Abstract: The details in the formation of the primary acoustic sources in voice production during phonation are not yet fully understood. Some acoustic sources are due to the unsteady flow evolving between the vocal folds, where a jet develops. The glottal jet flow downstream the vocal folds features characteristics which depend on the physiological conditions, e.g. the parameters in geometry, kinematics, material and fluid. A driven mechanical model of the vocal folds is used with the aim to study the flow details and the acoustic sources in the glottal jet. Numerical simulations and experimental measurements of the flow are carried out. The results show topological characteristics of the glottal jet flow. When prominent ventricular folds are included in the vocal folds model the jet evolves differently due to the interaction with these supraglottal structures. They lead to a changed distribution and abundance of the flow acoustic sources and changed spectral properties of the flow close to the glottis.

Keywords: mechanical vocal folds model, ventricular folds, flow simulation, higher harmonics, spectral analysis

I. INTRODUCTION

The production of the voice is a complex process, which is influenced by a wide range of parameters [1]. In general voicing is a more or less coupled process of fluid-structure-acoustic interaction. The singing and phonation regimes differ quite strongly from each other. Due to its complexity models of the respiratory system, in particular the trachea, the glottis and the vocal tract are generated in order to reduce the problem study to specific voicing aspects. The “pressed” and “breathy” voicing types indicate the importance of the detailed knowledge of the generation of the primary voice source. Singing is a strongly coupled problem and represents a special area in the voicing research with regard to professional singers; whereas phonation is more amenable and has a wider range of application in everyday life. Extensive investigations are nowadays carried out with the aim of tackling voice disorders in phonation. The generation of the primary acoustic signal at the glottis is the first link in the chain of voice production. Herein, the overall output voice signal spectrum is partly influenced by the nature of the flow field downstream of the glottis. Several models exist for the investigation of the primary voice sources: theoretical / lumped mass models, computational fluid dynamics (CFD) models, and mechanical models [2]. These are also classified with regard to the degree of idealization into static, dynamic driven and self-oscillating or 1-dimensional, 2-dimensional / axisymmetric or 3-dimensional (3-D) models. Most of these models incorporate only a very simplified geometry of the vocal tract.

The time-dependent 3-D flow field in a driven vocal folds model [3], which considers the 3-D shape change of the glottis during the cycle of phonation, is considered in the present paper.

The description of 3-D effects in vortex dynamics such as stretching and bending of vortex lines and the determination of the local pressure and velocity are of immense importance for the spectral characterization of the flow field. The temporal evolution of large and small scale flow structures including their interaction with supraglottal walls may change the spectral fingerprint of the flow field. Therefore, flow effects at different driving pressures, changed glottal and supraglottal configurations can be explored with regard to the resulting output flow patterns. In order to combine as best as possible their inherent advantages, experimental and numerical methods of flow investigation are applied simultaneous.

II. METHODS

The main physiological parameters of real vocal folds kinematics during phonation are replicated. The characteristic movement of the walls in the glottal region, e.g. the continuous deformation of the mucosal layer is achieved by means of two 3D contoured cams, which rotate in counter-direction [3]. Similarity of geometry, flow dynamics and fluid dynamic forces is kept in the model. Due to the low Mach number, the flow can be treated as incompressible. The time-dependent and 3D nature of the flow field downstream of the glottis requires flow analysis methods with appropriate temporal and spatial resolutions.

The experimental model of the vocal folds is shown in Fig. 1. The cams which are covered with a membrane can be seen on the left hand side. On the right hand side of
the photograph supraglottal elements e.g. ventricular folds (VFs) which are optionally inserted into the test section downstream of the vocal folds model are indicated. Two variations of models of the VFs are available: first, rigid transparent models allowing distortion-free optical access into the flow field; second, models with a compliant surface layer and incorporating an air-cushion. The latter are used for pressure sensing the higher harmonics from the glottal jet flow or for selective activation of the compliant surface layer.

![Image of vocal folds model](image1)

**Fig. 1:** Photograph of the model of the vocal folds.

In the experiments a global driving pressure head across the glottal orifice is imposed. The close-to-glottis flow-dependent pressure difference is recorded. The well established method of high speed flow visualization is used for accurate determination of temporal and spatial velocity information in selected measuring planes. The resulting glottal volume waveform is measured and given as input for the inlet boundary condition in the numerical model.

The Navier-Stokes equations for incompressible fluid flow are discretized with the Finite Volume method and solved numerically with the open source CFD code OpenFOAM. A block structured mesh of 1 million cells with variation in time according to the glottal kinematics is implemented. The solver uses a second order Crank Nicolson time stepping and as space discretization a second order TVD (total variation diminishing) scheme. The full transient 3-D flow field in the near-glottal region is simulated. The subgrid-scale turbulence is modeled implicitly. All simulations are carried out with the volume waveform synchronous to the imposed time-varying motion of the 3D glottis model.

### III. RESULTS

**A. Flow structures**

Fig. 2 shows the experimental visualization of the flow in the mid-coronal plane at the maximum opening instant \( t/T_0 = 0.25 \) of the glottal cycle. The case without (top) and with rigid (bottom) VFs has been studied. The character of the near field of the emerging glottal jet with its most energetic large coherent vortex structures is shown. Kelvin-Helmholtz instabilities are responsible for the roll-up of the jet edge. The jet front and the successive vortex structures are seen to interact with the VFs. The determination of the pressure fluctuations due to the jet edge instabilities is given in section C.

![Image of flow visualization](image2)

**Fig. 2:** Visualization of the flow in the mid-coronal plane for transglottal pressure of \( \Delta p = 6 \text{ cmH}_2\text{O} \) at maximum opening instant \( t/T_0 = 0.25 \) of the glottal cycle for two supraglottal configurations (a) and (b).

Further flow results are shown from the numerical simulations which resolve the full 3D flow field in space and time in the glottal model. A preliminary study on resolution requirements, accuracy and convergence of the model has been carried out in [4]. There exist several velocity or pressure based tools for vortex detection. In Fig. 3 the Q criterion [5] is used to illustrate the 3-D vortex structures. Elliptic vortex rings are generated at the glottal orifice. These are strongly deformed due to self-induction, interaction among each other and with the supraglottal walls, e.g. VFs.
Fig. 3: Representative isocontour of the Q-criterion of the flow field at the divergent closing instant $t/T_0 = 0.35$ of the glottal cycle for both supraglottal configurations (a) and (b) from Fig. 2 at the transglottal pressure of $\Delta p = 6$ cmH$_2$O.

The complex 3D unsteady vortex structures which are generated downstream of the glottal orifice are already subjected to break-down.

B. Primary acoustic sources

The distribution of the divergence of the Lamb vector $L$ can be computed from the velocity $u$ of the flow field and appears as a dominant acoustic source term in Lighthill’s wave equation [6]. The source term reads

$$\nabla \cdot L = \nabla \cdot ((\nabla \times u) \times u) \quad . \quad (1)$$

One example of this distribution is shown in Fig. 4.

C. Power spectrum

In Fig. 5 the normalized power spectra of the flow field velocity from the numerical simulation is compared for the cases without and with VFs in a center point at a downstream distance from the glottis corresponding to one vocal tract height. The differences are considerable.

In order to determine the pressure fluctuations of the jet edge instabilities in experiment, a VF has been replaced with a VF model with a compliant surface layer and air cushion. The small amplitude of the pressure fluctuations poses a challenge in the measurement and analysis of the data. Fig. 6 clarifies the actual situation. The frequency spectra of two pressure measurements are shown in dimensionless form with the help of a Strouhal number $Sr$ defined as

$$Sr = f \cdot \frac{w}{u_{\text{mean}}} \quad (2)$$

where $f$ is the frequency content of the pressure signal, $w$ is the maximum width of the glottal gap and $u_{\text{mean}}$ is the mean velocity in the glottal gap. On top the spectrum results from the reference pressure upstream of the vocal folds model. The spectrum below results from the integral pressure measured in the air-tight air cushion. The difference of both spectra yields the Strouhal number with a value of 0.27, which is supposed to be due to the shear layer instabilities interacting with the walls. This value is highlighted by the arrow in the spectrum and it correlates well with the frequency value from the numerical simulation in Fig. 5.
IV. DISCUSSION

The change in the vortex dynamics and the spectra at different supraglottal configurations are clearly shown. A negative slope of 3 dB per octave in the low frequency range up to the 10th harmonic is extracted the spectra in Fig. 5. Due to the jet edge interaction with the ventricular folds the higher frequency range of the spectrum in configuration (a) differs considerably from that in configuration (b).

V. CONCLUSION

Prominent ventricular folds leave a strong fingerprint in the spectra of the flow close to the glottis. The ventricular folds redirect part of the displacement flow into the lateral gap of the Morgani space which seems to stabilize the jet core at the exit of the glottis. In addition, the shear-layer roll-up is affected by the presence of the folds and vortex structures are interacting with the walls in this region. As a consequence, vortex dynamics and wall interaction is changed considerably when supraglottal structures are included in the models. These effects are well seen in the change of spectral content within the flow. Further studies in our lab now concentrate on possible feedback and jet-control by passive and active excitation of the ventricular folds wall.

REFERENCES

Abstract: Synthetic models are used to study vocal fold flow-induced vibration. Advantages include reproducibility and vibration frequencies typical of human phonation. Limitations of recent models include lack of a mucosal wave, excessive inferior-superior motion, and limited convergent-divergent motion. To overcome these limitations, a synthetic vocal fold model was developed that included separate epithelial and lamina propria layers. A corresponding finite element model was developed. High-speed imaging was used to quantify synthetic model motion, including videokymography and determination of three-dimensional marker trajectories. Both models exhibited similar characteristics in terms of vibration frequency (around 115 Hz) and maximum glottal width (just under 2 mm). The synthetic model onset pressure was 0.4 kPa, which is significantly lower than many previous synthetic models. These values are consistent with human phonation. Importantly, in both models mucosal wave-like motion was evident and alternating convergent-divergent intraglottal profiles were seen. These advantages will be useful in future experiments and simulations by providing synthetic models that exhibit more life-like response and motion. The two models are described, data are presented, and suggestions for future work are provided.

II. METHODS

A. Synthetic Model

The synthetic model geometry is shown in Fig. 1. Silicone interior layers were fabricated according to the multi-layer rapid prototyping, molding, and casting procedures described in [5,7]. The epithelial layer was created by pouring a silicone mixture over the assembled interior layers. The epithelial layer thickness was estimated to be approximately 0.1 mm. Layer Young’s modulus were controlled by varying the pre-cured silicone mixture content; values for the different layers were approximately 11.8 kPa (body), 1.6 kPa (ligament), 0.2 kPa (superficial lamina propria), and 49.8 kPa (epithelium). Tension was applied to a fiber thread that ran anteriorly-posteriorly within the ligament layer to reduce inferior-superior motion. High-speed video imaging (Photron SA3, 3000 frames per second) was used to capture model motion.
B. Finite Element Model

The finite element model consisted of two-dimensional, fully-coupled fluid and solid domains, as shown in Fig. 2. The solid model incorporated the same geometry as the synthetic model, but with a 50 μm-thick epithelium. The material properties were also the same, with the exception that the superficial lamina propria layer material property was based on a nonlinear stress-strain curve. This curve was governed by the equation

$$\sigma(\varepsilon) = 11.2(e^{10.5\varepsilon} - 1),$$  \hspace{1cm} (1)

where $\sigma$ is stress (Pa) and $\varepsilon$ is strain. This yielded a tangent modulus of 200 Pa at $\varepsilon = 0.05$ and 972 Pa at $\varepsilon = 0.2$. The finite element model did not include a fiber.

The fluid model used an incompressible, viscous, 2D, unsteady Navier-Stokes solver with a constant 600 Pa inlet pressure. The solid domain allowed for large strain and large deformation and included Rayleigh damping ($\alpha = 101.67$, $\beta = 0.0001073$) for energy dissipation.

Solution was accomplished using the commercial code ADINA. A time step size of $10^{-4}$ and a second-order composite time marching scheme were used. For computational efficiency, medial-lateral symmetry was assumed. The fluid domain consisted of 7340/7641 1st-order elements/nodes and the solid domain consisted of 2359/2582 1st-order elements/nodes (see Fig. 3). A solution for 1500 time steps required approximately 3.2 hours on a single 2.53 GHz Intel P9500 processor.

III. RESULTS

The synthetic model had an onset pressure of 400 Pa. At a pressure 20% above onset pressure (480 Pa), the vibration frequency was 114.5 Hz and the maximum glottal width was approximately 1.8 mm. These values compare well with those of human phonation.

Importantly, mucosal wave-like motion was evident and the inferior-superior motion appeared to be lower than with previous two-layer models. To capture this wave-like motion, a hemilarynx configuration and two synchronized high speed cameras (Photron SA3, 3000 frames per second) were used to track the medial surface position in a manner similar to that described in [8]. One sample image is shown in Fig. 4 in which ink dots placed on the model surface are visible. The medial-lateral trajectories (three-dimensional positions) of the dots in the center column were tracked over several oscillation periods, as shown in Fig. 5. A wave-like motion clearly propagated superiorly along the medial surface, and an alternating convergent-divergent medial surface profile was visible. Evidence of this convergent-divergent motion can also be seen in the kymogram shown in Fig. 6 (obtained using a single high-speed camera and a full larynx configuration).
The finite element model vibrated at 116 Hz with a maximum glottal width of 1.9 mm, which compares well with the synthetic model response. Glottal width vs. time is shown in Fig. 7. Steady-state vibration was achieved around 0.05 s. Still images of model motion are shown in Fig. 8, in which mucosal wave-like motion is evident.

IV. DISCUSSION

The synthetic and computational models exhibited similar characteristics in terms of vibration frequency and amplitude. Some differences in motion were observed; for example, unlike the synthetic model, the numerical model did not experience complete glottal closure during vibration. Differences in motion of the two models were attributed to four factors: differences in material properties (stress vs. strain relationships, Poisson’s ratio, and damping coefficients), three-dimensionality of the synthetic model versus two-dimensionality of the finite element model, difference in thickness of the epithelial layer, and presence of an anterior-posterior fiber in synthetic model.

V. CONCLUSION

Complementary synthetic and finite element models of the vocal folds have been developed and tested. The models were based on the same multi-layer geometry. Each included a cover layer that was comprised of a thin epithelial layer and a very flexible layer that was similar to the superficial lamina propria. Each also included
ligament and body layers, and the synthetic model included a fiber imbedded within the ligament layer.

In both models mucosal wave-like motion was evident. Alternating convergent-divergent intraglottal profiles were also seen. The vibration frequencies and glottal amplitudes were typical of adult human male phonation. Further, the synthetic model had an onset pressure that was much lower than previous models and that is comparable to human phonation. These advantages will be useful in future experiments and simulations by providing models that exhibit more life-like response and motion.

For both models future work includes the use of anisotropic materials. Incorporation of a downstream duct (to simulate the vocal tract) in the synthetic model will also be important. For the finite element model, future work includes performing extensive numerical verification studies (e.g., ensuring that the solutions are independent of grid density and time step size), extending the model to three dimensions, and removing the symmetry condition. The latter will enable the study of asymmetric aerodynamics and vocal fold properties. Potential future work also includes investigation of the influence of epithelial layer thickness and of the material properties of the different layers on model response.

VI. ACKNOWLEDGEMENTS

This study was supported by Grants R01DC9616 and R01DC5788 from the U.S. National Institute on Deafness and Other Communication Disorders (NIDCD). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIDCD or the National Institutes of Health.

REFERENCES

Session III:
Signal analysis
AUTOMATED TOOLS FOR IDENTIFYING SYLLABIC LANDMARK CLUSTERS THAT REFLECT CHANGES IN ARTICULATION

Suzanne Boyce¹, Harriet Fell², Lorin Wilde³, and Joel MacAuslan³

¹University of Cincinnati, OH, ²Northeastern University, Boston, MA, ³Speech Technology and Applied Research, Bedford MA, USA

Abstract: We have developed a set of software tools to detect articulatory changes in the production of syllabic units based on acoustic landmark detection and classification. Results from the application of this automated analysis system to studies of Parkinson’s Disease and Sleep Deprivation show the ability to detect subtle change. We are making these tools available as add-ons to systems such as Wavesurfer and R.

Keywords: speech-acoustic landmarks, syllabic landmark cluster, automatic vocalization processing.

I. INTRODUCTION

Acoustic evidence provides information on speech production, but that information is scattered across multiple frequency bands and multiple time scales. Landmark analysis [5,6] is one approach by which acoustic patterns characteristic of particular changes in speech movements are detected. In this paper, we describe an extension of the landmark method to the detection of articulatory complexity in the production of syllables, by using clusters of landmarks as a measure of whether a string of (intended) syllables is produced in its canonical form (dictionary pronunciation), in a less complex (CCVC -> CVC), or more lenited form (softened consonants). We refer to this measure as a syllabic landmark cluster measure. We have applied this approach successfully to measure speech articulation changes in Parkinson’s Disease, in infant speech development, in sleep deprivation, and other studies.

The notion of syllabic complexity is illustrated as follows. A word such as “interesting” can have four syllables in its canonical form, but when uttered as /ɪnˈɜːrɛstɪŋ/ it has three syllables with fewer consonants, and thus reduced articulatory complexity. In landmark systems in general, different types of types and combinations of speech sounds are detected as different patterns of landmarks. In our particular system, a syllabic landmark cluster is a sequence of consecutive landmarks grouped according to specific rules. For example:

1. A syllabic cluster must contain a voiced region of at least 30 ms, corresponding to a syllable nucleus.
2. A noisy sound such as “s” (/s/) must hit a threshold of loudness before being detected.

If uttered in a canonical fashion, the pronunciation of a word will show a characteristic pattern of landmarks for each syllable in that word. As long as the syllables are uttered with the same acoustical characteristics, our measures will detect the same pattern of landmarks. However, if the syllables are uttered less canonically—perhaps with less extreme articulatory movements, less precise timing, or reduced aerodynamic support—then fewer landmarks will be detected. Our version of the speech-acoustic landmark system thus can be used to detect two common effects in speech production: (1) simplification of syllable onsets (e.g. “string” /strɪŋ/ as /shrʊŋ/), nuclei (e.g. “diamond” /dəˈmaɪnd/ as /dəˈmaɪnd/) and rimes (e.g. “pelt” /pɛlt/ as /pɛl/, and (2) fewer uttered syllables.

II. METHODS: LANDMARK SYSTEM

Landmarks and Rules: Our landmark analysis system is based on Stevens et al. [6], especially as developed by Liu [5] and Howitt [4]. The speech signal is automatically partitioned into 5 frequency bands plus a voicing-status contour. Abrupt landmarks are identified as points where abrupt changes in the amplitude of several frequency bands coincide in a specified pattern [5,6]. These landmark patterns are identified by comparison between “coarse” and “fine” temporal resolution.

The system detects the following types of landmarks:

1. g: glottis. Marks a time point at which voicing begins (+g) or ends (-g), based on the harmonic spectrum.
2. s: syllabic. Marks sonorant consonantal releases (+s) and closures (-s).
3. b: burst. Marks frication onsets or affricate/stop bursts (+b) and points where aspiration or frication ends (-b) due to a stop closure.
4. V: vowel. Marks a time point corresponding to maximum harmonic power.
The +/- b and +/- s “abrupt” landmarks are identified from patterns of rapid change in the amplitude of several frequency bands. The +/-g and V landmarks are identified from the harmonic spectrum.

This system makes no attempt to identify phonemes, but it is sensitive to broad categories of speech sounds and to aspects of metrical structure. The features it detects are those known as “articulator free” [6] because they are independent of the specific articulator used to produce the segment. These features are instead associated with creation and release of constrictions in the vocal tract and with the acoustic consequences of those constrictions and releases.

An example of how abrupt landmarks are determined from patterns across frequency and voicing bands is shown in Fig. 1. An example of landmark location in the speech signal can be found in Fig. 2, which shows a spectrogram of the nonsense word /pʌtəkə/ repeated 10 times in two breath groups by a native speaker of American English with moderate dysarthria due to Parkinson’s Disease.

Figure 1. Spectral analysis of an utterance: voicing (bottom) and five frequency bands’ energy waveforms. (a) Too few bands show large, simultaneous changes in energy. (b) All bands show large, simultaneous energy increases immediately before the onset of voicing, identifying a +b (burst) landmark. (c) All bands show large, simultaneous energy increases during ongoing voicing, identifying a +s (syllabic) landmark.

Figure 2. Ten repetitions of /pʌtəkə/ by an American English speaker with moderate dysarthria due to Parkinson’s Disease. Vertical lines above the waveform pane show +/- b, +/-s and +/- g landmarks. Vertical lines below the waveform pane show Vowel landmarks as V. The period of silence shows the pause between breath groups.

Use of the Landmark System to Characterize Differences in Speech Production: The landmark system operates with empirically derived threshold values. As discussed, abrupt landmarks are determined by the patterns of abrupt change across frequency and voicing bands; if the amplitude value of the signal in a particular set of frequency and voicing bands meets the predetermined threshold for abruptness, then a landmark is detected. If the amplitude value of the signal in any of the frequency/voicing bands does not meet this criterion, then no landmark is detected.

The operation of this system is shown by the pattern of V landmarks in Fig. 2. As noted above, the speaker produced /pʌtəkə/ in two breath groups; the first seven repetitions belong to the first breath group, and the
following three repetitions belong to the second breath group. This speaker showed a tendency to dysphonia typical of Parkinson’s patients, characterized subjectively as causing a harsh and breathy voice, and the dysphonic phonation was more marked in the late portions of a breath group—presumably because reduced breath support made it more difficult to sustain normal periodic vocal fold vibration. Because the V landmarks are computed on the basis of harmonic power, and dysphonic vowels are produced with less harmonic power, fewer V landmarks will be detected on dysphonic voices. This effect is shown in Fig. 2, where the first few repetitions in the first breath group are marked with V landmarks on the stressed syllable, while the last few repetitions in the same breath group show that no such landmarks have been detected. Note that these repetitions were produced with vowels—this is evident in the spectrogram—but the vowels had too little harmonic power to be registered as V landmarks.

Grouping Landmarks to Characterize “Syllabic Clusters”: Fell & MacAuslan originally developed the syllabic cluster measure to detect the increasing syllabic complexity of utterances by young children [2, 3]. More recently, we have applied this method, termed the Syllabic Cluster analysis, to speech uttered under normal and sleep-deprived conditions, and to speech by Parkinson’s Disease patients undergoing Deep Brain Stimulation (DBS) therapy.

Cluster Rules: The Syllabic Cluster analysis works by grouping sequences of detected landmarks into clusters that roughly correspond to syllabic units in the acoustic speech signal. The grouping rules include categorical dependencies as well as dependencies of timing, and were empirically determined from datasets of speech.

For example, one such rule states that a gap of 30 ms in voicing, with whatever [+b]’s immediately follow it, identifies a type of syllable cluster endpoint. In contrast, burst-like noise that does not occur within 120 ms before a voiced region, or 80 ms after, is not part of a cluster—presumably because reduced breath support made it more difficult to sustain normal periodic vocal fold vibration. Because the V landmarks are computed on the basis of harmonic power, and dysphonic vowels are produced with less harmonic power, fewer V landmarks will be detected on dysphonic voices. This effect is shown in Fig. 2, where the first few repetitions in the first breath group are marked with V landmarks on the stressed syllable, while the last few repetitions in the same breath group show that no such landmarks have been detected. Note that these repetitions were produced with vowels—this is evident in the spectrogram—but the vowels had too little harmonic power to be registered as V landmarks.

Grouping Landmarks to Characterize “Syllabic Clusters”: Fell & MacAuslan originally developed the syllabic cluster measure to detect the increasing syllabic complexity of utterances by young children [2, 3]. More recently, we have applied this method, termed the Syllabic Cluster analysis, to speech uttered under normal and sleep-deprived conditions, and to speech by Parkinson’s Disease patients undergoing Deep Brain Stimulation (DBS) therapy.

Cluster Rules: The Syllabic Cluster analysis works by grouping sequences of detected landmarks into clusters that roughly correspond to syllabic units in the acoustic speech signal. The grouping rules include categorical dependencies as well as dependencies of timing, and were empirically determined from datasets of speech.

For example, one such rule states that a gap of 30 ms in voicing, with whatever [+b]’s immediately follow it, identifies a type of syllable cluster endpoint. In contrast, burst-like noise that does not occur within 120 ms before a voiced region, or 80 ms after, is not part of a cluster. Indeed, we have found it useful to designate these types of isolated bursts as non-speech noise. The syllabic grouping procedures are described in more detail in Fell et al. [2,3] and Boyce et al. [1]. The following is a list of examples of some common types of syllabic cluster that occur in speech:

• (+g,-g)- singleton V [vowel] or CV [consonant-vowel] syllables, where C is voiced;
• (+g,-s) - V or voiced-CV syllables followed by a sonorant consonant and syllabic cluster;
• (+s,-g) - V or voiced-CV syllables, preceded by a syllabic cluster;
• (+g,-s,-g) - VS syllable, where S is a sonorant consonant or voiced obstruent adjacent to the +g or -g;
• (+b,+g,-g) - syllable beginning with fricative: (+b) marks the presence of frication;
• (+b,-b,+g,-g) - syllables with an initial plosive: (+b , -b) mark the beginning and end of the release.

III. METHODS: APPLICATION

Parkinson’s Disease Study: In one study using the Syllabic Cluster measure, we contrasted speech as produced by Parkinson’s Disease (PD) patients who were receiving Deep Brain Stimulation (DBS). In the typical progression of Parkinson’s Disease, patients show clinically significant levels of unintelligible speech later than they show gross motor symptoms. Thus, patients in DBS programs may not be showing clinically overt signs of dysarthric speech. However, the application of DBS therapy can sometimes cause their speech intelligibility to worsen, and this is both a matter of clinical concern and scientific interest. The data described in Fig. 3 come from a study of 12 Control vs 15 PD patients who had undergone surgery for Deep Brain Stimulation (DBS) repeating the syllable /ka/. The aim of the study was to detect subtle and/or overt changes in speech production when DBS stimulus was OFF vs. ON.

Sleep Deprivation: In another study, we used the Syllabic Cluster analysis to test whether speech articulation changes as a result of sleep deprivation. Studies of both speech articulation per se, and listener perceptions of change, have shown conflicting results to date [1]. In our study, the speech of 17 speakers of American English (9 female, 8 male) was recorded at 8 hour intervals over 32-40 hours without sleep. (Not all subjects completed the final session.) Subjects read aloud the Rainbow Passage each time. To control for the possible effect of familiarity with the speech materials, another set of 15 subjects (7 male and 8 female) read aloud the Rainbow Passage at 8-hour intervals while maintaining a normal sleep schedule.

III. RESULTS AND DISCUSSION

Parkinson’s Disease Study: The mean cluster rate in rapid repetitions of the syllable /ka/ decreases (a) between Control vs. PD speakers, and (b) as a result of DBS. The differences were significant at the .01 level.

Sleep Deprivation: The first two sessions were combined as the Early, or Rested, condition. The last two sessions were combined as the Late, or Sleep Deprived condition. As Fig. 4 shows, Syllabic Cluster rate decreased between the Early and Late sessions. This difference was significant at the p < .05 level by a binomial (sign) test. In contrast, the Early vs. Late sessions were not significantly different for speakers performing the identical task while following their normal sleep schedule (p > .10 by a binomial (sign) test.)
VI. CONCLUSION

The Syllabic Cluster analysis based on acoustic landmark detection appears to be sensitive to articulatory differences in speech production scattered across multiple frequency bands and multiple time scales. The Parkinson’s Disease results suggest this analysis provides a rough measure of a speaker’s ability to repeat speech materials with a certain level of articulatory precision at a particular speech rate. The Sleep Deprivation results suggest that speech articulation does indeed change with sleep deficit in a way that reduces the rate at which well-formed syllabic clusters are produced and that this change is not due to familiarity with the speech materials. Both sets of results suggest the analysis is sensitive to very subtle changes that listeners do not always detect. The automatic nature of the analysis facilitates evaluation of large amounts of data.

We are currently developing a set of software tools for automatic landmark detection, and classification into syllabic cluster patterns, to be available as add-ons to systems such as Wavesurfer and R.

ACKNOWLEDGEMENT

Support: R43 DC010104, R42 AG033523, R21 HL086689, R42 HD34686 from the U. S. National Institutes of Health.

REFERENCES


