

機械装置の稼働率を高める監視技術、 アナログ・デバイセズのAIで音の意味を把握する

著者: Sebastien Christian

はじめに

機械装置の保守の必要性を実感している人であれば、装置が発する音や振動がいかに重要な意味を持つか知っています。音や振動を利用して適切に装置の状態を監視することにより、装置の保守にかかるコストを1/2に削減し、寿命を2倍に延伸することができます。状態基準保全（Conditional Based Maintenance）向けのシステムを実現する上では、音のライブ・データを収集／解析する手段を実装することが、1つの重要なポイントになります。

そうしたシステムでは、まず装置の正常な稼働音を把握します。音に変化が生じたら、それを異常として検知します。その結果、問題を特定することができれば、その音を特定の問題と関連づけられます。異常の検知は、わずか数分のトレーニングで行えるようになります。ただ、音や振動をその原因と結び付けて診断できるようになるまでには、かなりの時間を要します。そのような知識を身につけた経験豊富な技術者は存在しないわけではありません。しかし、決して多くはありません。音だけで問題を直感的に認識するというのは、蓄積されてきた記録や詳しく解説された枠組みがあったり、専門家から個人的な指導を受けたりしたとしても、難易度が高い可能性があります。

そこで、筆者が所属するアナログ・デバイセズでは、この20年間、人間が音や振動の意味を理解する仕組みを解明すべく、取り組みを進めてきました。その目的は、装置の音や振動を把握し、その意味を解釈して異常な動作を検出できる診断システムを構築することでした。本稿では、装置の状態監視を行うためのアーキテクチャ／プラットフォームである「OtoSense」について解説します。OtoSenseは、「コンピュータ聴覚（Computer Hearing）」という機能を実現します。この機能は、装置の状態を表す音と振動を主な指標とし、その意味をコンピュータによって解釈するというものです。

OtoSenseをベースとするシステムは、任意の装置に適用可能です。また、ネットワークに接続することなく、リアルタイムに動作します。産業用アプリケーション向けに構築されており、装置の状態を監視するためのスケラブルで効率的なシステムを実現できます。

本稿では、まずOtoSenseの開発方針について説明します。続いて、設計に際して参考にした人間の聴覚の仕組みについて詳しく解説します。そのうえで、音または振動の特徴（Feature）を表現する方法、それらから意味を導出する方法を紹介します。最後に、より高い精度でより複雑な診断を実行できるように、時間の経過に応じてOtoSenseを進化／改善するための継続的な学習プロセスについて説明します。

開発方針

堅牢かつアグノスティックで効率的なシステムを構築するために、OtoSenseは以下に示すいくつかの方針に基づいて設計しました。

- ▶ 神経学から着想を得る：人間は耳にする任意の音を学習し、それに意味を持たせることができます。また、非常に高いエネルギー効率でそれを行います。
- ▶ 定常音と過渡音を学習できるようにする：これには、適切な特徴の抽出と継続的な監視が必要です。
- ▶ センサーの近くのエッジで認識を実行する：リモート・サーバとのネットワーク接続を必要とすることなく、判断が行えるようにします。
- ▶ 専門家に負荷がかからないようにする：専門家と意見を交換したり、専門家から学んだりする必要性が生じた場合に、専門家の日々の仕事量に及ぼす影響を最小限に抑えます。また、そうしたやり取りの過程を楽しいものに努めます。

人間の聴覚システムをOtoSenseに変換

人間の聴覚は、生存を維持するために備わっている感覚です。この感覚は、遠方で生じた事象や目に見えない事象を全体的に捉えるためのものであり、生まれる前の段階で成熟します。

人間が音に意味を持たせるプロセスは、アナログでの音の取得、デジタル信号への変換、特徴の抽出、解釈という4つのなじみ深いステップで表すことができます。以下では、各ステップについて、人間の耳とOtoSenseベースのシステムを対比して説明します。

- ▶ アナログでの音の取得、デジタル信号への変換：人間は、中耳の鼓膜と耳小骨によって音を取得します。そのうえで、インピーダンスを調整し、リンパ液で満たされた蝸牛に振動を伝達します。蝸牛管内の基底膜は、信号に含まれる周波数成分に応じて選択的に変位します。その結果、柔軟な細胞が屈曲し、その曲がり具合と強さに応じたデジタル信号が出力されます。個々の信号は、周波数成分ごとにふり分けられ、平行する神経経路を通して1次聴覚野に到達します。

- OtoSenseでは、センサー、アンプ、コーデックによって、上記の内容と同等の処理を行います。デジタル信号への変換処理には、250Hz～196kHzの範囲で調整可能な固定のサンプル・レートを使用します。信号は16ビットで符号化され、128～4096個のサンプル・データがバッファに格納されます。

- ▶ 特徴の抽出：1次聴覚野において、支配的な周波数、高調波、スペクトル形状といった周波数領域の特徴と、約3秒間の時間ウィンドウ内における衝撃、強度の変動、主要な周波数成分といった時間領域の特徴が抽出されます。

- OtoSenseでは、チャンク（Chunk）と呼ばれる固定のステップ・サイズで移動する時間ウィンドウを使用します。チャンクのサイズとステップは、認識しなければならない事象とサンプル・レートに応じて23ミリ秒～3秒の範囲内で設定され、特徴はエッジで抽出されます。OtoSenseによって抽出される特徴については、次のセクションで詳しく説明します。

- ▶ 解釈：解釈が行われる連合野は、すべての知覚と記憶を融合し、言語などによって音に意味を付与します。つまり、人間の知覚を形成する上で非常に重要な役割を果たします。解釈は、事象に対して単に名前を付けるだけの作業ではありません。それをはるかに超えた描写を整理するプロセスです。事柄、音、出来事などに名を付与することで、人間はそれらに対してより深く重層的な意味を与えることができます。名称や意味によって、専門家はその環境についてより深く理解できるようになります。

- OtoSenseと人間の相互作用が、人間の神経学に基づき、教師なし（Unsupervised）の視覚的な音声マッピングから始まる理由はここにあります。OtoSenseは、取得したすべての音や振動をグラフィカルに表現し、類似性に基づいて分類します。但し、厳密なカテゴリを設けようとはしません。そのため、専門家は、有界のカテゴリを人工的に作成することなく、画面上に表示された分類集合を整理して命名することができます。また、専門家はOtoSenseの最終出力に関する自らの知識、知覚、期待に合致したセマンティック・マップを構築することが可能になります。自動車の整備士、航空宇宙分野の技術者、冷間鍛造プレスの専門家といった分野ごとに、あるいは同じ分野でも企業ごとに、同じサウンドスケープが異なる形で分割、整理、ラベル付けされる可能性があります。OtoSenseは、人間が言語を使用するのと同様に、ボトムアップのアプローチで意味を形成します。

音と振動から特徴を抽出

特徴には、一定の時間（時間ウィンドウ、つまりはチャンクで決まる）における音または振動の属性／質を表す番号が割り当てられます。OtoSenseは、以下の方針に従って特徴を選択します。

- ▶ 特徴は環境を表すものでなければならない：特徴は、周波数領域と時間領域の両方において、できるだけ完全かつ詳細に環境を表すものであることが必要です。「ブーン」、「カチカチ」、「ガタガタ」、「キーキー」といった定常音と、任意の種類の過渡的な不安定性を表せるものでなければなりません。
- ▶ 特徴は、できるだけ直交した集合を構成しなければならない：例えば、1つの特徴が「チャンク内の平均振幅」として定義されている場合、「チャンク内の合計スペクトル・エネルギー」といった具合に、それと強く相関を持つ別の特徴が存在してはなりません。当然のことながら、完全な直交性は達成できませんが、他の特徴の組み合わせとして表現される特徴が存在するのは許されません。言い換えれば、各特徴には何らかの特異な情報が含まれている必要があります。
- ▶ 特徴は、演算を最小限に抑えられるものでなければならない：人間の脳にできるのは、加算、比較、ゼロへのリセットだけです。OtoSenseのほとんどの特徴は、インクリメント変数として設計されています。新たなサンプルが加わるごとに、特徴は単純な演算によって変更されます。そのため、チャンク全体はもちろん、バッファ全体を対象として特徴を再計算する必要はありません。演算の最小化は、標準的な物理単位について配慮する必要がないということも意味します。例えば、強度を表すのにdBAを単位とする値を使う必要はありません。そうした値が必要であれば、出力する際にそのための演算を行うだけで済みます。

OtoSenseでは、時間領域に対応する2～1024の特徴を扱います。それらは信号から直接抽出されるか、またはチャンク内の他の特徴から導出されます。それらの特徴には、平均または最大の振幅、信号の線形長から導出した複雑さ、振幅の変動、衝撃の存在とその特性、最初と最後のバッファ間の類似性としての安定性、畳み込みを行わない状態での皮相的な自己相関、主要なスペクトルのピークの変動などが含まれます。

周波数領域で使われる特徴は、FFT（高速フーリエ変換）の結果から抽出されます。FFTの演算はバッファ単位で行われ、128～2048の周波数成分が生成されます。続いて、必要な次元数のベクトルを作成します。次元数は、当然のことながら、FFTのサイズよりもはるかに小さいわけですが、環境を包括的に表現することができます。OtoSenseでは、まずアグノスティックな方法により、対数スペクトルに対する等サイズのバケットを作成します。続いて、環境と識別したい事象に応じてエントロピーを最大化する教師なしの観点、またはラベル付けされた事象を指針として利用する半教師あり（Semi-supervised）の観点のうちいずれかに基づき、情報の密度が高いスペクトル部分に焦点が絞られるようにバケットを調整します。これは、音声情報が最大の部分で密度が高くなる、人間の内耳細胞の構造を模倣したものです。

アーキテクチャ：エッジとオン・プレミスのデータに力を与える

OtoSenseによる異常値の検出と事象の認識は、リモートのアセットを全く利用することなくエッジで行います。このようなアーキテクチャであることから、システムはネットワーク障害の影響を受けません。また、分析を行うためにすべての未加工データを送信する必要がなくなります。OtoSenseの処理を実行するエッジ・デバイスは、音を使ってリアルタイムに装置の問題を把握する自己完結型のシステムとなります（図1）。

通常、AI（Artificial Intelligence：人工知能）とHMI（Human Machine Interface）の機能を実行するOtoSenseのサーバは、オン・プレミスの形で運用されます。意味のある複数のデータ・ストリームをOtoSenseに対応するデバイスの出力としてまとめるという目的に対しては、このクラウド・コンピューティングのアーキテクチャが合理的な選択肢となります。OtoSenseのAIは、大量のデータを処理して1つのサイト内にある数百台のデバイスとやり取りするだけです。この目的に対して、ホスティング型のクラウド・コンピューティングを利用するのはあまり意味がありません。

特徴に基づく異常の検出

正常か異常かの評価には、さほど専門家が介入する必要はありません。専門家の介入が必要になるのは、装置の正常な音と振動の基準を確立する際だけです。その基準が、OtoSenseのサーバで使用する異常値のモデルに変換され、デバイスに適用されます。

入力された音や振動が正常な場合のものかどうかを評価するには、以下に示す異なる2つの戦略を適用します。

- ▶ 1つ目の戦略は、正常性と呼ばれるものです。特徴空間に新たに入力された音について、その環境、基準点やクラスタとの距離、クラスタの大きさを確認します。距離が遠く、クラスタが小さいほど、その音の異常性は高く、異常値のスコアは高くなります。異常値のスコアが、専門家が定義した閾値を超えている場合、該当するチャンクは「異常」としてラベル付けされ、専門家が確認できるようにサーバに送信されます。
- ▶ 2つ目の戦略は、とてもシンプルなものですが。チャンクに対する入力の特徴の値が、基準を定義する全特徴の最大値を上回るか最小値を下回る場合、そのチャンクは「極度の異常」としてラベル付けされ、サーバに送信されます。

「異常」と「極度の異常」の2つを組み合わせることにより、異常な音や振動を十分に網羅し、漸進的に進行する摩擦と、予期せぬ過酷な事象を検出することができます。

特徴に基づく事象の認識

特徴は物理的な領域に属し、意味は人間の認識力に属します。特徴と意味を関連づけるためには、OtoSenseのAIと人間である専門家の間の相互作用が必要になります。当社は、かなりの時間を費やして顧客からのフィードバックを取り入れ、技術者がOtoSenseと効率的にやり取りして、事象の認識モデルを設計することを可能にするHMIを開発しました。このHMIを使えば、データの調査を行い、ラベル付けを実施し、異常値のモデルと音の認識モデルを作成し、それらのモデルのテストを行うことができます。

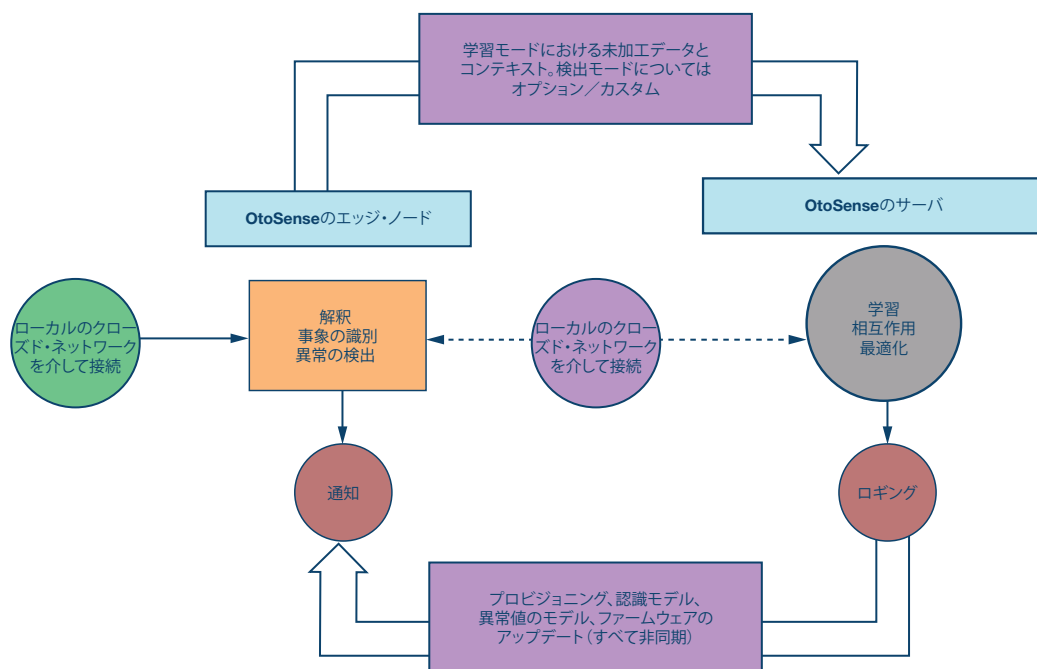


図1. OtoSenseをベースとするシステム

「OtoSense Sound Platter」（splatterとも呼ばれます）を使えば、データ・セットの包括的な概観を把握しつつ、音の調査とタグ付けを行うことができます。splatterは、データ・セット全体の中から関心を持つべき代表的な音を選択し、ラベル付けされた音とされていない音が入り混じった2次元の類似性マップとしてそれらを表示します（図2）。

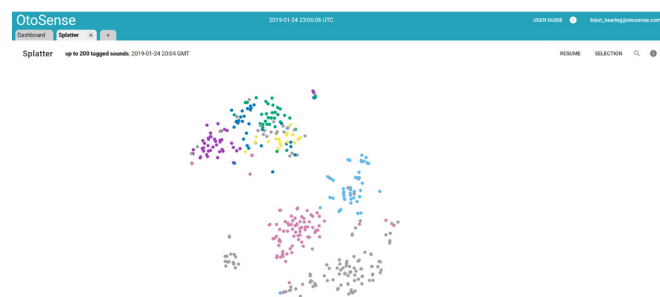


図2. OtoSense Sound Platterによる2次元マップ（splatterマップ）

OtoSenseでは、任意の音または振動を、コンテキストと共にいくつかの異なる方法で視覚化することができます。その1つは「Sound Widget」（swidgetとも呼ばれます）を使用する方法です（図3）。

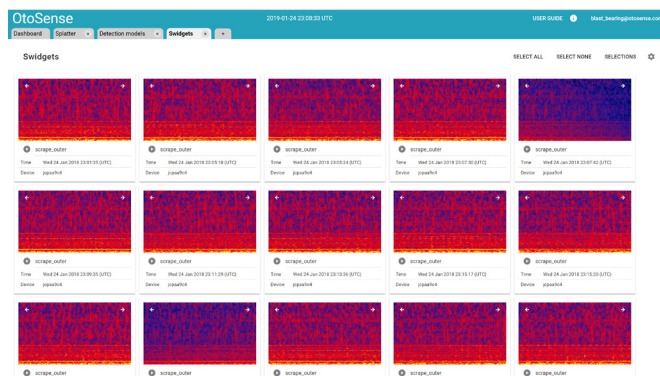


図3. OtoSenseのSound Widget（swidget）

また、異常値のモデルまたは事象の認識モデルは、任意の時点で作成できます。事象の認識モデルは、円形の混同行列として表現されます（図4）。OtoSenseのユーザーは、それを使用することにより、混同事象について調査することができます。

MODEL LIST	PARAMETER				STATISTICS							
tag	kl_jarvis	kl_jarvis	tp	tp	ts	ts	accuracy	precision	recall	f1		
jump_off	88	88	88	0	263	0	1.00	1.00	1.00	1.00		
straps_jarvis	81	80	80	0	270	1	1.00	1.00	0.99	0.99		
oil_buck	74	84	68	16	261	6	0.94	0.81	0.92	0.86		
normal	73	53	51	2	276	22	0.93	0.96	0.70	0.81		
straps_inner	35	46	35	11	305	0	0.97	0.76	1.00	0.86		

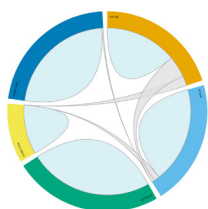


図4. 必要な事象に基づいて作成された事象の認識モデル

異常値に対しては、一定の時間内のすべての異常音と極度な異常音を表示するインターフェースを介して、調査とラベル付けが行えます（図5）。



図5. 時系列での音の分析。OtoSenseにおいて異常値を視覚化することで実現しています。

異常の検出から、より複雑な診断までの継続的な学習プロセス

OtoSenseは、複数の専門家から学習し、時間の経過に伴って、より複雑な診断ができるように設計されています。通常のプロセスは、OtoSenseと専門家の間の再帰的ループによって行われます。その詳細は以下のようになります。

- ▶ 異常値のモデルと事象の認識モデルはエッジで実行されます。それらによって、異常値のスコアと共に、潜在的な事象が発生する確からしさを表す出力が生成されます。
- ▶ 定義された閾値を超える異常な音または振動によって、異常値の通知がトリガされます。これを受けて、OtoSenseを使用する技術者は、その音とコンテキストを確認することができます。
- ▶ 続いて、技術者は、その異常な事象にラベル付けを行います。
- ▶ その新たな情報を含む新たな認識モデルと異常値のモデルが計算され、エッジ・デバイスに送られます。

まとめ

アナログ・デバイセズのOtoSenseは、任意の装置に対して音と振動に関する専門的な知見を継続的に提供し、ネットワークに接続することなく、異常値の検出と事象の認識を実施できるようにすることを目的としたものです。この技術は、航空宇宙、自動車、産業の各分野において、装置の状態監視向けに広く採用されつつあります。かつては人間の専門的な知見を必要とし、特に複雑な装置に関しては組み込みアプリケーションが必要だった環境において、優れた実績を積み重ねています。

参考資料

Sebastien Christian「How Words Create Worlds（言葉が創り出す世界）」TEDxCambridge、2014年



著者：

Sebastien Christian (sebastien.christian@analog.com) は、アナログ・デバイスでの製品化に向けて、OtoSenseの開発を統括しています。

Christianはかなり早い時期から「人間が自らの感覚を利用して、世界の内面モデルを共有可能なものとして構築し、そのモデルを使って人間が住む世界を表現する方法を理解したい」と強く願っていました。量子物理学の修士号に続いて神経科学の修士号を取得した後、セマンティクスに関する学位も取得しています。これらの分野において、研究、開発、フィールド実験を融合した形で学問を修めてきました。その後、精神障害や聴覚障害を抱える子供を対象とする発話と言語の病理学者として、10年間就業しました。それを通して、センサーをベースとし、聴覚に重点を置いて意味の形成／共有を実現する手法について理解を深めていきました。Christianは、「若い患者たちと何年も向き合った当時の経験によって、ばらばらだった知識のかげらのすべてが首尾一貫した1つの形にまとまった」と述べています。

同じ時期に、フランス保健省のエキスパートとなり、難聴に関する制度に対して助言を行っていました。医科大学やパリ・ソルボンヌ大学で教鞭を取り、2011年には、知覚障害や認知障害を抱える人々に対し、AIに基づくイノベーションをもたらすことを専門とする初の独立系民間R&D施設を創設しました。

2013年には、機械聴覚（Machine Hearing）プロジェクトにおいて完全なプロトタイプを完成させ、米マサチューセッツ州ケンブリッジで行われた技術コンペティション「NETVA」で賞を授与されました。MIT（マサチューセッツ工科大学）の同僚や産業界からかなり好意的なフィードバックを得たことを受けて、2014年初頭にOtoSenseを創設しました。そして、任意の音に意味づけすることに焦点を絞った初のAIを開発しました。機械聴覚に対応するこのプラットフォームは、複雑な環境や複雑な機械の監視にうまく適応できるということが明らかになりました。

開発したAIは、2015年の「GSMA Mobile World Congress」において、「Best App of the Year」を受賞しました。その他にいくつかの賞を授与された後、OtoSenseの開発は、産業分野と輸送分野における機械の監視に焦点を絞って進められるようになりました。その先にある広範かつ潜在的な用途への展開も見据えています。