JACC: HEART FAILURE VOL. ■, NO. ■, 2021

Remote Speech Analysis in the Evaluation of Hospitalized Patients With Acute Decompensated Heart Failure

Offer Amir, MD,[a,b] William T. Abraham, MD,[c] Zaher S. Azzam, MD,[d,e] Gidon Berger, MD,[d,e] Stefan D. Anker, MD, PhD,[f] Sean P. Pinney, MD,[g] Daniel Burkhoff, MD, PhD,[h] Ilan D. Shallom, PhD,[i] Chaim Lotan, MD,[a] Elazer R. Edelman, MD, PhD[j]

ABSTRACT

OBJECTIVES This study assessed the performance of an automated speech analysis technology in detecting pulmonary fluid overload in patients with acute decompensated heart failure (ADHF).

BACKGROUND Pulmonary edema is the main cause of heart failure (HF)–related hospitalizations and a key predictor of poor postdischarge prognosis. Frequent monitoring is often recommended, but signs of decompensation are often missed. Voice and sound analysis technologies have been shown to successfully identify clinical conditions that affect vocal cord vibration mechanics.

METHODS Adult patients with ADHF (n = 40) recorded 5 sentences, in 1 of 3 languages, using HearO, a proprietary speech processing and analysis application, upon admission (wet) to and discharge (dry) from the hospital. Recordings were analyzed for 5 distinct speech measures (SMs), each a distinct time, frequency resolution, and linear versus perceptual (ear) model; mean change from baseline SMs was calculated.

RESULTS In total, 1,484 recordings were analyzed. Discharge recordings were successfully tagged as distinctly different from baseline (wet) in 94% of cases, with distinct differences shown for all 5 SMs in 87.5% of cases. The largest change from baseline was documented for SM2 (218%). Unsupervised, blinded clustering of untagged admission and discharge recordings of 9 patients was further demonstrated for all 5 SMs.

CONCLUSIONS Automated speech analysis technology can identify voice alterations reflective of HF status. This platform is expected to provide a valuable contribution to in-person and remote follow-up of patients with HF, by alerting to imminent deterioration, thereby reducing hospitalization rates. (Clinical Evaluation of Cordio App in Adult Patients With CHF; NCT03266029. (J Am Coll Cardiol HF 2021;■:■-■) © 2021 The Authors. Published by Elsevier on behalf of the American College of Cardiology Foundation. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

From the [a]Department of Cardiology, Hadassah Medical Center, Faculty of Medicine, Jerusalem, Israel; [b]Azrieli Faculty of Medicine, Bar-Ilan University, Zfat, Israel; [c]Division of Cardiovascular Medicine, The Ohio State University, Columbus, Ohio, USA; [d]Department of Internal Medicine "B", Rambam Health Care Campus, Haifa, Israel.; [e]The Bruce Rappaport Faculty of Medicine, Technion, Israel Institute of Technology, Haifa, Israel; [f]Department of Cardiology (CVK) and Berlin Institute of Health Center for Regenerative Therapies (BCRT); German Centre for Cardiovascular Research (DZHK) partner site Berlin; Charité Universitätsmedizin Berlin, Berlin, Germany; [g]Section of Cardiology, University of Chicago, Chicago, Illinois, USA; [h]Cardiovascular Research Foundation, New York City, New York USA; [i]Cordio Medical, Or Yehuda, Israel; and [j]Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. Drs Amir and Abraham contributed equally to this paper.

The authors attest they are in compliance with human studies committees and animal welfare regulations of the authors' institutions and Food and Drug Administration guidelines, including patient consent where appropriate. For more information, visit the Author Center.

Manuscript received February 2, 2021; revised manuscript received July 21, 2021, accepted August 19, 2021.

ISSN 2213-1779 https://doi.org/10.1016/j.jchf.2021.08.008

ABBREVIATIONS AND ACRONYMS

ADHF = acute decompensated heart failure

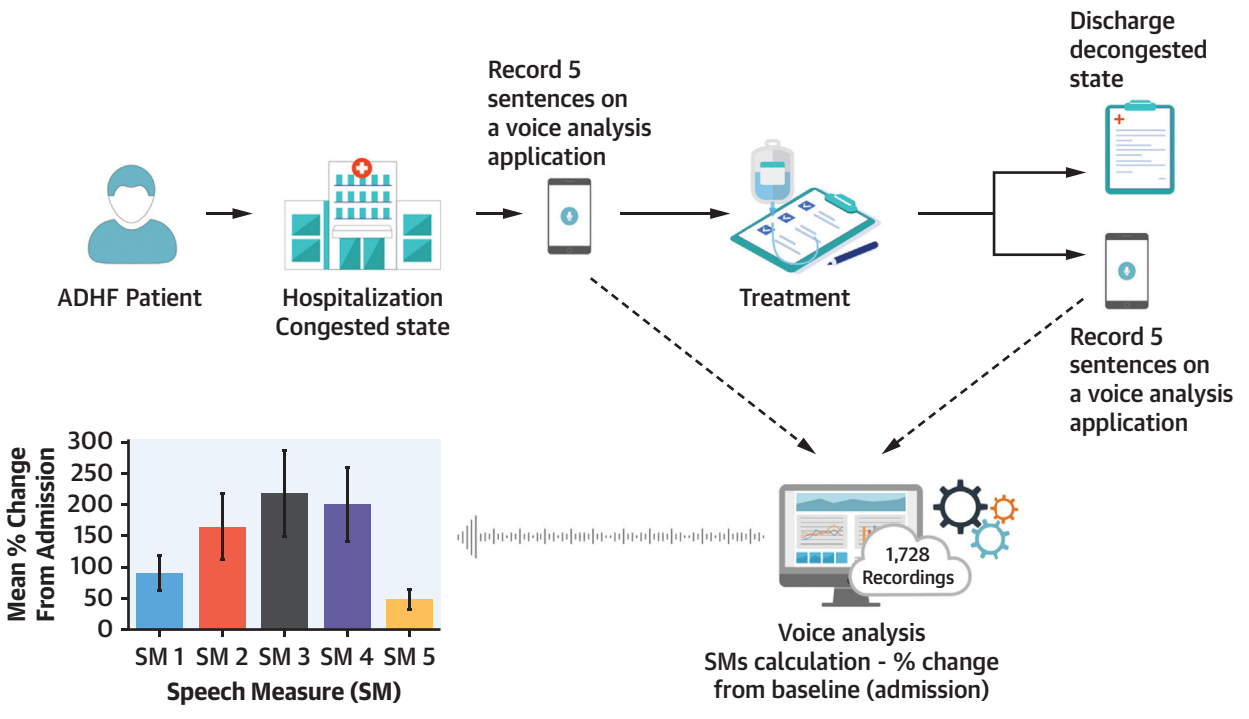
HF = heart failure

SM = speech measure

Heart failure (HF) is a progressive condition that affects approximately 6.5 million adults in the United States alone and 26 million people worldwide (1,2). It is a leading cause of hospitalization in the United States and Europe (3). Most patients with HF present to the hospital with fluid retention, which manifests as worsening dyspnea caused by pulmonary edema (4). Because pulmonary congestion is not only the predominant contributor to HF hospitalization but also a major predictor of poor postdischarge outcomes (3), frequent monitoring for pulmonary congestion has been proposed as a means to keep patients well and out of the hospital (5). However, in its early stages, clinical detection of HF decompensation is often difficult for patients and for clinicians. Various monitoring approaches have been proposed, with variable success (6,7). The Food and Drug Administration–approved implantable wireless pulmonary artery pressure monitoring CardioMEMS HF System (Abbott) significantly improves HF management through pulmonary artery pressure–guided HF management of pulmonary congestion, with reduced patient readmissions and enhanced quality of life (8). Because of its invasive nature, this approach is used in patients with more advanced HF with a recent history of HF hospitalization. A simple noninvasive approach to detecting worsening pulmonary congestion in its early stages presents an attractive value proposition in the management of patients with HF.

Recent progress in speech, voice, and sound analysis enabled the identification of speech features of clinical significance. A variety of speech processing algorithms have been designed to use such features in screening for depression, pneumonia-asthma,

CENTRAL ILLUSTRATION Speech Analysis in Heart Failure



Significant change ($P < 0.0001$) in all five tested SMs (median) between the congested (admission) and decongested (discharge) state

Amir, O. et al. J Am Coll Cardiol HF. 2021;■(■):■–■.

A novel system based on speech processing is demonstrated in this heart failure patient study. It showed an ability to identify voice alterations reflective of changes in acute decompensated heart failure clinical status, showing large changes in speech measures (SMs) between congested (admission) and decongested (discharge) states.

coronary artery disease, and autism spectrum disorder (9). Several works that assessed the impact of hydration on vocal cord vibration mechanics demonstrated an inverse relationship between hydration and phonation threshold pressure, ascribed to reduced viscosity of the vocal cord tissue (10-12). In the context of pulmonary congestion, it has been projected that the fluid overload may be detected by distinctly altered phonation patterns. In their voice analysis of 10 patients with acute decompensated HF (ADHF) who underwent standard treatment with diuretics, Murton et al (13) identified measurable changes in voice features that correlated with HF status. Maor et al (14) showed that a vocal biomarker, computed from recorded patient speech by a proprietary system, correlated with the probabilities of hospitalization and of death within a given time from the recording.

To test the hypothesis that distinct speech measures (SMs) reflect 2 different clinical states of pulmonary congestion, we performed a single-arm, open-label observational analysis of speech recordings of hospitalized patients with ADHF. Specifically, we sought to identify objective SMs distinguishing between ADHF on admission ("wet") versus at discharge ("dry") by using a novel speaker verification, speech processing, and analysis technology, implemented within a proprietary Smartphone application (app) (HearO Cordio Medical Ltd, Or Yehuda, Israel).

METHODS

PATIENT POPULATION AND STUDY DESIGN. This was a single-arm, observational open-label study. The study was approved by the institutional review board of Rambam Medical Center. Signed informed consent was obtained before any protocol-related procedures were initiated.

Patients eligible to participate had to meet the following inclusion criteria: age 18 years or older, current hospitalization caused by ADHF, as determined by the admitting physician based on medical history, clinical findings and chest x-ray suggestive of pulmonary congestion, which was assessed by a senior physician blinded to the recording analysis. Patients were excluded from the study if they had any of the following: acute coronary syndrome; significantly impaired glomerular filtration rate (<25 mL/min); dialysis treatment; congenital heart disease; mechanical heart valve; symptoms or signs of respiratory infection; sepsis; hemodynamic instability that required ionotropic therapy; or mechanical circulatory support.

TABLE 1 Characteristics of the 5 SMs Used in This Analysis

	SM1	SM2	SM3	SM4	SM5
High temporal resolution	✓	✓			✓
High spectral resolution			✓	✓	
Autoregressive model for the spectrum	✓	✓	✓	✓	
Nonlinear amplitude and frequency mapping[a]					✓
Symmetric version of a nonlinear spectra ratio	✓	✓	✓	✓	
Euclidean distance					✓

[a]Emulates cochlear auditory signal transduction.
SM = speech measure.

The medical decisions regarding admissions, time of discharge, and treatment options during hospitalization were all made based on clinical guidelines [European Society of Cardiology and American College of Cardiology/American Heart Association guidelines (15)] by a senior physician who was blinded to the recording analysis results.

Speech recordings were performed by each patient at the time of hospital admission and discharge.

VOICE RECORDING. The speech analysis system was a HearO speech capturing app (version 0.85), which runs on a smartphone; the analysis was executed on a cloud-based server. The HearOApp was used to record and digitize the patients' speech and then to upload the speech files to the server, where they were stored and analyzed. The app was installed on an iPhone 6 Plus (Apple, Cupertino, California, USA) provided by the study sponsor. For the purpose of the study, a software audio interface calibration (setting) procedure was performed using the recordings acquired from the first 10 successive patients. These patients were not included in the final study analysis.

In each recording session, the patient was prompted to repeat 5 sentences 3-4 times. The sentences were read (spoken) in the patient's native language (Hebrew, Arabic, or Russian). The duration of each recorded sentence was 2-5 s. Recording sessions were performed on the day of admission and on the day of discharge. The sentences for each language were composed by a phonetician, with the aim of representing, as much as possible, the phonetic variety in the language.

SPEECH ANALYSIS. The speech analysis scheme (**Central Illustration**) is based on the paradigm of text-dependent speaker verification (16,17). The premise behind the HearO System is that subtle physiological changes associated with HF decompensation affect the patient's speech and render them a "different person" (voiceprint). These changes are much more subtle than those found between different speakers,

but are nonetheless detectable using algorithms derived from those used in text-dependent speaker verification. Specifically in this study, the reference set for each sentence consisted of the recordings of a single sentence, by the same patient, at admission. The test recordings were the recordings of the same sentence, by the same patient, at discharge. This section describes the computation of the similarity score between a test utterance and a reference utterance.

The generation of speech sounds can be modeled as the passage of an excitation signal, originating from the lungs, through a linear filter. The filter represents the shape of the vocal tract, which includes the air pathways from the glottis upwards to the lips and nostrils. The excitation signal can be a periodic (voiced) excitation $S_p(\omega)$, generated by the vocal cords, or an aperiodic (unvoiced) excitation $S_a(\omega)$, generated by partially or fully obstructed air flow through the vocal tract, or a combination of $S_p(\omega)$ and $S_a(\omega)$. The model is particularly useful for short-term, frequency-domain analysis, where we may consider the excitation and filter to be time invariant (ie, the states of the vocal cords and of the vocal tract do not change). If $S(\omega)$ is the frequency-domain representation of the excitation and $H(\omega)$ is the frequency response of the filter, then the frequency-domain representation of the generated speech is the product of the 2: $X(\omega) = H(\omega)\ S(\omega)$. For example, the periodic excitation $S_p(\omega)$ appears in the frequency domain as a sequence of evenly spaced, equal-amplitude pulses, and $X(\omega)$ is a sequence of spikes with the same spacing but different amplitudes, determined by $H(\omega)$. Thus, the shape of the vocal tract determines the instantaneous spectral envelope.

Because the previous analysis applies only in the short term, in which the state of the articulatory organs is approximately constant, the spectral envelope $H(\omega)$ is estimated in short-term intervals (several tens of milliseconds) called "frames." The stream of speech is converted into a sequence of evenly spaced frames, after which a "feature vector" that represents the estimated spectral envelope during that frame is computed for each frame (18-21). Consequently, each utterance is represented by a sequence of frequency and perceptual feature vectors (22,23): $\boldsymbol{V} = \{\boldsymbol{v}[n],\ n = 1, \ldots, N\}$, where $\boldsymbol{v}[n]$ is the feature vector of the frame ending at time $n\delta$, where δ is the frame duration. Thus, $\boldsymbol{V}$ may be viewed as a sequence of "snapshots" of the vocal tract shape.

Let R and T be 2 utterances of the same sentence, with feature vector sequences $\boldsymbol{V}^R = \{\boldsymbol{v}^R[m],\ m = 1, \ldots, M\}$, $\boldsymbol{V}^T = \{\boldsymbol{v}^T[n],\ n = 1, \ldots, N\}$, respectively. We defined a similarity measure between the 2 utterances, or equivalently, between the 2 feature vector sequences, which indicated how R and T were dissimilar utterances (larger values mean greater dissimilarity). The distortion measure is defined by:

$$D(R,T) = D(\boldsymbol{V}^T, \boldsymbol{V}^R) = \min_{\psi} D(\boldsymbol{V}^T, \boldsymbol{V}^R; \psi) = \sum_{n=1}^{N} d(\boldsymbol{v}^T[n], \boldsymbol{v}^R[\psi(n)]),$$

where $d(\boldsymbol{v}_1, \boldsymbol{v}_2)$ is a local dissimilarity between 2 feature vectors: $\boldsymbol{v}_1, \boldsymbol{v}_2$ and $\psi : \{1, \ldots, L\} \rightarrow \{1, \ldots, N\} \times \{1, \ldots, M\}$ maps each feature vector in $\boldsymbol{V}^T$ to a corresponding feature vectors in $\boldsymbol{V}^R$, which represents the same state of the articulatory organs state, at the same point in the sentence and $D(\boldsymbol{V}^T, \boldsymbol{V}^R; \psi)$ is the sum of the local dissimilarities according to the mapping ψ. We considered the set Ψ of all mappings ψ, which satisfied the following constraints: $\psi(1) = (1,1)$, $\psi(L) = (N, M)$, which were monotonically rising and local slope bounded between 0.5 up to 2. The dissimilarity measure $D(R,T)$is the minimum of $D(\boldsymbol{V}^T, \boldsymbol{V}^R; \psi)$ over all the mapping ψ, which satisfied the constraints (18).

The 5 SMs used in this work (24) (**Table 1**) were selected by the Cordio Medical team as the ones considered to be most promising.

The similarity measures are unit-less, and, with appropriate normalization, are of similar scale, and can be simply defined as follows:

$$SM[\%] = \left(\frac{average\ of\ all\ dry - wet\ distortions\ of\ the\ same\ sentence}{average\ of\ all\ wet - wet\ distortions\ of\ different\ utterances\ of\ the\ same\ sentence} - 1 \right) \times 100$$

Ultimately, the algorithm compared similarity measures between utterances of the same physiological state to similarity measures between utterances of different physiological states; therefore, the scale is irrelevant and can be set arbitrarily.

STATISTICAL ANALYSES. The recordings were processed using a range of algorithms to report 5 different SMs, denoted SM1 SM2, SM3, SM4, and SM5. Descriptive statistics (mean, SD, median, minimum and maximum) of the relative (%) SMs change from admission day to discharge day were calculated, and 95% CIs were also calculated.

The percent change of each SM was found to exhibit an asymmetrical distribution and deviated from a normal distribution as assessed by the Shapiro-Wilk test. Therefore, the nonparametric sign test was used to test the null hypothesis that the median percent change of an SM from admission to the day of

discharge was 0. The number (%) of true positive alarms (hits) for each SM and overall were calculated, and *P* values of the binomial exact test of percent of hits = 0, and binomial exact 95% CIs were calculated.

RESULTS

A total of 47 patients with ADHF were enrolled; approximately 40 recordings were obtained from each patient (approximately 20 on admission and approximately 20 on discharge), which resulted in a total of 1,728 recordings. Seven patients were excluded from the analysis because of the technical insufficiency of the speech recordings, which left 1,484 recordings from 40 patients for analysis. An average (SD) 19.00 ± 2.90 admission recordings and 18.00 ± 4.09 discharge recordings were analyzed per patient. Baseline characteristics of the remaining 40 patients (Table 2) included an equal number of women and men; mean age was 75 ± 12 years. Most patients had a left ventricular ejection fraction of >40% (78%) with comorbidities of diabetes mellitus type 2 (63%) and hypertension (93%). Mean N-terminal pro-B-type natriuretic peptide level at admission was 7,002 ± 6,416 pg/mL, consistent with substantial fluid overload.

Interpatient comparisons of collected recordings identified significant differences in all 5 tested SMs of wet (admission) vs dry (discharge) recordings ($P < 0.0001$; sign test and paired Student's t-test), with SM1 showing a mean (median) 91% (68%) change from baseline, followed by SM2 with a change of 165% (108%), SM3 of 218% (158%), SM4 of 200% (143%), and SM5 of 49% (34%) (Table 3, Figure 1).

Overall, 94% of the discharge recordings were distinctly different from their respective baseline recordings, with all 5 SMs showing significant change, with 35 of 40 (87.5%) patients identifying these differences (Table 4). Partial detection (4 of 5 SMs) was achieved for 3 patients (7.5%). In 2 (5%) patients, none of the tested SMs showed any significant change from baseline. The breakdown over SMs was as follows: both SM2 and SM4 were successful in 38 (95%) patients, whereas each of SM1, SM3, and SM5 was successful in 37 (92.5%) patients. When we excluded the 2 patients who were not detected by any SM, SM2 and SM4 succeeded in all 38 remaining patients, whereas each of SM1 and SM3 was incorrect for 1 of the 38 patients. This was consistent with the fact that each SM focused on different information in the signal and thus could perform somewhat differently in different patients. The application was equally effective in detecting SM shifts in recordings of all 3 tested languages, despite the distinctly different phonetics of Russian as compared to Hebrew and Arabic. This was consistent with the general experience in text-dependent speaker verification, which was implemented in dozens of languages with no significant differences in accuracy among them (17).

TABLE 2 Baseline Demographics and Clinical Characteristics (n = 40)

Female	20 (50) (33.8-66.2)[a]
Age (yrs)	75 ± 12 (71.1-78.6)
Language	
Hebrew	13 (33) (18.6-49.1)
Russian	17 (43) (27.0-59.1)
Arabic	10 (25) (12.7-41.2)
Echo LVEF >40%	31 (78) (61.5-89.2)
History of coronary artery disease	22 (55) (38.5-70.1)
History of MI	10 (25) (12.7-41.2)
S/P CABG/ PCI	19 (48) (31.5-63.9)
History of diabetes mellitus	25 (63) (45.8-77.3)
History of hypertension	37 (93) (79.6-98.4)
NT-proBNP (pg/mL)	7,002 ± 6,416 (4,950-9,054)
SBP (mm Hg)	142 ± 26 (134-150)
Sodium (mEq/L)	137 ± 4 (136-140)
BUN (mg/dL)	35.7 ± 21.6 (28.7-42.7)
Serum creatinine (mg/dL)	1.6 ± 0.9 (1.4-1.9)

Values are n(%), mean ± SD, and (95% CI). [a]Binomial exact 95% CI.

BUN = blood urea nitrogen; CABG = coronary artery bypass grafting; LVEF = left ventricular ejection fraction; MI = myocardial infarction; NT-proBNP = N-terminal pro–B-type natriuretic peptide; PCI = percutaneous coronary intervention; SBP = systolic blood pressure.

As a complementary test, 72 untagged admission and discharge recordings from 9 patients (8 recordings per patient) were reanalyzed by the clustering algorithm, in a blinded manner. The system successfully segregated the recordings into 2 distinct unknown sets, which, when unblinded, were shown to correspond to the 2 different clinical statuses (ie, admission/discharge), with the exception of only 1 recording (2.2%). Taken together, exact matching between the 2 blindly computed clusters and the admission/discharge tagging was achieved in 44 (97.8%) of 45 cases (5 distortion measure variants for 9 patients) (data not shown).

DISCUSSION

This first study using a novel speech recognition system in patients demonstrated its ability to identify voice alterations reflective of changes in ADHF clinical status and showed large changes in SMs between the congested (admission) and decongested (discharge) states. Specifically, distinct speech vocalization features were identified in hospital admission versus discharge recordings of 40 patients with ADHF, with 95% of patients showing significant shifts in at least 4 of the 5 tested SMs.

TABLE 3 Descriptive Statistics of Change in Speech Model Variables

Speech Model Variable (% Change) (n = 40)	Mean ± SD	Min	Median	Max	95% CI	*P* Value (paired Student's t-test)[a]	*P* Value (sign test)[b]
SM1	91.2 ± 87.8	−30.0	67.9	387.8	63.1-119.3	<0.0001	<0.0001
SM2	164.9 ± 166.0	−59.4	108.4	665.9	111.8-218.0	<0.0001	<0.0001
SM3	217.7 ± 213.6	−47.9	157.8	983.4	149.4-286.0	<0.0001	<0.0001
SM4	200.1 ± 182.9	−55.4	143.2	719.9	141.6-258.6	<0.0001	<0.0001
SM5	48.5 ± 48.3	−26.3	33.5	197.4	33.1-64.0	<0.0001	<0.0001

[a]*P* value of paired Student's t-test for mean = 0. [b]*P* value of nonparametric sign test for median = 0.
Abbreviation as in Table 1.

The initial impetus for development of a speech analysis system for early warning of imminent HF decompensation came from informal observations by some of the investigators, that they could assess a patient's HF status by listening to the patient's voice. This anecdotal evidence was confirmed by a rigorous subjective test, in which 20 naïve listeners were presented with utterances recorded by 3 patients with HF in wet and dry conditions. For each patient, 1 dry utterance was provided to the listeners as a reference, after which, they were asked to classify 7 more utterances of the same patient as dry or wet, based on their similarity to the reference. The experimental design followed the MUSHRA (MUltiple Stimuli with Hidden Reference and Anchor) standard for subjective voice assessment (25). In 67% of the cases, the recording was correctly classified (95% CI: ±0.05). Because the success rate of random guessing would be 0.50, it was evident that patient speech contained information about HF status. However, the listeners could not point to any specific audible feature that helped them distinguish between the 2 classes. The current observations provided substantial proof of concept that this novel automated speech processing and analysis approach can reliably identify these differences between 2 states of pulmonary congestion in patients with HF at the time of hospitalization for ADHF and following a full course of inpatient treatment. In this context, this speaker verification–based concept has the potential to serve as a new tool in the in-hospital and the remote armamentarium for assessment of pulmonary congestion in patients with HF. If further validated in studies of ambulatory outpatients with chronic HF, this speech-based analysis could provide a simple, noninvasive approach for the remote monitoring and management of such patients.

TABLE 4 Number (%) of Hits[a] for Each SM and Overall

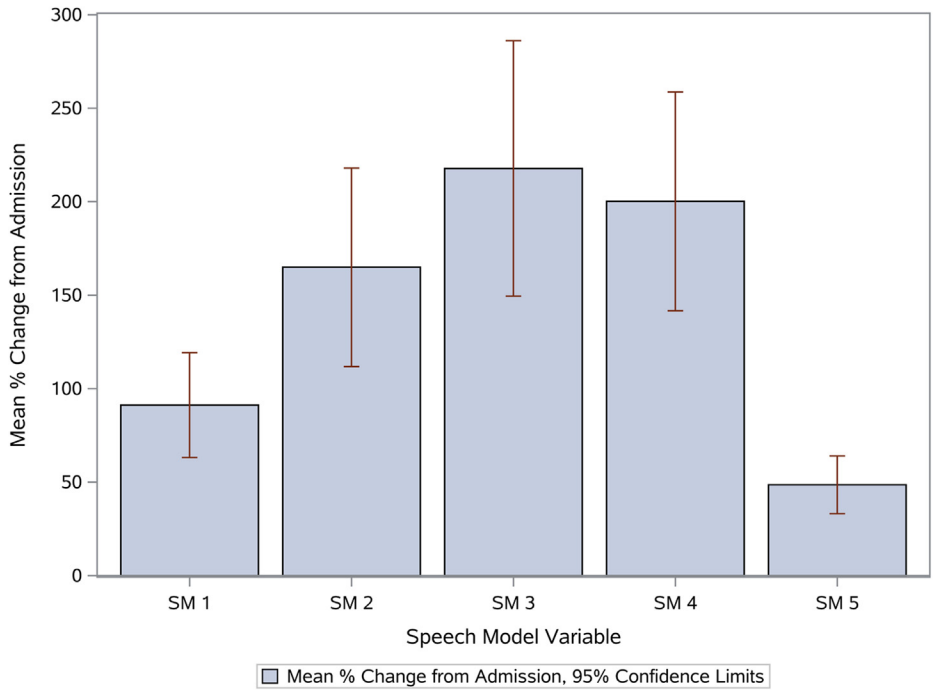
	Hit	*P* Value[b]	Binomial Exact 95% CI
SM1	37/40 (92.5)	<0.0001	79.6-98.4
SM2	38/40 (95.0)	<0.0001	83.1-99.4
SM3	37/40 (92.5)	<0.0001	79.6-98.4
SM4	38/40 (95.0)	<0.0001	83.1-99.4
SM5	37/40 (92.5)	<0.0001	79.6-98.4
Overall	35/40 (87.5)	<0.0001	(73.2-95.8)

[a]Hit = true positive alarm. [b]*P* value of the binomial exact test of percent of hits = 0.
Abbreviation as in Table 1.

Although investigation of the pathophysiological mechanism underlying the voice changes associated with ADHF status extended beyond the scope of this study, several plausible explanations could be provided. First, ADHF-related pulmonary congestion is often correlated with general edema, which, in turn, can alter the functioning of the articulatory organs, from the vocal cord upward (10). The changes in general hydration that cause voice changes have primarily been attributed to the level of fluid in the vocal cord, but it is possible that systemic edema also induces swelling of the epiglottis, soft palate, uvula, tongue, lips, inner cheeks, and soft tissue of the nasal cavity. These changes can modify the frequency responses of the vocal tract and possibly the aperiodic excitation, as opposed to the changes in the periodic excitation triggered by vocal cord swelling (26). Another possible mechanism is based on the effect of fluid accumulation in the alveoli or interstitial space on the acoustic properties of the lungs, as manifested by bronchophony, whispered pectoriloquy, and egophony (27), as well as on the reflective properties of the alveoli and air pathways. Such changes may result in altered acoustic wave reverberation inside the tracheobronchial tree (28), which may be reflected in speech. Thus, it is possible that the results of the present study represent changes not only in pulmonary congestion but also in edema that affects the upper airway, inclusive of the anatomy just mentioned.

The present findings also aligned with earlier studies, which pursued extractable speech features

FIGURE 1 **Change in Speech Model Variables from Admission Day to Discharge Day**



Patients hospitalized for acute decompensated heart failure, recorded several sentences into their Smartphone using the voice analysis application on the day of admission and at discharge. Mean percent change from baseline is presented, along with 95% confidence limits. Note that a positive change represents less pulmonary congestion. SM = speech measure.

discriminative of hydration-related pathophysiological states. For instance, pre- versus post-hemodialysis speech recordings showed significant differences in vocal acoustic parameters, including shimmer, fundamental frequency, maximum phonation time, and the noise-to-harmonics ratio (29-31), suggested to be the result of modified vocal cord thickness (31).

STUDY LIMITATIONS. Several limitations of the present study should be noted. The study tested a small cohort of patients, was a single-center and open-label study, and postdischarge follow-up was not performed via this specific protocol. In addition, using the admission and discharge as proxies for 2 distinct ADHF states was not ideal because some patients might have been released while still experiencing some level of pulmonary edema, and the extent of such practice might vary across hospitals. Furthermore, the cumulative recording time, per patient, was approximately 1.75 min at admission and 1.75 min at discharge, capturing only a short segment of their condition. Moreover, although use of the SMs provided a compelling proof of concept, SMs are not clinical-grade detectors. The latter should provide, in addition to a binary wet/dry decision, additional information, such as confidence level or degree of "wetness." In addition, such a tool should combine the information from all 5 SMs to reach a more accurate decision, and it should probably use a method more sophisticated than simple averaging to combine all the available distortion measures (as is often done in the speaker verification field). The development of a clinical-grade verifier requires a significantly larger volume of data than available from our experiments.

CONCLUSIONS

Although further research is needed to determine its potential contribution in both hospital and ambulatory settings, the present study highlighted automated speech voice analysis as a potential clinical tool and served as a first step toward implementation of an objective, highly accessible, nonintrusive, and low-cost pulmonary congestion monitoring tool to identify differences between wet

and dry states. Because the present analysis compared recordings at relatively extreme conditions (ie, admission vs discharge), confirmation of performance in more subtly different clinical states is still required. Similarly, applicability of SM profiles specific to defined HF subgroups and other clinical disorders, remains to be determined.

ACKNOWLEDGMENTS The authors acknowledge the substantial contributions of key team members:

Mrs Maria Goldshmidt, Dr Geula Klorin, and Dr Katya Dolnikov, MD, PhD, Internal Medicine "B", Rambam Health Care Campus, Haifa, Israel.

FUNDING SUPPORT AND AUTHOR DISCLOSURE

The study was supported by Cordio Medical Ltd. Dr Amir has been a paid consultant to Cordio Medical Ltd. Dr Abraham has received consulting fees from Abbott, Boehringer Ingelheim, CVRx, Edwards Lifesciences, Respicardia; has received salary support from V-Wave Medical; and has received research support from the U.S. National Institutes of Health/National Heart, Lung, and Blood Institute. Dr Anker has received grant support from Abbott and Vifor Pharma; and has received fees from Abbott, Bayer, Boehringer Ingelheim, Cardiac Dimension, Impulse Dynamics, Novartis, Servier, and Vifor Pharma. Dr Pinney has received consulting fees from Abbott, CareDx, Medtronic, NuPulse, and Procyrion. Dr Shallom is the Chief Technology Officer of Cordio Medical. Dr Lotan has been a board member of Cordio Medical Ltd; and has received lectures fee from Boehringer Ingelheim. Dr Edelman has been supported in part by a grant from the National Institutes of Health (NIH R01 49039); and has been a paid consultant to Cordio Medical Ltd. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

ADDRESS FOR CORRESPONDENCE: Dr William T. Abraham, Division of Cardiovascular Medicine, The Ohio State University, 473 West 12th Avenue, Columbus, Ohio, 43210, USA. E-mail: William.Abraham@osumc.edu.

PERSPECTIVES

COMPETENCY IN MEDICAL KNOWLEDGE: Seasoned clinicians may recognize worsening HF from patients' voices and speech. The technology automates quantification of this distinction using speaker verification–based algorithms for in-person and remote (telemedicine) assessment of HF clinical status.

TRANSLATIONAL OUTLOOK: The present study highlights automated speech voice analysis as a potential clinical tool for the assessment of clinical congestion in patients with HF. This tool serves as a first step toward implementation of an objective, highly accessible, nonintrusive, and low-cost pulmonary congestion monitoring tool, to identify differences between wet and dry states. With further clinical validation, this approach to HF monitoring may represent a useful addition to our armamentarium for HF management.

REFERENCES

1. Benjamin EJ, Muntner P, Alonso A, et al. Heart disease and stroke statistics-2019 update: a report from the American Heart Association. *Circulation*. 2019;139:e56–e528.

2. Savarese G, Lund LH. Global public health burden of heart failure. *Card Fail Rev*. 2017;3:7–11.

3. Fonarow GC. The Acute Decompensated Heart Failure National Registry (ADHERE): opportunities to improve care of patients hospitalized with acute decompensated heart failure. *Rev Cardiovasc Med*. 2003;4(Suppl 7): S21–S30.

4. Gheorghiade M, Abraham WT, Albert NM, et al. Systolic blood pressure at admission, clinical characteristics, and outcomes in patients hospitalized with acute heart failure. *JAMA*. 2006;296: 2217–2226.

5. Adamson PB, Abraham WT, Aaron M, et al. CHAMPION trial rationale and design: the long-term safety and clinical efficacy of a wireless pulmonary artery pressure monitoring system. *J Card Fail*. 2011;17:3–10.

6. Yang XW, Hua W, Ding LG, et al. OptiVol fluid index predicts acute decompensation of heart failure with a high rate of unexplained events. *J Geriatr Cardiol*. 2013;10:253–257.

7. Sasaki S, Ishida Y, Kinjo T, et al. Telediagnosis of heart failure with continuous intrathoracic impedance monitoring by Medtronic CareLink Network: importance of the elevation pattern of OptiVol Fluid Index. *J Arrhythmia*. 2013;29:347–352.

8. Abraham WT, Adamson PB, Bourge RC, et al. Wireless pulmonary artery haemodynamic monitoring in chronic heart failure: a randomised controlled trial. *Lancet*. 2011;377:658–666.

9. Bonneh YS, Levanon Y, Dean-Pardo O, Lossos L, Adini Y. Abnormal speech spectrum and increased pitch variability in young autistic children. *Front Hum Neurosci*. 2011;4:237.

10. Verdolini K, Min Y, Titze IR, et al. Biological mechanisms underlying voice changes due to dehydration. *J Speech Lang Hear Res*. 2002;45: 268–281.

11. Verdolini K, Titze IR, Fennell A. Dependence of phonatory effort on hydration level. *J Speech Hear Res*. 1994;37:1001–1007.

12. Solomon NP, Glaze LE, Arnold RR, van Mersbergen M. Effects of a vocally fatiguing task and systemic hydration on men's voices. *J Voice*. 2003;17:31–46.

13. Murton OM, Hillman RE, Mehta DD, et al. Acoustic speech analysis of patients with decompensated heart failure: a pilot study. *J Acoust Soc Am*. 2017;142:EL401.

14. Maor E, Perry D, Mevorach D, et al. Vocal biomarker is associated with hospitalization and mortality among heart failure patients. *J Am Heart Assoc*. 2020;9:e013359.

15. Yancy CW, Jessup M, Bozkurt B, et al. 2017 ACC/AHA/HFSA focused update of the 2013 ACCF/AHA guideline for the management of heart failure. *J Am Coll Cardiol*. 2017;70:776–803.

16. Bimbot F, Bonastre J-F, Fredouille C, et al. A tutorial on text-independent speaker verification. *EURASIP J Appl Signal Proc*. 2004;2004:101962.

17. Furui S, Rosenberg AE. Speaker verification. In: Madisetti VK, Williams DB, eds. *Digital Signal Processing Handbook*. CRC Press; 1999:1–22.

18. Deller JR, Proakis JG, Hansen JHL. *Discrete Time Processing of Speech Signals*. Macmillan; 1993.

19. Furui S. *Digital Speech Processing, Synthesis and Recognition*. Marcel Dekker; 1989.

20. Rabiner LR, Juang BH. *Fundamentals of Speech Recognition*. Prentice Hall; 1993.

21. Tirumala SS, Shahamiri SR, Garhwal AS, Wang R. Speaker identification features extraction methods: a systematic review. *Exp Syst Appl*. 2017;90:250–271.

22. Oppenheim AV, Schafer RW. *Discrete-Time Signal Processing*. Pearson; 2010.

23. Rabiner LR, Schafer RW. *Digital Processing of Speech Signals*. Prentice Hall; 1978.

24. Acero A, Hon H, Huang X. *Spoken Language Processing: A Guide to Theory, Algorithm and System Development*. Prentice Hall; 2001.

25. International Telecommunication Union. *ITU-R BS.1534-1: method for the subjective assessment of intermediate quality level of audio systems*. International Telecommunication Union; 2015:1–36.

26. Zeitels SM, Hillman RE, Bunting GW, Vaughn T. Reinke's edema: phonatory mechanisms and management strategies. *Ann Otol Rhinol Laryngol*. 1997;106:533–543.

27. Sarkar M, Madabhavi I, Niranjan N, Dogra M. Auscultation of the respiratory system. *Ann Thorac Med*. 2015;10:158–168.

28. Henry B, Royston TJ. A multiscale analytical model of bronchial airway acoustics. *J Acoust Soc Am*. 2017;142:1774.

29. Sagiroglu S, Doganer A. The effect of electrolyte balance on the voice in hemodialysis patients. *Eur Arch Otorhinolaryngol*. 2018;275:2755–2761.

30. Unver S, Hardal U, Esertas K, Sezen A, Celikbilek F, Altundag A. Objective analysis of voice changes in a hemodialysis session and its correlation with ultrafiltration. *Ren Fail*. 2015;37:268–272.

31. Jung SY, Ryu JH, Park HS, Chung SM, Ryu DR, Kim HS. Voice change in end-stage renal disease patients after hemodialysis: correlation of subjective hoarseness and objective acoustic parameters. *J Voice*. 2014;28:226–230.

KEY WORDS **acute decompensated heart failure (ADHF), remote speech analysis, speech measure (SM)**