White Paper

aculabcloud
A true cloud telecoms platform

Telephone-based speech analysis in the healthcare industry

Executive summary

Speaking is something that most of us tend to take for granted. Unfortunately, there are those for whom it becomes a problem. And as clinicians and speech therapists will tell you, such problems can strike at any moment in a person's life.

Speech dysfunction is evaluated by medical professionals such as specialist clinicians and therapists. The means at their disposal for the screening and monitoring of patients usually requires face-to-face meetings. For some health authorities, budget permitting, it involves a certain amount of technology, much of which is dated. Regular meetings between patient and clinician are needed in order to monitor the progress of any treatment. That is undoubtedly costly to both parties and offers a significant incentive for innovation.

A telephone-based voice and speech analysis system is proposed as a tool for clinicians, speech therapists and medical researchers. It is intended to enable medical specialists and professionals to offer an outpatient service for monitoring voice characteristics as a cost-effective means of contributing to the management of certain disorders and diseases. It is also feasible for the system to be considered for screening of disorders affecting a person's speech capability.

The system envisaged will provide an automatic, objective, efficient and effective evaluation of subjects' voice and speech characteristics, reducing the need for specialist, on-site equipment or facilities.

The evaluation process is conducted remotely and is fully automated. No specialist hardware is required, and no special training need be undertaken by the patient. The use of a cloud-based telephony platform means that overall system costs are minimal, and the solution is fundamentally flexible and scalable. The proposed system ensures that regular monitoring as an aid to the management of speech disorders is economically viable.

For IT developers working in healthcare, such as those businesses who already offer specialist IT services including patient management and advisory (PMA) and electronic health record (EHR) systems, the proposed system can provide an opportunity to enhance their service offering, by integrating APIs for voice and speech analysis in a cost-effective and scalable manner.

Contents

Contents

Introduction

Introduction

The ability to articulate the sounds needed to speak and communicate, regardless of language, requires fine motor control, both in terms of timing and precision of movement. That is something that most of us tend to take for granted.

Unfortunately, there are those for whom it becomes a problem. And as clinicians and speech therapists will tell you, such problems can strike at any moment in a person's life. An individual's degree of control can be adversely affected by many diseases and disorders, such as Parkinson's disease and cerebral palsy. The same can be said as a result of physical injury leading to brain damage or through suffering a stroke, for example.

The myriad causes of dysfunction of speech production can affect more than one vocal characteristic. As a result, medical professionals, such as specialist clinicians and speech therapists, evaluate their patients' speech quality using many factors. The methods of evaluation involve a mixture of subjective and objective analysis, and both qualitative and quantitative data.

Typically, the means clinicians have at their disposal for the screening and monitoring of patients usually requires face-to-face meetings. For some health authorities, budget permitting, it involves a certain amount of technology, much of which is dated. One result of the essentially non-portable nature of the technology is that, other than during regular therapy sessions at the clinic, the clinician has no ready means of monitoring the patient's condition.

Furthermore, any monitoring carried out during a visit to a clinic will fail to produce any information other than data related to the patient's condition 'on the day'. It may be that the patient displays less acute symptoms at that specific time, perhaps during a moderate period of remission in a chronic case. In any event, there can be no intermediate, day-to-day history of a patient's speech function.

All told, the situation presents a significant incentive for innovation. Opportunities for advancing the cause exist through applying technology to enable clinicians to offer an alternative, outpatient service that can be accurate, efficient and cost-effective – without compromising ongoing treatment or diagnosis. An automatic and objective speech analysis system that uses the telephone network to provide quantitative data to assist clinicians in diagnosing, monitoring and managing many diseases and medical conditions that affect patients' voice or speech must be a welcome advance.

Speech dysfunction

Speech dysfunction

Speaking can be adversely affected by many diseases and disorders. Conditions such as apraxia and dysarthria can result from a variety of causes, including:

- Stroke, brain injury or progressive illnesses often result in apraxia of speech, which involves inconsistent production of sounds or rearranging sounds in words, and may also cause substitution, subtraction or addition of new sounds in words
- Dysarthria, which means the sounds in words are slurred or less prominent, is often caused by neurological disorders
- Parkinson's Disease, strokes, cerebral palsy, head or neck injuries, and anything that causes paralysis or weakness of speech muscles due to damage to the nerves and/or brain, can be causal factors of dysarthria
- Voice disorders such as spasmodic dysphonia and oral or laryngeal cancer, are caused by physical impairments involving the use of vocal resonance or the larynx

The causes of dysfunction in the production of speech can affect more than one clinical voice/speech characteristic. As a result, clinicians evaluate speech quality using many factors[1]. Those can include:

Breathiness	Creakiness	Phoneme quality
Phoneme sustainability	Pitch	Range
Rhythm	Speed	Steadiness
Tone	Volume	Word initiation

Table 1 - Clinical voice/speech characteristics

Currently, dysfunction in the production of speech is evaluated by a clinician or speech therapist. Notwithstanding routine therapy sessions, that usually requires the patient and the clinician to meet on a regular basis to monitor the progression of any treatment, which can be costly to the medical practice, in terms of time and resources, and to the patient.

The clinician uses specialist recording and analysis equipment, such as the Pentax Medical (previously Kay Elemetrics) Computer Speech Laboratory (CSL)[2], in order to measure the various factors. That equipment is expensive and not readily portable, confining its use to the clinic or therapy lab. In addition, such systems typically only analyse a limited number of traditional voice parameters[3] based on sustained phonation, hence there is a certain degree of subjective analysis[4] required of the clinician.

Less costly alternatives, such as OperaVOX[5] and other smartphone 'apps', which are obviously far more portable, are nevertheless comparable in the sense that they analyse a similarly limited number of traditional voice parameters, again solely based on sustained phonation.

Speech dysfunction cont

Integration

In summary, there are a number of issues with the current methods of evaluation and treatment. Those include:

- Existing equipment is restricted to a limited number of voice parameters
- The clinician's evaluation is subjective and, therefore, may be inconsistent
- Not all voice or speech parameters are audible to the clinician
- Voice/speech can change from day to day as well as at different times of the day
- Clinicians can only capture the state of the condition at the time of the clinical visit
- There is a high demand on clinicians' time and the time per patient may be limited
- Early stage deterioration or mild conditions can be missed

The proposed system is intended to remove those issues. It will fundamentally improve the possibilities and scope of monitoring, evaluation and treatment.

Proposed solution

Proposed solution

The proposed solution provides healthcare system developers with the software building blocks to create an automatic and objective voice/speech analysis service, intended to assist clinicians in the monitoring, screening and treatment of patients with many diseases and medical conditions that affect their voice or speech.

Those building blocks comprise a cloud-based communications platform-as-a-service (CPaaS) and specialist speech analysis software. From strategically located cloud infrastructure, the CPaaS provides the essential building blocks of worldwide connectivity for making and receiving calls, or sending/receiving faxes and SMS messages, and APIs for invoking an analysis of a caller's speech. The platform is operated on a pay-as-you-go approach – a healthcare service provider need only pay for the system when it is being used – and the elastic nature of the network means that, when necessary, it can support large traffic spikes.

The appetite for cloud-based solutions amongst healthcare providers is evidenced by the 2016 Healthcare Cloud Survey[6], which states that 73% of organisations are aggressively using or planning to move to cloud to host, amongst other things, patient engagement and empowerment tools. As further endorsement that suppliers of patient management and advisory (PMA) and electronic health record (EHR) systems have embraced cloud-based technologies, real-world case studies[7] exist, involving organisations such as eClinicalWorks.

To the end users (clinicians and outpatients), the system consists of a tailored application, accessed either online, via a browser using click-to-call functionality (e.g., WebRTC), or from any landline phone, SIP phone or mobile/cell/smartphone. Such a system is scalable, cost-effective, and globally accessible – 24 hours per day, every day of the year.

For software developers providing solutions into the healthcare industry, the proposition consists of simple, high-level REST APIs accessible from a variety of common languages (e.g., PHP, Python, Java, .NET or Ruby) that enable the quick and easy development of many communications related applications, including specifically, speech analysis applications for healthcare service providers. The APIs provide both the speech analysis capabilities and the means to communicate with, and provide direction to, patients using the service.

The cloud-based CPaaS is used to provide prompts and instructions to therapy subjects or patients, using voice, synthetic speech and speech recognition, in addition to SMS texts, and faxes. An architectural model, based on Aculab Cloud, is presented in Figure 1.

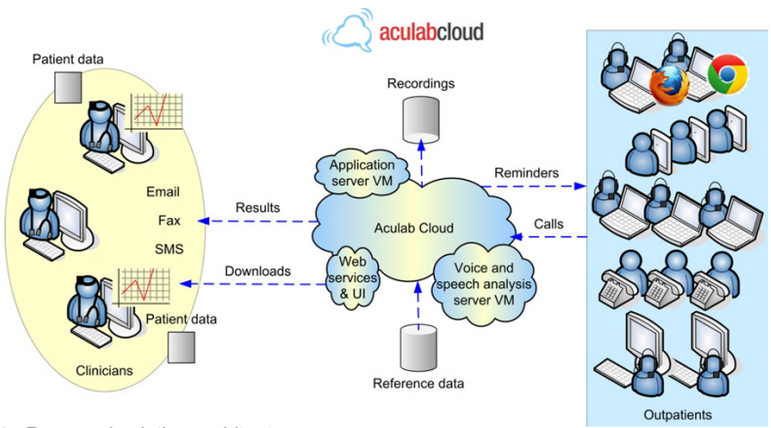

Figure 1 - Proposed solution architecture

How it works

How it works

Telephone speech contains enough information for a caller's speech to be understood, the speaker to be recognised, and for many speech pathologies to be assessed. Each patient's speech is recorded in real-time, during a call they make into the system. Calls can be scheduled and as frequent as necessary, and the system can be programmed to issue reminders to the patient e.g., for schedule reminders and appointment confirmations.

All recordings are analysed by the system before a result is produced, which can then be presented to the practicing therapist or responsible clinician via a suitable web page or web-based user interface. Naturally, secure log-ins and encryption will be used to protect sensitive patient data. Alternatively, the results and recordings can be downloaded, or forwarded to the clinician or researcher.

Note that the scope for presenting the analysis data is determined by the application developer, and ideally, the content and format will be configurable to suit the healthcare provider ultimately offering the patient service. Typically, the output will be presented as graphs, histograms, waveforms, or charts.

Unlike current systems, where the patient's speech pathology 'on the day' is hit or miss, coinciding with the timing of the clinical appointment, the proposed system can be used on a regular basis so that both 'good' and 'bad' days can be measured. In fact, as many calls as needed can be made over as many days as necessary in order for a thorough assessment to be made, and for effective monitoring of progress through periods of deterioration and remission. That means underlying trends can be identified far more quickly and effectively through the availability of historical information.

Beneficially, the system can be used anywhere a phone call can be made. However, it should be obvious that limiting background noise is desirable for best accuracy results, notwithstanding the fact that the system is designed to be immune to normal background noise, and robust against environmental variations. It goes without saying that it is better to call from the comfort of your own home than from a busy high street during rush hour. Of course, the system can be accessed also by the patient while attending the local clinic e.g., during any routine appointment for treatment or therapy.

In addition to the system described above, a cloud-based CPaaS has other uses in healthcare such as to automatically remind patients of clinical appointments and confirm their availability, to remind patients of scheduled times for self-medication procedures or of home visits and check-ups, to confirm that they have taken regular medication at the scheduled time, and to advise of the availability of lab results or prescriptions for collection.

Speech analysis capabilities

Speech analysis capabilities

The solution proposed is readily deliverable, with the necessary speech technology being available from Aculab, via Aculab Cloud and its CPaaS API suite. Regarding the provenance of its speech analysis capability, Aculab collaborated with the Phonetics Laboratory in the Faculty of Linguistics, Philology and Phonetics at Oxford University and MIT to develop its software. The resulting capability enables the analysis of recordings of voice responses made to questions posed over the phone, searching for acoustic characteristics indicative of a range of conditions. Further research[8] into the related benefits of automatic speech analysis for diagnosis and assessment of speech disorders has also been published by speech scientists employed by Aculab.

Uniquely, Aculab's system measures over 170 voice and speech parameters. Those include the parameters in Table 1 as well as other conventional measures, such as shimmer, jitter, harmonics to noise ratio (HNR), and formant trajectories, and in addition, further novel, perceptually undetectable, but still measurable parameters.

Advanced algorithms are used to analyse the movement of the vocal cords and the larynx in a way that is robust to variations due to the recording environment and the communication channel. Lexical or phonetic alignment based on automatic speech recognition (ASR), as used in studies resulting in a number of papers[9], is also used to quantify speech parameters that are beyond the capabilities of existing clinical systems, and which can be used to improve the detection of conditions such as Parkinson's Disease.

The objective results of the system's analysis, together with audio recordings, can aid clinicians in their evaluation of the subject's voice/speech. The results of the analysis can be presented in several ways:

- Instant results, compared with data already collected from thousands of healthy subjects
- Graphs showing the progression of the parameters for an individual subject, allowing longitudinal studies and monitoring
- Application of a machine learning system for screening and detection of abnormalities

Instant results may be all that are required by a therapist engaged in monitoring an individual patient. If a result shows zero abnormality, the patient is effectively given a lease until the next scheduled call to the system. Downloads on the other hand, including graphs, recordings and other data, can be used by a clinician for more detailed analysis on an ongoing basis and in ways of their choosing.

Moreover, the measured parameters can be fed into a machine learning algorithm to give an estimate of the probability of a speech/voice disorder being present. Alternatively, they can be compared to reference distributions covering thousands of healthy subjects, to aid a clinician in their diagnosis. The system is most effective when used for monitoring of an individual patient's voice and speech characteristics throughout the day, or longitudinally over a period of weeks, months or even years.

Use case example

Parkinson's Disease data analysis

Aculab collected telephone recordings of over 25,000 volunteers from nine countries around the world, together with a small amount of demographic information, such as age, gender, and known speech/voice pathology. That data has been used to provide a baseline to compare the system's parameters against, so a clinician can identify those subjects most likely to suffer from some form of speech/voice pathology. Table 2 summarises the content of the data collected from the USA; data from the other countries is similarly divided.

USA Participants	Over 10,000
Age	18 to 95 years
Gender	45% female / 55% male
Pathology	20% pathalogical / 80% healthy

Table 2 - Statistics regarding the data collection

The collected data set was then used to produce reference distributions for each of the system's speech/ voice parameters. Those are used to identify anomalies in a subject's voice/speech. An example is shown in Figure 2, which shows the range of variation in the parameter 'log jitter'. That is a measure of the relative variation in timing between consecutive openings of the vocal folds. Unusually high values of that parameter can indicate the presence of voice pathology.

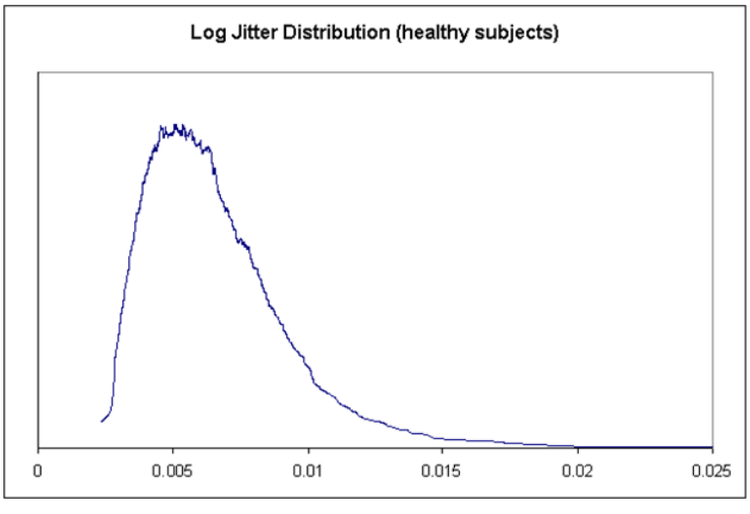

Figure 2 - Log jitter distribution

Use case example cont

The data can also be presented as cumulative distributions as in Figure 3, allowing a speaker to be characterised numerically in terms of percentiles of the healthy population.

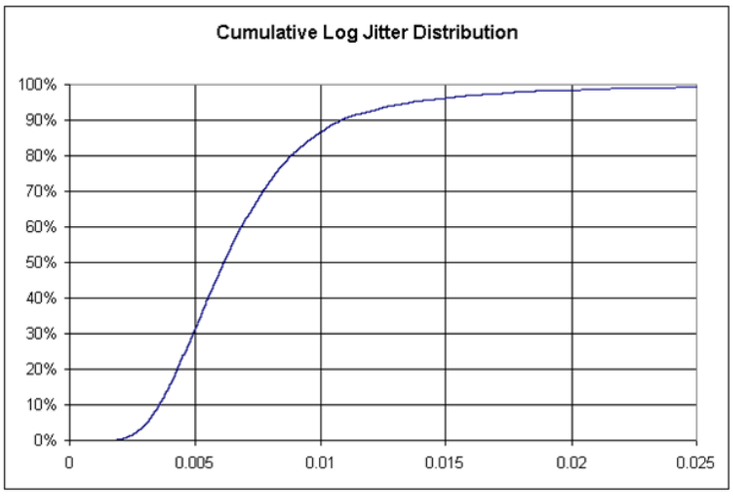

Figure 3- Cumulative log jitter distribution

In the cumulative log jitter plot shown in Figure 3, a log jitter value of 0.015 corresponds to a value beyond the 95th percentile, and would therefore be worthy of consideration for further tests.

In-depth examples

Examples

Taking Parkinson's Disease as an example, Figures 4 and 5 show a comparison of two, typical, 59 years old, male speakers, with perceptually similar voices. Those were recorded as part of the data collection. One participant was healthy and the other suffered from Parkinson's Disease. The voice characteristics of the sustained phonation, "aah", are apparent in the waveforms below:

Healthy

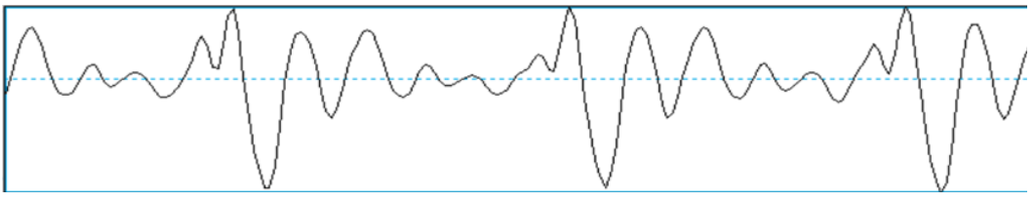

Figure 4- Healthy Male speaker

With Parkinson's Disease

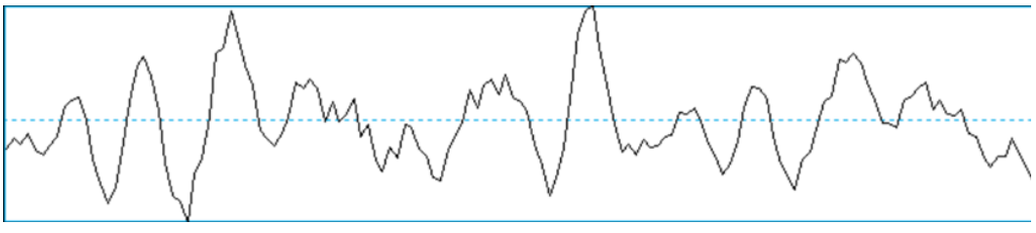

Figure 5 - Male speaker with Parkinson's disease

The healthy speech is very regular. In terms of voice parameters, there is little variation in amplitude between one repetition and the next (low 'shimmer'), and each repetition has very nearly the same duration (low 'jitter'). The detailed shape of the waveform is also very similar from one repetition to the next, indicating a high HNR. In contrast, the speech from the Parkinson's sufferer is very irregular in all these respects, with high values of jitter and shimmer, and a low HNR.

In-depth examples

Examples

As well as the voice characteristics apparent during sustained phonation, the dynamics of a subject's speech are also affected by Parkinson's Disease. Figure 6, compares the same two speakers, saying the first part of the phrase "my grandfather…". The darkness at each point denotes the intensity of the sound at the corresponding time and frequency.

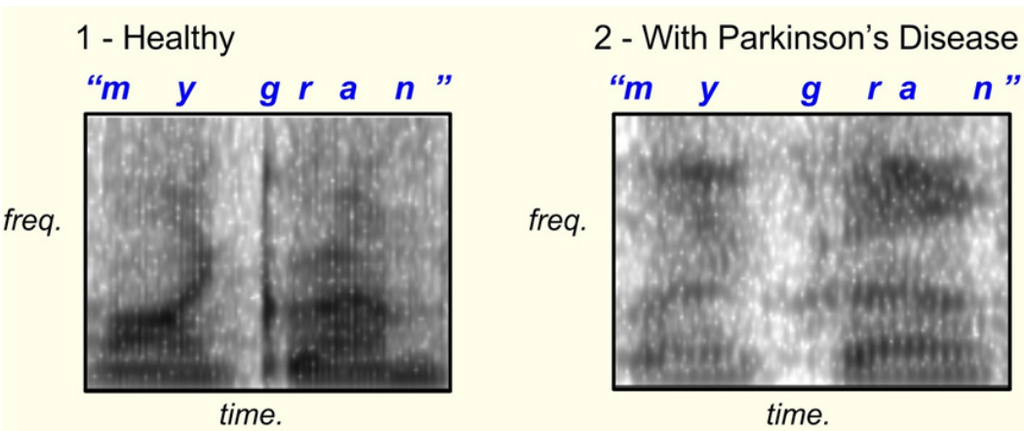

Figure 6 - Comparison of two speakers

The quality of the phonemes is much lower in the subject with Parkinson's Disease, and in particular the 'glide' at the end the word "my", and the "g" of "grandfather", are only partially articulated. The respective regions of different phoneme quality are circled in blue, in Figure 7:

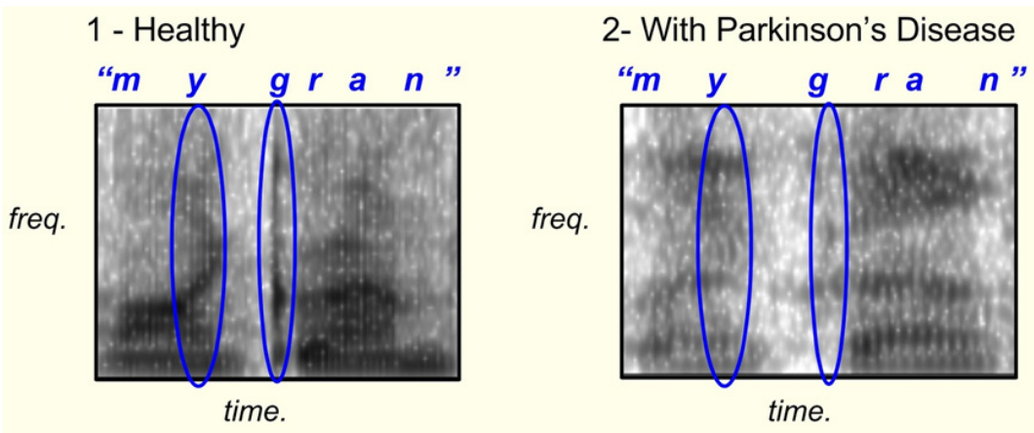

Figure 7 - Comparison of two speakers

In-depth examples

In the images in Figure 8, the red lines indicate the boundaries between predictable segments. The healthy subject has more segments and there is more change within each segment.

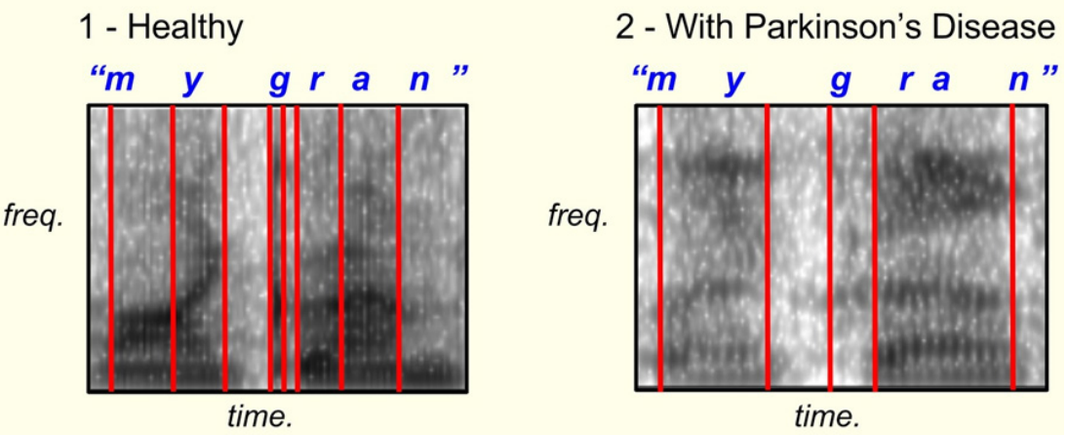

Figure 8 - Predictable segment differentiation

The system's analysis quantifies these high-level aspects of speech dynamics, providing the clinician with insights into the range and precision of articulation, as well as providing more fundamental speech parameters, such as phoneme quality, and the lower-level (voice) parameters mentioned previously: jitter, shimmer, HNR, etc.

Recommendations

Recommendations

When considering a system such as the one proposed, developers offering specialist IT services to the healthcare service provider market e.g., PMA and EHR systems, will no doubt readily come to appreciate that a sizeable opportunity exists. Should they determine to seek advantage from that opportunity, through enhancing their service offering and developing suitable applications for healthcare service providers, they will need to evaluate the core technology building blocks needed to produce a viable end user system.

Table 3 is offered as a quick reference guide or checklist of to what to look for in a clinical speech analysis system.

Clinical speech analysis system recommended requirements	Check
Cloud-based platform with a suite of easily programmable APIs	
Scalable, cost-effective, and globally accessible CPaaS – 24x7x365(6)	
Inbound (and outbound) calling	
IVR and self-service functionality	
Text-to-speech synthesis (for outpatient user prompts)	
Call recording capability	
SMS and 2-factor authentication	
Fax transmit and receive	
HIPAA and HITECH compliant system	
Large database of baseline speech reference data	
Automatic speech recognition	
Automatic, objective speech analysis producing quantitative data	
Advanced algorithms for analysis of movement of vocal cords and larynx	
Measures over 170 voice and speech parameters	
Measures standard deviation of pitch (rather than max and min values)	
Additional measures e.g, speech dynamics, phoneme duration and validity scores	
Detection accuracy close to 90%.	
Comparable parameters correlate to MDVP and OperaVOX at Pearson 0.9 or better	
Analysis useable as input to machine learning system	
Immune to background noise	
Robust against variations in recording environment and communication channel	
Ready access via any form of telephony end-point	
Calls can be scheduled and as frequent as necessary	
No special hardware, training, or travel is required	
Reminders and confirmations issued by SMS or voice message	
Instant results via: email; fax; or SMS	
Download option for: patient data; histograms; graphs; charts; and recordings	

Table 3 - Requirements check list

White Paper

About Aculab

More knowledge, more choice, more innovation

Aculab is an innovative company that offers deployment proven technology for any telecoms related application. Its enabling technology serves the evolving needs of automated and interactive systems, whether on-premise, data centre hosted, or cloud-based.

Over 1000 customers in more than 80 countries worldwide, including developers, integrators, and solutions and service providers, have adopted Aculab's technology for a wide variety of business critical services and solutions.

Aculab offers development APIs for voice, data, fax and SMS, on hardware, software and cloud-based platforms, giving a choice between capital investment and cost-effective, 'pay as you go' alternatives.

.

For more information

To learn more about Aculab Cloud and Aculab's extensive telephony APIs, visit:

www.aculab.com

EMEAA
+44 (0) 1908 273802 | sales@aculab.com

AMERICAS
+1 (781) 352 3550 | sales@aculab.com

References

1. See table 2.5 of the Motor Speech Disorders: Clues to Neurologic Diagnosis chapter by Duffy in Parkinson's Disease and Movement Disorders - Diagnosis and Treatment Guidelines for the Practicing Physician, published in 2000, for an indication of what type of conditions are associated with each measure, including different forms of Dysarthria and Apraxia of speech.

2. Vendor product information can be found at: http://pentaxmedical.com/pentax/en/99/1/ENT-Speech

3. Such systems calculate numerical measures for a subset of the parameters listed in table 2.5 of the Motor Speech Disorders: Clues to Neurologic Diagnosis chapter by Duffy in Parkinson's Disease and Movement Disorders - Diagnosis and Treatment Guidelines for the Practicing Physician, and display them in a graphical form.

4. See the early section of Parkinson's Disease and Movement Disorders - Diagnosis and Treatment Guidelines for the Practicing Physician for a description of the overall process that a speech and language pathologist (SLP) or therapist (SLT) will go through to assess their subjects. See also Reliability of perceptions of voice quality: evidence from a problem asthma clinic population, by Sellers et al, and published in 2009 in The Journal of Laryngology & Otology.

5. Information on OperaVOX is available at: http://www.operavox.co.uk/wp/

6. From the 2016 Healthcare Cloud Survey; research and analysis carried out by HIMSS Analytics and sponsored by Level 3.

7. See the Aculab Cloud case study Aculab and eClinicalWorks – Building bridges to better healthcare at: http://www.aculab.com/about/news-events/

8. See Automatic Speech Signal Analysis For Clinical Diagnosis And Assessment Of Speech Disorders by Ladan Baghai-Ravary and Steve W. Beet, published in 2013.

9. As example, see Diagnostic Assessment of Childhood Apraxia of Speech Using Automatic Speech Recognition (ASR) Systems, by Shriberg et al and published in 2003.

For more information, visit Aculab's website at www.aculab.com
Aculab plc,
Lakeside, Bramley Road,
Mount Farm,
Milton Keynes.
Buckinghamshire.
MK1 1PT. United Kingdom.

Tel: +44 (0) 1908 273 800 Fax: +44 (0) 1908 273 801