフトウェアの違いによりパラメータの数値が大きく異なることが確認された。しかしながら、APQにおいては APQ_{MDVP} においてやや高い値が算出されるものの、近似した数値が得られることが確認された。その理由として、PPQの算出には抽出されたF0を直接的にPPQの算出に用いるためF0抽出における誤差がパラメータの数値の違いとして出現しやすいが⁶⁹⁾、APQの算出に用いられる振幅の値は抽出された1サイクルの波形内における最大音圧と最小音圧との差により決定される値であるため、F0抽出の誤差の影響を受けにくいこと^{52, 54)}が考えられた。

マイクロフォン間の比較においては、いずれのソフトウェアを用いた場合もダイナミック型マイクロフォンを用

図8 両ソフトウェアにより計算した音響分析パラメータの散布図。回帰直線と相関係数を示す。

表2 マイクロフォンおよびソフトウェア別の、各音響分析パラメータと嗄声の程度との相関および各音響分析パラメータによる嗄声の診断精度

		Praat					MDVP				
		平均Gスコアとの相関	AUC	カットオフ値	感度	特異度	平均Gスコアとの相関	AUC	カットオフ値	感度	特異度
PPQ	コンデンサ	0.66	0.853	0.29	0.74	0.85	0.70	0.861	0.53	0.79	0.81
	ダイナミック	0.68	0.873	0.27	0.77	0.86	0.70	0.868	0.92	0.67	0.91
APQ	コンデンサ	0.67	0.872	2.49	0.85	0.83	0.68	0.875	3.14	0.72	0.89
	ダイナミック	0.71	0.885	3.01	0.84	0.87	0.74	0.901	3.80	0.87	0.85
HNR	コンデンサ	-0.68	0.858	20.96	0.72	0.90	-0.54	0.781	8.76	0.72	0.78
	ダイナミック	-0.70	0.876	17.25	0.77	0.88	-0.63	0.831	8.12	0.69	0.86

AUC : receiver operatorating characteristic curveにおける曲線下面積 (area under the curve)

いた場合にPPQおよびAPQが高値あるいはHNRが低値となった。同じ音声を同時録音した場合のHNRが低値であるということから、ダイナミック型マイクロフォンによる録音の方が録音精度が悪く（＝録音の信号雑音比が低く）、そのためPPQおよびAPQが高値となったと考えられた。

一方、聴覚心理的評価との相関性や音声障害の診断精度については、いずれのソフトウェアおよびマイクロフォンを使用した場合も、少なくともPPQおよびAPQにおいては大きく損なわれることはなく、各施設で使用しやすいマイクロフォンおよびソフトウェアを用いたしても、臨床上の問題はないと考えられた。

結 語

- 1) 音響分析の概念について解説し、実際に施行する際に必要となる機器および設備、また、対象とする発声タスクについて解説した。
- 2) マイクロフォンやソフトウェアの違いがあったとしても、F0についてはほぼ同一の値が得られるが、PPQ, APQ, HNRなどの音声障害の指標については異なる値が得られる事が示された。
- 3) 今回使用したマイクロフォンやソフトウェアにおいては、いずれを使用しても音響分析パラメータと聴覚心理的評価との相関性および音声障害の診断精度については影響が少ないことが判明した。

本論文の要旨は、第28回日本喉頭科学会総会学術講演会(2016年3月大阪)・教育パネルディスカッション「音声障害評価の普及に向けて」において口演した。

本論文に関連し、利益相反に該当する事項はない。

参 考 文 献

- 1) Kreiman J, Gerratt BR, Kempster GB et al : Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. J Speech Hear Res 36 : 21-40, 1993.
- 2) 一色信彦 : 嗄声の分類記載法. 音声言語医学 7 : 15-21, 1966.
- 3) Isshiki N, Okamura H, Tanabe M et al : Differential diagnosis of hoarseness. Folia Phoniatr (Basel) 21 : 9-19, 1969.
- 4) Hirano M : Psycho-acoustic evaluation of voice. Disorders of Human Communication 5. Clinical Examination of Voice. (Arnold GE et al ed), 81-84 : Springer-Verlag, Vienna, 1981.
- 5) Lieberman P : Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. J Acoust Soc Am 35 : 344-353, 1963.
- 6) Koike Y : Application of some acoustic measures for the evaluation of laryngeal dysfunction. Studia Phonologica 7 : 17-23, 1973.
- 7) Takahashi H, Koike Y : Some perceptual dimensions and acoustical correlates of pathologic voices. Acta Otolaryngol Suppl 338 : 1-24, 1976.
- 8) Yumoto E, Gould WJ, Baer T : Harmonics-to-noise ratio as an index of the degree of hoarseness. J Acoust Soc Am 71 : 1544-1549, 1982.
- 9) 湯本英二, 岡本和憲, 佐々木裕美ほか : H/N比測定に影響を及ぼす諸因子について. 音声言語医学 28 : 145-151, 1987.
- 10) Kasuya H, Ogawa S, Mashima K et al : Normalized noise energy as an acoustic measure to evaluate pathologic voice. J Acoust Soc Am 80 : 1329-1334, 1986.
- 11) Hillenbrand J, Cleveland RA, Erickson RL : Acoustic correlates of breathy vocal quality. J Speech Hear Res 37 : 769-778, 1994.
- 12) Hillenbrand J, Houde RA : Acoustic correlates of breathy vocal quality : dysphonic voices and continu-

- ous speech. *J Speech Hear Res* 39 : 311-321, 1996.
- 13) Dejonckere PH, Remacle M, Fresnel-Elbaz E et al : Reliability and clinical relevance of perceptual evaluation of pathological voices. *Rev Laryngol Otol Rhinol* 119 : 247-248, 1998.
 - 14) Yamaguchi H, Shrivastav R, Andrews ML : A comparison of voice quality ratings made by Japanese and American listeners using the GRBAS scale. *Folia Phoniatr Logop* 55 : 147-157, 2003.
 - 15) 阿部博香, 米川紘子, 太田文彦ほか : 嗄声の聴覚心理的評価の再現性. *音声言語医学* 27 : 168-177, 1986.
 - 16) 溝尻源太郎, 柴 裕子, 米川紘子ほか : GRBAS尺度の再現性の向上に関する事柄. *音声言語医学* 37 : 305-311, 1996.
 - 17) Kreiman J, Gerratt BR, Precoda K : Listener experience and perception of voice quality. *J Speech Hear Res* 33 : 103-115, 1990.
 - 18) Gerratt BR, Kreiman J, Antonanzas-Barroso N et al : Comparing internal and external standards in voice quality judgments. *J Speech Hear Res* 36 : 14-20, 1993.
 - 19) De Bodt MS, Wuyts FL, Van de Heyning PH et al : Test-retest study of the GRBAS scale : influence of experience and professional background on perceptual rating of voice quality. *J Voice* 11 : 74-80, 1997.
 - 20) Sofranko JL, Prosek RA : The effect of experience on classification of voice quality. *J Voice* 26 : 299-303, 2012.
 - 21) Chan KMK, Yiu EML : The effect of anchors and training on the reliability of perceptual voice evaluation. *J Speech Hear Res* 45 : 111-126, 2002.
 - 22) Awan SN, Lawson LL : The effect of anchor modality on the reliability of vocal severity ratings. *J Voice* 23 : 341-352, 2009.
 - 23) Hosokawa K, Barsties B, Iwahashi T et al : Validation of the acoustic voice quality index in the Japanese language. *J Voice*: in press.
 - 24) Wolfe VI, Ratusnik DL : Acoustic and perceptual measurements of roughness influencing judgments of pitch. *J Speech Hear Disord* 53 : 15-22, 1988.
 - 25) Hirano M, Hibi S, Yoshida T et al : Acoustic analysis of pathological voice. Some results of clinical application. *Acta Otolaryngol* 105 : 432-438, 1988.
 - 26) Wolfe V, Fitch J, Cornell R : Acoustic prediction of severity in commonly occurring voice problems. *J Speech Hear Res* 38 : 273-279, 1995.
 - 27) Halberstam B. Acoustic and perceptual parameters relating to connected speech are more reliable measures of hoarseness than parameters relating to sustained vowels. *ORL* 66: 70-73, 2004.
 - 28) Vaz Freitas S, Melo Pestana P, Almeida V et al : Integrating voice evaluation: correlation between acoustic and audio-perceptual measures. *J Voice* 29 : 390 e1-7, 2015.
 - 29) Zeitels SM, Hochman I, Hillman RE : Adduction arytenopexy : A new procedure for paralytic dysphonia with implications for implant medialization. *Ann Otol Rhinol Laryngol* 107 : 2-24, 1998.
 - 30) Tokashiki R, Hiramatsu H, Tsukahara K et al : A "fenestration approach" for arytenoid adduction through the thyroid ala combined with type I thyroplasty. *Laryngoscope* 117 : 1882-1887, 2007.
 - 31) Kaneko M, Hirano S, Tateya I et al : Multidimensional analysis on the effect of vocal function exercises on aged vocal fold atrophy. *J Voice* 29 : 1-7, 2015.
 - 32) Parsa V, Jamieson DG : Acoustic discrimination of pathological voice: sustained vowels versus continuous speech. *J Speech Lang Hear Res* 44 : 327-339, 2001.
 - 33) Godino-Llorente JI, Osma-Ruiz V, Saenz-Lechon N et al : Acoustic analysis of voice using WPCVox: a comparative study with Multi Dimensional Voice Program. *Eur Arch Otorhinolaryngol* 265 : 465-476, 2008.
 - 34) Hosokawa K, Ogawa M, Hashimoto M et al : Statistical analysis of the reliability of acoustic and electroglottographic perturbation parameters for the detection of vocal roughness. *J Voice* 28 : 263 e9-16, 2014.
 - 35) Deliyski DD, Shaw HS, Evans MK : Adverse effects of environmental noise on acoustic voice quality measurements. *J Voice* 19 : 15-28, 2005.
 - 36) Titze IR. Workshop on Acoustic Voice Analysis. Summary Statement. Salt Lake City, UT, USA : National Center for Voice and Speech ; 1995.
 - 37) Barsties B, De Bodt M : Assessment of voice quality: current state-of-the-art. *Auris Nasus Larynx* 42 : 183-188, 2014.
 - 38) Švec JG, Granqvist S : Guidelines for selecting microphones for human voice production research. *Am J Speech Lang Pathol* 19 : 356-368, 2010.
 - 39) Zhu Y, Witt RE, MacCallum JK et al : Effects of the voice over internet protocol on perturbation analysis of normal and pathological phonation. *Folia Phoniatr Logop* 62 : 288-296, 2010.
 - 40) Vogel AP, Morgan AT : Factors affecting the quality of sound recording for speech and voice analysis. *Int J Speech Lang Pathol* 11 : 431-437, 2009.
 - 41) 今泉 敏 : 声の検査 - 音響的・聴覚心理的評価を巡っ

- て－. 音声言語医学 51:261-262, 2010.
- 42) Awan SNS, Roy N, Zhang D et al : Validation of the cepstral spectral index of dysphonia (CSID) as a screening tool for voice disorders: development of clinical cutoff scores. J Voice 29 : 1-15, 2015.
 - 43) Maryn Y, Corthals P, Van Cauwenberge P et al : Toward improved ecological validity in the acoustic measurement of overall voice quality : combining continuous speech and sustained vowels. J Voice 24 : 540-555, 2010.
 - 44) Barsties B, Maryn Y : External validation of the acoustic voice quality index version 03.01 with extended representativity. Ann Otol Rhinol Laryngol: 125: 571-583, 2016.
 - 45) Boersma P : Praat, a system for doing phonetics by computer. Glot Int 5 : 341-345, 2001.
 - 46) Milenkovic P : Least mean square measures of voice perturbation. J Speech Hear Res 30 : 529-538, 1987.
 - 47) 水田匡信, 庄司和彦, 児嶋 剛ほか : 音響分析ソフト「VA」の開発－軽度嗄声の音響分析－. 耳鼻臨床 104 : 297-302, 2011.
 - 48) Kay Elemetrics Corporation : Multi-Dimensional Voice Program (MDVP) model 5105, version 2.0 : Software Instruction Manual. Kay Elemetrics Corporation, Lincoln Park NJ, 1999.
 - 49) Milenkovic P : TF32 User's Manual. Paul Milenkovic, Madison WI, 2001.
 - 50) Titze IR, Liang H : Comparison of F0 extraction methods for high-precision voice perturbation measurements. J Speech Hear Res 36 : 1120-1133, 1993.
 - 51) Karnell MP, Hall KD, Landahl KL : Comparison of fundamental frequency and perturbation measurements among three analysis systems. J Voice 9 : 383-393, 1995.
 - 52) Bielamowicz J, Kreiman BR, Gerratt MS et al : Comparison of voice analysis systems for perturbation measurement. J Speech Hear Res 39 : 126-134, 1996.
 - 53) Roark RM : Frequency and voice : perspectives in the time domain. J Voice 20 : 325-354, 2006.
 - 54) Maryn Y, Corthals P, De Bodt M et al : Perturbation measures of voice: a comparative study between Multi-Dimensional Voice Program and Praat. Folia Phoniatri Logop 61 : 217-226, 2009.
 - 55) Boersma P : Should jitter be measured by peak picking or by waveform matching? Folia Phoniatri Logop 61 : 305-308, 2009.
 - 56) Dejonckere P, Schoentgen J, Giordano A et al : Validity of jitter measures in non-quasi-periodic voices. Part I : perceptual and computer performances in cycle pattern recognition. Logoped Phoniatr Vocol 36 : 70-77, 2011.
 - 57) Winholtz WS, Titze IR : Miniature head-mounted microphone for voice perturbation analysis. J Speech Hear Res 40 : 894-899, 1997.
 - 58) Ma EPM, Yiu EML : Suitability of acoustic perturbation measures in analysing periodic and nearly periodic voice signals. Folia Phoniatri Logop 57 : 38-47, 2005.
 - 59) 西尾正輝, 田中康博, 新美成二 : 加齢に伴う音声の変化. 音声言語医学 50 : 6-13, 2009.
 - 60) Lovato A, De Colle W, Giacomelli L et al : Multi-dimensional voice program (MDVP) vs Praat for assessing euphonic subjects : a preliminary study on the gender-discriminating power of acoustic analysis software. J Voice : in press.
 - 61) 藤本保志, 岩田義弘, 小澤喜久子ほか : 喉頭癌音声温存術式の音声・嚥下機能の比較. 喉頭 19 : 70-74, 2007.
 - 62) 兒玉成博, 讃岐徹治, 東家 完ほか : 披裂軟骨内転術と併用術式の発声機能評価. 音声言語医学 52 : 149-157, 2011.
 - 63) 金子真美, 平野 滋, 楯谷一郎ほか : 歌唱者の音声障害に対する音声治療の効果. 音声言語医学 55 : 201-208, 2014.
 - 64) 川村直子, 城本 修, 望月隆一ほか : Vocal Function Exercise (VFE) の訓練効果に影響を及ぼす要因. 音声言語医学 55 : 137-145, 2014.
 - 65) 羽石英里, 斎藤 毅, 城本 修ほか : 音楽療法士を対象とした発声訓練プログラム. 音声言語医学 54 : 186-196, 2013.
 - 66) 足立千浪, 吐師道子, 城本 修ほか : カラオケ歌唱による声質変化と水分摂取の関係. 音声言語医学 55 : 195-200, 2014.
 - 67) 松本かおり, 中川秀樹 : 職業の違いによる音声障害のリスクについての検討－自覚的評価および客観的評価による分析－. 音声言語医学 56 : 315-320, 2015.
 - 68) Landis JR, Koch GG : The measurement of observer agreement for categorical data. Biometrics 33 : 159-174, 1977.
 - 69) Manfredi C, Giordano A, Schoentgen J et al : Perturbation measurements in highly irregular voice signals: Performances/validity of analysis software tools. Biomed Signal Process Control 7 : 409-416, 2012.