Evaluation of Speech and Pause Alterations in Patients With Acute and Chronic Heart Failure

Dario Schöbi, PhD;* Yan-Ping Zhang, PhD;* Joelle Kehl, BSc; Meriam Aissani, BSc; Otmar Pfister, MD; Martin Strahm, PhD; Paul van Haelst, MD, PhD; Qian Zhou, MD

BACKGROUND: Acute heart failure is the most frequent cause of unplanned hospital admission in elderly patients. Various biomarkers have been evaluated to better assess the status of these patients and prevent decompensation. Recently, voice has been suggested as a cost-effective and noninvasive way to monitor disease progression. This study evaluates speech and pause alterations in patients with acute decompensated and stable heart failure. Specifically, we aim to identify a vocal biomarker that could be used to monitor patients with heart failure and to prevent decompensation.

METHODS AND RESULTS: Speech and pause patterns were evaluated in 68 patients with acute and 36 patients with stable heart failure. Voice recordings were performed using a web-browser based application that consisted of 5 tasks. Speech and pause patterns were automatically extracted and compared between acute and stable patients and with clinical markers. Compared with stable patients, pause ratio was up to 14.9% increased in patients with acute heart failure. This increase was largely independent of sex, age, and ejection fraction and persisted in patients with lower degrees of edema or dyspnea. Furthermore, pause ratio was positively correlated with NT-proBNP (N-terminal pro-B-type natriuretic peptide) after controlling for acute versus stable heart failure. Collectively, our findings indicate that the pause ratio could be useful in identifying acute heart failure, particularly in patients who do not display traditional indicators of decompensation.

CONCLUSIONS: Speech and pause patterns are altered in patients with acute heart failure. Particularly, we identified pause ratio as an easily interpretable vocal biomarker to support the monitoring of heart failure decompensation.

Key Words: acute heart failure ■ chronic heart failure ■ digital health ■ voice analysis

Heart failure (HF) is a chronic condition caused by structural or functional cardiac abnormalities, resulting in symptoms such as breathlessness, fatigue, and edema. HF affects around 26 million people worldwide, and the numbers are expected to increase exponentially.¹ Patients with HF can be stable, but progression of the underlying conditions or other changes in health can result in acute decompensated HF, which accounts for the majority of HF hospitalizations.²³ Each HF-related hospitalization increases the risk for subsequent hospitalization events⁴ and despite current treatments, rates of hospital admissions for HF are still high. Therefore, improvements in outpatient management are needed to address the increasing burden of HF.

Recently, voice assessment has been suggested as a cost-effective and noninvasive tool for coronary artery disease.⁵ Changes in voice patterns have been identified in patients with decompensated HF⁶ or pulmonary congestion.⁷ Additionally, changes in vocal characteristics were associated with adverse outcomes among patients with HF.⁸ These findings suggest a potential of voice as a novel biomarker to monitor patients with HF noninvasively. However, its application in clinical practice remains challenging, as simplified standardized
voice tests and easily interpretable vocal biomarkers are yet underdeveloped.

In our study, we hypothesized that worsening of the HF condition is reflected in an altered speech rate resulting from changed speech and pause time. We used simple text-reading and number counting and adopted the Stroop test\(^9\) to evaluate speech and pause time patterns in patients with acute decompensated HF and stable chronic HF. Specifically, we aimed to identify reproducible and informative voice biomarkers for the assessment of speech and pause alterations in HF.

**METHODS**

The data that support the findings of this study are available from the corresponding author upon reasonable request. Because of the sensitive nature of voice recordings, only processed data can be made available.

**Study Design and Population**

Patients admitted to the University Hospital Basel because of acute HF (AHF) were enrolled between March 2020 and May 2021. Patients with AHF had either preexisting HF or new-onset HF and were included within 48 hours after admission if they fulfilled the inclusion criteria. AHF was defined as the presence of at least 1 of the following symptoms: (1) progressive dyspnea on exertion, (2) orthopnea, (3) paroxysmal nocturnal dyspnea, (4) fatigue, (5) peripheral edema, or (6) weight gain. Additionally, at least 1 of the following parameters had to be fulfilled: (1) signs with high specificity of AHF—bilateral pulmonary rales and/or peripheral edema and/or jugular vein dilatation, positive hepatojugular reflux; (2) NT-proBNP (N-terminal pro-B-type natriuretic peptide) >450 pg/mL if age <50 years, >900 pg/mL if age 50 to 75 years, >1800 pg/mL if age >75 years (cutoffs should be reduced by 50% in case of severe obesity [body mass index >35 kg/m\(^2\)]\(^{10}\); (3) treatment with intravenous (furosemide) or increased doses of oral diuretics (furosemide, torsemide, metolazone) within 24 hours of presentation; or (4) radiological features with high specificity for AHF—interstitial or alveolar edema with or without bilateral pleural effusion.

Patients with stable HF were recruited from the HF outpatient clinic (OP) at the University Hospital Basel during their routine follow-up visits. These patients were defined as those who (1) took stable guideline-directed HF medication during the past 3 months, (2) had not been hospitalized for HF during the past 3 months, and (3) showed no clinical signs of AHF.

Every patient ≥18 years and meeting the inclusion criteria was invited to participate in the study irrespective of their left ventricular ejection fraction (LVEF). Specific exclusion criteria included acute respiratory infection, acute worsening of pulmonary diseases that required intensifying the prior treatment regime, hemodynamic instability requiring inotropes or vasopressors, sepsis, history of diagnosed vocal cord pathology or dysfunction, inability to follow procedures, or inability to give consent. Blood levels of NT-proBNP (N-t) were collected at admission and before discharge for hospitalized patients and during the OP visit for patients with stable HF.

Patients were excluded from baseline analyses for any of the following reasons: (1) missing levels of NT-proBNP at admission, (2) impaired ability to read German, or (3) impaired ability to discriminate between red and green colors used in the Stroop task. Additionally, patients were excluded from the analysis of change between admission and discharge for any of the following reasons: (1) missing levels of NT-proBNP at discharge and (2) fewer than 2 available voice recordings.
In a subset of patients, the severity of dyspnea was assessed using the visual analog scale (VAS)\textsuperscript{11} and the American Thoracic Society Shortness of Breath Scale.\textsuperscript{12} The study was approved by the Swiss Association of Research Ethics Committees (Ethikkommission Nordwest-und Zentralschweiz) and complied with the Declaration of Helsinki. Written informed consent was obtained from each patient before inclusion.

**Voice Recording and Speech Analyses**

The voice tasks were performed using a web-browser based application on a Samsung Galaxy Tablet (S6). The application was developed in house. Hospitalized patients with AHF performed the voice tasks daily or every other day. OP patients with stable HF performed the tasks only once at the time of their (single) visit.

The speech protocol consisted of 5 tasks (Figure 1A): (1) reading a patient consent text, (2) number counting in ascending order, (3) number counting in descending order, (4) Stroop word reading (“red,” “green,” “blue”) displayed in black ink, and (5) Stroop color word reading with randomized word-ink combinations. In contrast to the traditional use of Stroop in cognitive neuroscience,\textsuperscript{13} both Stroop tasks were used as pure reading tests. In 4, patients had to read the color words as written. In 5, the color words were displayed in a different ink color for example, the word “red” was displayed in blue ink. Here, patients were asked to ignore the ink color and read out the color word. The word order in 4 and 5 was randomized daily to prevent effects of familiarity with the text. All tests were performed in German. Patients were instructed to read out the displayed words as clearly and quickly as possible.

The voice recordings were automatically segmented into articulation and pause segments using a custom written algorithm. This algorithm is based on the relative change of energy in the signal during articulation and pausing. A pause segment was defined as the pause between different utterances, that is, pauses at the end of words and pauses within polysyllabic words.\textsuperscript{14} An exemplary segmentation of the audio signal is illustrated in Figure 1B.

We extracted 4 voice features related to speech and pause time from the segmented recordings. Pause ratio (PR) was defined as the ratio between the total pause duration and total task time (sum of articulation and pause segments). PR can be broken down into (mean) articulation duration (MAD) and (mean) pause duration (MPD). MAD was calculated as the total duration of articulation divided by the number of articulation segments. MPD was calculated as the total duration of pauses divided by the number of pause segments (ie, the number of articulation segments minus 1). PR was further compared with rate of speech timing (RST), that is, the number of syllables per second (speech rate).
RST was computed as the slope of the linear fit of the articulation segment onset to the segment number.\textsuperscript{15}

**Statistical Analysis**

We assessed the clinical and vocal baseline characteristics of patients with AHF and those from the OP, as defined. Values are reported as mean±SD, count (%) or median [interquartile range], unless otherwise indicated. Significance for continuous clinical variables was assessed with the nonparametric Mann–Whitney U test. Effect sizes were estimated in terms of the common language effect size, which equals the proportion of pairs where a value in group A is larger than a value in group B.\textsuperscript{16} For variables referring to frequencies (ratios), we used χ\textsuperscript{2} tests.

To account for differences in covarying variables, we performed propensity score (PS) matching.\textsuperscript{17} Age, sex, and LVEF entered the PS for AHF. We then matched OP patients with stable HF and patients with AHF based on their PS and recomputed the vocal characteristics for the subset of matched patients. For the matched subset, the covariates are, on average, balanced between AHF and stable OP, and potential differences in the vocal features not because of differences in the covariates. This covariate balance was evaluated visually using the histogram of the PS and based on Cohen’s d\textsuperscript{0} before and after matching. Cohen’s d\textsuperscript{0} relates differences between cohorts to the pooled SD (eg, d\textsuperscript{0}=0.5 implies that the difference in cohort means is half as large as the pooled SD). Commonly used thresholds to interpret Cohen’s d are d\textsuperscript{0}=0.2 (small), d\textsuperscript{0}=0.5 (medium), and large d\textsuperscript{0}=0.8 (large) effects.\textsuperscript{18}

To account for further confounding effects, we investigated speech changes by matching and restriction. This analysis looks at the vocal characteristics within subgroups of the clinical categories that showed a significant difference between patients with AHF and stable OP patients. Within these restricted subgroups, differences in vocal characteristics between AHF and stable OP are not because of differences in the severity or presence of the clinical category.

To investigate potential associations of PR with symptom expression and degree of congestion, we visualized PR across the levels of New York Heart Association (NYHA) functional classification and edema, the presence of lung auscultation findings, NT-proBNP, and the experience of dyspnea as assessed by the VAS. To interpret the associations beyond the average PR difference between AHF and stable OP, we controlled for cohort in a linear regression of the form: \( PR = \beta_0 + \beta_1 \cdot COHORT + \beta_2 \cdot Clinical + \epsilon \). Here, the regression coefficient \( \beta_0 \) reflects the average PR and \( \beta_1 \) reflects the difference of PR between the 2 cohorts. The coefficient \( \beta_2 \) reflects the linear association between PR and the clinical feature (eg, NT-proBNP).

In order to increase the statistical efficiency, levels of NYHA and edema were transformed to scalar values (NYHA: I, II, III, IV→1, 2, 3, 4; and edema: no, mild, moderate, severe→0, 1, 2, 3). All linear regressions were visually assessed.

Finally, the changes in NT-proBNP, body mass index, and PR of patients with AHF, from admission to discharge, were investigated using the Wilcoxon signed rank test.

The statistical analyses were performed with the Python package `pingouin` ver. 0.5.1. PS matching was performed using `psmpy` ver. 0.2.8. Multivariate regression analyses were performed with the Python package `statsmodels` ver. 0.13.2.

**RESULTS**

**Study Population Characteristics**

In total, 78 hospitalized patients with AHF and 38 patients with stable chronic HF from the OP were included. Ten patients with AHF and 2 OP patients with stable HF were excluded from the analyses, because 1 or more exclusion criteria were met as described in the methods section. The baseline characteristics of the remaining study population is summarized in Table 1. Overall, patients with AHF were older than OP patients (76.1±9.2 versus 67.2±12.8 years, \( P<0.001 \)). Patients were predominantly male in both cohorts (72% and 89%, \( P=0.086 \)). LVEF was higher in patients with AHF (44.3±15.7 versus 36.7±9.4%, \( P=0.010 \)). Patients with AHF were categorized in higher NYHA classes III and IV (72% versus 6%, \( P<0.001 \)), showed higher levels of circulating NT-proBNP (4292 [2477, 8175] versus 578 [224, 3730] pg/mL, \( P<0.001 \)), and showed higher levels of dyspnea (≥3) as measured by the American Thoracic Society scale (54% versus 0%, \( P<0.001 \)). These higher levels of dyspnea were consistent with higher levels of self-reported dyspnea on the VAS. Patients with AHF and OP patients with stable HF reported a VAS of 4 [3, 6] versus 1 [0, 2] cm, respectively (\( P<0.001 \)). The presentation of moderate to severe peripheral edema differed significantly across the 2 groups (34% versus 11%, \( P=0.023 \)). Likewise, abnormal lung auscultation, defined as the presence of crackles or rales, was significantly more frequent in patients with AHF (63% versus 3%, \( P<0.001 \)). Only 2 patients (1 from each cohort) presented with crackles.

**Reduced Speech Rate and Elevated Pause Ratio in AHF**

All patients were instructed to perform 5 speech tests as clearly and quickly as possible. From each test, we extracted 4 voice features measuring speech and pause time (RST, PR, MAD, and MPD as described...
in the methods section, Table 2). Compared with OP patients with stable HF, patients with AHF showed a decrease in RST in the text reading and Stroop tests at admission (Figure 2A). PR was markedly increased in these 3 tests and showed the highest effect sizes overall, with common language effect size >0.9 in both Stroop tasks (Table 2).

The increased PRs in text reading and both Stroop tests were mostly mediated by higher MPD in patients with AHF (Figure 2C). Interestingly, our data

| Table 1. Basic Characteristics at Admission of Patients With AHF or Outpatient Visit for OP Patients With Stable HF |
|---------------------------------------------------------------|
| **AHF** | **Stable OP** | **P value** |
|---------|---------------|-------------|
| N | Value | N | Value |  |
| Age, y | 68 | 76.1±9.2 | 36 | 67.2±12.8 | <0.001 |
| Sex, male | 68 | 49 (72) | 36 | 32 (89) | 0.09 |
| Body mass index, kg/m² | 68 | 27.1±5.1 | 36 | 28.4±4.7 | 0.14 |
| New York Heart Association | 68 |  | 36 |  |
| I | 2 (3) | 19 (28) | 13 (36) | 34 (84) | <0.001 |
| II | 17 (25) | 21 (58) |  |
| III | 37 (54) | 49 (72) | 2 (6) | 2 (6) |  |
| IV | 12 (18) | 0 (0) |  |
| LV, % | 68 | 44.3±15.7 | 36 | 36.7±9.4 | 0.010 |
| HF with reduced EF, LVEF <40% | 25 (37) | 21 (58) | 0.001 |
| HF with mildly reduced EF, LVEF 40–49% | 11 (16) | 11 (31) |  |
| HF with preserved EF, LVEF >50% | 32 (47) | 4 (11) |  |
| Hypertension | 68 | 40 (59) | 36 | 16 (44) | 0.23 |
| Tachyarrhythmia | 68 | 41 (60) | 36 | 1 (3) | <0.001 |
| Hypertrophy | 68 | 17 (25) | 36 | 0 (0) | 0.003 |
| Myocarditis | 68 | 0 (0) | 36 | 1 (3) | 0.75 |
| Ischemic | 68 | 20 (29) | 36 | 21 (58) | 0.008 |
| Dilated cardiomyopathy | 68 | 32 (47) | 36 | 12 (33) | 0.25 |
| Chronic obstructive pulmonary disease | 68 | 5 (7) | 36 | 4 (11) | 0.78 |
| Active smoking | 68 | 14 (21) | 36 | 7 (19) | 0.90 |
| Edema | 68 |  | 36 |  |
| No | 20 (29) | 45 (66) | 27 (75) | 32 (89) | 0.023 |
| Mild | 25 (37) | 5 (14) |  |
| Moderate | 13 (19) | 26 (34) | 4 (11) | 4 (11) |  |
| Severe | 10 (15) | 0 (0) |  |
| N-terminal pro-B-type natriuretic peptide, pg/mL | 68 | 4292 [2477, 8175] | 36 | 578 [224, 3730] | <0.001 |
| Estimated glomerular filtration rate, mL/min per 1.73 cm² | 64 | 50 [38, 69] | 36 | 70 [47, 80] | 0.007 |
| Abnormal lung auscultation | 67 | 42 [63] | 35 | 1 (3) | <0.001 |
| Dyspnea (American Thoracic Society) | 41 |  | 15 |  |
| 0 | 1 (2) | 19 (48) | 7 (47) | 15 (100) | <0.001 |
| 1 | 4 (10) | 6 (40) |  |
| 2 | 14 (34) | 2 (13) |  |
| 3 | 16 (39) | 22 (54) | 0 (3) | 0 (0) |  |
| 4 | 6 (15) | 0 (0) |  |
| Visual analog scale, cm | 41 | 4 [3, 6] | 15 | 1 [0, 2] | <0.001 |

N indicates the number of patients with available measures in the respective cohort. Values are reported as mean±SD, N (%), and median [interquartile range]. AHF indicates acute heart failure; HF, heart failure; LVEF, left ventricular ejection fraction; and OP, outpatient clinic.
also indicated a decrease in MAD in patients with AHF (Figure 2D) with the largest differences in MAD observed in the number counting normal and reverse task, as well as Stroop color word reading.

Because we observed a significant age difference between patients with stable and acute HF, we repeated the analysis of the voice characteristics after matching OP patients with stable HF and patients with AHF patients based on their PS. Age, sex, and LVEF entered the PS. The differences between the 2 cohorts for each of the 3 values were reduced after PS matching (Cohen’s $d$, before versus after): LVEF: 0.55 versus −0.06, age: 0.84 versus 0.43 and sex: 0.42 versus −0.20. Therefore, no moderate or large differences were observed in any of the 3 variables for the matched patients (N=72). Importantly, all significant differences observed for PR remained significant with $P<0.001$ for reading text, Stroop word reading, and Stroop color word reading.

To exclude further confounding effects we investigated PR by matching and restriction. The critical categories were based on the significant clinical features in Table 1. For this analysis, we averaged PR for each patient across the 2 Stroop tasks. The results are summarized in Table 3. Across all considered subgroups, PR remained elevated among patients with AHF. PR was significantly higher in patients with AHF versus OP patients with stable HF in NYHA classes I and II and for patients with no or mild edema. Importantly, PR remained elevated in patients with AHF who were in sinus rhythms, had normal lung auscultations, and for patients with low dyspnea scores (American Thoracic Society scale score <3). Furthermore, despite all considered patients having high levels of NT-proBNP, OP patients with stable HF showed lower PR than patients with AHF.

**Association of PR With Clinical Features**

To further characterize the role of PR in patients with HF, averaged PR for each patient across the 2 Stroop tasks was compared with the NYHA classes, severity of peripheral edema, lung auscultation, NT-proBNP, and self-reported levels of dyspnea on the VAS.

---

**Table 2. Characteristics of Pause Ratio Related Voice Features Across All 5 Tasks and Both Cohorts**

|                           | AHF (N=68) Mean±SD | Stable OP (N=36) Mean±SD | $p$ (Bonferroni) | Common language effect size |
|---------------------------|--------------------|--------------------------|-----------------|-----------------------------|
| Rate of speech timing     |                    |                          |                 |                             |
| Number counting normal    | 2.05±0.45          | 1.93±0.58                | 0.39            | 0.61                        |
| Number counting reverse   | 2.02±0.46          | 1.92±0.43                | 1               | 0.58                        |
| Reading text              | 2.25±0.31          | 2.67±0.31                | <0.001          | 0.17                        |
| Stroop color word reading | 1.78±0.38          | 2.06±0.32                | 0.008           | 0.31                        |
| Stroop word reading       | 1.80±0.32          | 2.11±0.31                | <0.001          | 0.26                        |
| Pause ratio               |                    |                          |                 |                             |
| Number counting normal    | 41.10±12.15        | 45.17±9.55               | 0.30            | 0.39                        |
| Number counting reverse   | 39.30±11.48        | 40.51±9.92               | 1               | 0.46                        |
| Reading text              | 43.70±7.51         | 32.04±8.74               | <0.001          | 0.84                        |
| Stroop color word reading | 45.07±9.85         | 30.19±5.62               | <0.001          | 0.91                        |
| Stroop word reading       | 43.04±8.73         | 29.19±5.36               | <0.001          | 0.92                        |
| Mean pause duration       |                    |                          |                 |                             |
| Number counting normal    | 0.21±0.12          | 0.27±0.12                | 0.030           | 0.34                        |
| Number counting reverse   | 0.19±0.10          | 0.23±0.11                | 0.15            | 0.37                        |
| Reading text              | 0.20±0.07          | 0.13±0.05                | <0.001          | 0.82                        |
| Stroop color word reading | 0.27±0.13          | 0.15±0.04                | <0.001          | 0.87                        |
| Stroop word reading       | 0.25±0.09          | 0.14±0.04                | <0.001          | 0.89                        |
| Mean articulation duration|                    |                          |                 |                             |
| Number counting normal    | 0.25±0.03          | 0.28±0.04                | <0.001          | 0.27                        |
| Number counting reverse   | 0.24±0.04          | 0.29±0.04                | <0.001          | 0.17                        |
| Reading text              | 0.24±0.02          | 0.25±0.02                | 0.028           | 0.33                        |
| Stroop color word reading | 0.30±0.04          | 0.34±0.04                | 0.003           | 0.30                        |
| Stroop word reading       | 0.31±0.04          | 0.33±0.04                | 0.048           | 0.34                        |

Statistical significance assessed using Mann–Whitney U tests with Bonferroni correction. AHF indicates acute heart failure; and OP, outpatient clinic.
Table 4 summarizes the bivariate association between PR and the clinical features and laboratory measures while considering the main effect of cohort. This analysis tests for an additional linear association of PR with severity, taking into account an average difference between the cohorts.

PR was consistently elevated in patients with AHF for all clinical subgroups and levels of NT-proBNP (Figure 3A through 3E). The first column in Table 4 shows the main effect of cohort, estimating the difference between patients with AHF and OP patients with stable HF to range between 12.5 and 15.4%. Furthermore, PR remained positively correlated with (log-) NT-proBNP after controlling for the cohort difference (Coeff [SE]=1.16 [0.56], \( P=0.041, N=104 \)). However, for NYHA, edema, lung auscultation, and the VAS, no significant linear associations of PR were observed (Table 4).

### PR Change From Admission to Discharge

The difference in PR between patients with stable HF and AHF was not limited to distinct diagnostic subgroups. This promotes PR as an independent marker of the disease trajectory. We next evaluated the temporal change of PR with the course of AHF treatment. Additionally, we compared the PR change to the reduction in NT-proBNP and loss of weight, which are 2 typical clinical indicators of overall improvement of the HF condition. A total of 51% of the patients showed lower PRs at discharge than at admission (mean change=−1.2±7.1%, \( N=55, P<0.05, \) Figure 4A). NT-proBNP in patients with AHF decreased in 87% of the patients (median decrease=1241 [481, 2974] pg/mL, \( N=55, P<0.001 \)) but remained elevated at discharge compared with NT-proBNP in OP patients with stable HF at their routine visit (2316 [985, 4246] versus 578 [224, 3730] pg/mL). This suggests a persistent

| Table 3. Comparing PR Within Clinical Subgroups of Similar Severity |
|----------------------|---------------------------------|----------------------|---------------------------------|---------------------------------|----------------------|---------------------------------|
|                      | AHF                             | Stable OP            |                      |                                |                      |                                |
|                      | Subgroup N Mean±SD              | N Mean±SD            |                      |                                |                      |                                |
|                      |                                 |                      |                      |                                |                      |                                |
| New York Heart Association: I-II |                                 |                      |                      |                                |                      |                                |
| I-II                 | 19 45.8±8.6                     | 34 29.5±5.2          | <0.001              | 0.97                           |                      |                                |
| N-terminal pro-B-type natriuretic peptide | >1800 ng/mL | 56 45.2±9.0 | 14 29.6±4.4 | <0.001 | 0.96 |                      |                                |
| Edema                | No-mild 45 44.9±9.0             | 32 30.0±5.3          | <0.001              | 0.93                           |                      |                                |
| Tachyarrhythmia      | No 27 43.6±8.2                  | 35 29.5±5.0          | <0.001              | 0.95                           |                      |                                |
| Abnormal lung auscultation | No 26 43.2±9.5 | 34 29.6±5.2 | <0.001 | 0.91 |                      |                                |
| Dyspnea (American Thoracic Society) | <3 19 46.1±9.6 | 15 28.7±5.3 | <0.001 | 0.95 |                      |                                |

PR values were averaged across the 2 Stroop tasks. Clinical categories were identified based on the significant categories in Table 1. Subgroups were chosen based on largest overlap between patients with AHF and stable patients. Statistical significance assessed using Mann–Whitney U tests. AHF indicates acute heart failure; OP, outpatient clinic; and PR, pause ratio.
hemodynamic congestion at discharge (Figure 4B). Body mass index decreased in 85% of patients with AHF (mean change=−1.15±1.12 kg/m², N=55, \( P < 0.001 \), Figure 4C) as excess fluid was reduced under diuretic treatment.

We did not observe an association between the change in NT-proBNP, weight, and PR. Of the 7 patients who showed an increase in NT-proBNP, only 1 patient also showed weight gain; 4 patients had higher PR at discharge (Figure 4D).

DISCUSSION

Recently, noninvasive voice biomarkers have been suggested for early recognition of HF decompensation. We performed automated speech and pause analyses in patients with stable HF and AHF. Compared with

| Table 4. Bivariate Association Between Clinical Features and PR |
|---------------------------------------------------------------|
| Clinical                                              | N  | Main effect Cohort: Coef (SE) | \( P \) value | Main effect Clinical: Coef (SE) | \( P \) value |
|---------------------------------------------------------|----|-------------------------------|--------------|-------------------------------|--------------|
| NYHA (ord.)                                             | 104| −14.4 (2.1)                   | <0.001       | −0.05 (1.15)                  | 0.96         |
| Edema (ord.)                                            | 104| −15.0 (1.7)                   | <0.001       | −0.77 (0.84)                  | 0.36         |
| Lung auscultation (bin.)                                | 102| −13.4 (2.0)                   | <0.001       | 1.38 (1.96)                   | 0.71         |
| (log) - N-terminal pro-B-type natriuretic peptide (cont.) | 104| −12.5 (4.7)                   | <0.001       | 1.16 (0.56)                   | 0.041        |
| Visual analog scale (cont.)                             | 56 | −15.4 (3.1)                   | <0.001       | 0.15 (0.55)                   | 0.78         |

Coefficients were estimated for each clinical feature individually using a linear model of the form \( \text{PR} = \text{Intercept} + \text{Cohort} + \text{Clinical} \). The intercept is not reported. The entries in the predictor Cohort are 1 for stable outpatient patients and 0 otherwise. The entries in the predictor Clinical correspond to the clinical feature values of a patient. For categorical clinical features (NYHA, Edema), we estimated the linear trend by transforming the categories to an ordinal variable (ord.). Coeff indicates mean regression coefficient; NYHA, New York Heart Association; and PR, pause ratio.

Figure 3. Visualization of PR across clinical categories and features. A through C, Association between PR and the categorical clinical features (A) NYHA class, (B) edema grade, and (C) normal lung auscultation findings. The numbers above the boxes summarize the number of patients (single dots) in the respective group. D and E Association between PR and the continuous clinical features (D) NT-proBNP (the gray shaded area) represents the diagnostic criteria (see methods section) and (E) self-rated dyspnea (VAS, the gray shaded area represents the interquartile range of VAS reported by stable patients, see Table 1). AHF indicates acute heart failure; NT-proBNP, N-terminal pro-B-type natriuretic peptide; NYHA, New York Heart Association; OP, outpatient clinic; PR, pause ratio (averaged across the 2 Stroop tasks); and VAS, visual analog scale.
patients with stable HF, patients with AHF spoke slower and showed a reduction in articulation time as well as an increased pause time to accommodate longer speech tasks. Across the 5 different speech tests, the Stroop reading tests showed the most consistent alterations of RST and PR in patients with AHF, thus suggesting that these tests be used to detect measurable changes in patients with HF. The shortened articulation and prolonged pause time in AHF make PR more sensitive to detect these differences than MAD, MPD, and RST.

HF is a chronic disease, and long-term management can be very challenging. Much effort has been put into maintaining patients’ stability and preventing episodes of decompensation. Studies on BNP- or NT-proBNP-guided treatment in chronic HF have demonstrated promising but inconsistent results.19,20 Remote hemodynamic monitoring has shown great promise in identifying increases in pulmonary pressures that might precede the symptom worsening by days to weeks.21–23 However, this requires invasive implantation of the pressure sensor and is hence mainly reserved for patients with more advanced HF or recent HF-related hospitalization. In addition, the approach is very costly. In view of the prevalence of HF, traditional biomarker analyses and implantation of pressure sensors would cause significant economic burden.

Despite patients with HF being trained to carefully monitor their clinical symptoms such as weight gain, peripheral edema, and dyspnea, a previous study demonstrated that 1 week before hospitalization, only around 50% of patients experienced worsening of edema or dyspnea, and only 30% had weight gain.24 We found PR, the proportion of pause time in Stroop tasks, to be elevated in AHF, accounting for >40% of the total task time. Importantly, PR ≥40% was specific to patients with AHF and largely independent of clinical symptoms. Furthermore, we observed a significant association of PR with levels of NT-proBNP, indicating that PR is related to the degree of congestion. These findings suggest that PR could be a useful additional marker in the monitoring of AHF and potentially serve as an early indicator of HF decompensation in otherwise asymptomatic patients.

Previous studies have demonstrated changes of different vocal characteristics in individual patients as they undergo treatment.6,7,25,26 However, the change of PR between admission and discharge was only modest in our study. This dichotomy between large group differences but modest within-patient change is in line with a recent finding by Sara and colleagues.27 The authors found that a vocal biomarker was associated with the presence of coronary artery disease, but they...
did not observe a change in the biomarker from baseline to follow-up. They speculated that voice may reflect an integrated index of well-being.\textsuperscript{27} In our study, one possible explanation is that PR also captures cognitive changes that go alongside congestion, like mental fatigue, which may persist at discharge. In fact, it has been reported that Stroop task performance is altered in chronic fatigue syndrome\textsuperscript{28} and is affected by a number of demographic factors, including age.\textsuperscript{29} Crucially, PR remained significantly elevated in patients with AHF when controlling for the age difference in the 2 cohorts (PS matching analysis). Another explanation could be that the patients with AHF were not yet fully compensated at the time of discharge. Indeed, NT-proBNP remained elevated in most of the patients with AHF at discharge, both indicating an incomplete level of decongestion and highlighting the need for a means to closely monitor these patients after discharge.

Study Limitations
Several limitations of the study should be noted. First, this is an observational study with a cohort that was mostly male, and the patient cohorts naturally differed in their disease characteristics. Second, this is a single-center study with a small sample size. Third, tests were performed in German. Future studies are needed to investigate PR in different patient populations using different languages. Finally, voice might change differently during the period of decompensation compared with after stabilization. Understanding these changes for individual patients will require further longitudinal investigations.

CONCLUSIONS
In summary, we evaluated speech and pause alterations in patients with acute and chronic HF and identified PR as an important additional marker of AHF decompensation. Our findings provide novel opportunities for future research to validate the usefulness of PR as a vocal biomarker to monitor patients with HF and prevent acute decompensated HF.

ARTICLE INFORMATION
Received June 3, 2022; accepted September 19, 2022.

Affiliations
Pharma Research and Early Development Informatics, Roche Innovation Center, Basel, Switzerland (D.S., Y.Z., M.S.); Division of Cardiology, Department of Medicine, University Hospital Basel, Basel, Switzerland (J.K., M.A., O.P., P.v.H., Q.Z.); Roche Diabetes Care, Basel, Switzerland (P.v.H.); and Department of Cardiology and Angiology I, Heart Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany (Q.Z.).

Acknowledgments
We thank Detlef Wolf for supporting the web application development and Mischa Huber for providing the devices for this project. Furthermore, we thank Eoin O’Connor and Betty Tai Lorentzen for their critical reading of the article.

Sources of Funding
Dr Zhou has received research support from Roche for the submitted work. Dr Schöbi was funded by the Roche Postdoctoral Fellowship program.

Disclosures
Dr Zhang, M. Aissani, and Dr van Haelst are current full-time employees at F. Hoffman-La Roche Ltd. Dr Pfister had speaker engagements or received research support from Abbott, Astra Zeneca, Bayer, Boehringer Ingelheim, Daiichi Sankyo, Novartis, Vifor, and Pfizer outside the submitted work. Dr Zhou had speaker engagements with Abbott, Novartis, and Vifor and has received research support from Abbott outside the submitted work. All other authors have no competing interest to declare.

REFERENCES
1. Ponikowski P, Anker SD, Alhajib KB, Cowie MR, Force TL, Hu S, Jaarsma T, Krum H, Rastogi V, Rohde LE, et al. Heart failure: preventing disease and death worldwide. ESC Heart Fail. 2014;1:4–25. doi: 10.1002/ehf2.12005
2. Crespo-Leiro MG, Anker SD, Maggioni AP, Coats AJ, Filippatos G, Ruschitzka F, Ferrari R, Pippiol MF, Delgado Jimenez JF, Metra M, et al. European Society of Cardiology Heart Failure Long-Term Registry (ESC-HF-LT). 1-year follow-up outcomes and differences across regions. Eur J Heart Fail. 2016;18:613–625. doi: 10.1002/ejhf.566
3. Ambrosy AP, Fonarow GC, Butler J, Chioncel O, Greene SJ, Vaduganathan M, Nodari S, Lam CSP, Sato N, Shah AN, et al. The global health and economic burden of hospitalizations for heart failure: lessons learned from hospitalized heart failure registries. J Am Coll Cardiol. 2014;63:1123–1133. doi: 10.1016/j.jacc.2013.11.083
4. Setoguchi S, Stevenson LW, Schneeweiss S. Repeated hospitalizations predict mortality in the community population with heart failure. Am Heart J. 2007;154:260–266. doi: 10.1016/j.ahj.2007.01.041
5. Maor E, Sara JD, Orbelo DM, Lerman LO, Levanon Y, Lerman A. Voice signal characteristics are independently associated with coronary artery disease. Mayo Clin Proc. 2018;93:840–847. doi: 10.1016/j.mayocp.2017.12.025
6. Murton OM, Hillman RE, Mehta DD, Semigram M, Daher M, Cunningham T, Verkouw K, Tabbatabi S, Steiner J, Dec GW, et al. Acoustic speech analysis of patients with compensated heart failure: a pilot study. J Accust Soc Am. 2017;142:EL401–EL407. doi: 10.1121/1.5007092
7. Amir O, Abraham WT, Azzam ZS, Berger G, Anker SD, Pinney SP, Burkhoff D, Shaloum ID, Lotan C, Edelman ER. Remote speech analysis in the evaluation of hospitalized patients with acute decompensated heart failure. JACC Heart Fail. 2022;10:41–49. doi: 10.1016/j.jchf.2021.08.008
8. Maor E, Perry D, Mevorchach T, Tablumb Y, Luz Y, Mazin I, Lerman A, Koren G, Shalev V. Vocal biomarker is associated with hospitalization and mortality among heart failure patients. J Am Heart Assoc. 2020;9:e013359. doi: 10.1161/JAHA.119.013359
9. Stroop JR. Studies of interference in serial verbal reactions. J Exp Psychol. 1935;18:643–662. doi: 10.1037/h0054561
10. Singh S, Pandey A, Neeland IJ. Diagnostic and prognostic considerations for use of natriuretic peptides in obese patients with heart failure. Prog Cardiovasc Dis. 2020;63:649–655. doi: 10.1016/j.pcad.2020.09.006
11. Wilson RC, Jones PW. A comparison of the visual analogue scale and modified Borg scale for the measurement of dyspnoea during exercise. Clin Sci. 1989;76:277–282. doi: 10.1042/cs0760277
12. Mahler DA, Weinberg DH, Wells CK, Feinstein AR. The measurement of dyspnea. Contents, interobserver agreement, and physiologic correlates of two new clinical indexes. Chest. 1984;85:751–758. doi: 10.1378/chest.85.6.751
13. Scarpina F, Tagini S. The stroop color and word test. Front Psychol. 2017;8:557. doi: 10.3389/fpsyg.2017.00557
14. Skodda S, Schlegel U. Speech rate and rhythm in Parkinson's disease. Mov Disord. 2008;23:985–992. doi: 10.1002/mds.21996
15. Hlavnička J, Čmejla R, Tykalová T, Šonka K, Růžička E, Ruzs J. Automated analysis of connected speech reveals early biomarkers of Parkinson’s disease in patients with rapid eye movement sleep behaviour disorder. Sci Rep. 2017;7:12. doi: 10.1038/s41598-017-00047-5
16. Vargha A, Delaney HD. A critique and improvement of the “CL” common language effect size statistics of McGraw and Wong. J Educ Behav Stat. 2000;25:101–132. doi: 10.2307/1163239
17. Elze MC, Gregson J, Baber U, Williamson E, Sartori S, Mehran R, Nichols M, Stone GW, Pocock SJ. Comparison of propensity score methods and covariate adjustment; evaluation in 4 cardiovascular studies. J Am Coll Cardiol. 2017;69:345–357. doi: 10.1016/j.jacc.2016.10.080

18. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol. 2013;4:663. doi: 10.3389/fpsyg.2013.00863

19. Troughton RW, Frampton OM, Brunner-La Rocca H-P, Pfisterer M, Eurlings LWM, Emterl H, Persson H, O’Connor CM, Moertl D, Karlström P, et al. Effect of B-type natriuretic peptide-guided treatment of chronic heart failure on total mortality and hospitalization: an individual patient meta-analysis. Eur Heart J. 2014;35:1559–1567. doi: 10.1093/eurheartj/ehu090

20. Felker GM, Anstrom KJ, Adams KF, Ezekowitz JA, Fiuza M, Houston-Miller N, Januzzi JL Jr, Mark DB, Pita LL, Passmore G, et al. Effect of natriuretic peptide-guided therapy on hospitalization or cardiovascular mortality in high-risk patients with heart failure and reduced ejection fraction: a randomized clinical trial. JAMA. 2017;318:713–720. doi: 10.1001/jama.2017.10565

21. Abraham WT, Stevenson LW, Bourge RC, Lindenfeld JA, Bauman JG, Adamson PB. Sustained efficacy of pulmonary artery pressure to guide adjustment of chronic heart failure therapy: complete follow-up results from the CHAMPION randomised trial. Lancet. 2016;387:453–461. doi: 10.1016/S0140-6736(15)00723-0

22. Abraham WT, Adamson PB, Bourge RC, Aaron MF, Costanzo MR, Stevenson LW, Strickland W, Neelagari S, Raval N, Krueger S, et al. Wireless pulmonary artery haemodynamic monitoring in chronic heart failure: a randomised controlled trial. Lancet. 2011;377:658–666. doi: 10.1016/S0140-6736(11)60101-3

23. Angermann CE, Assmus B, Anker SD, Asselbergs FW, Brachmann J, Brett M-E, Brugts JJ, Ertl G, Ginn G, Hilker L, et al. Pulmonary artery pressure-guided therapy in ambulatory patients with symptomatic heart failure: the CardioMEMS European Monitoring Study for Heart Failure (MEMS-HF). Eur J Heart Fail. 2020;22:1891–1901. doi: 10.1002/ejhf.1943

24. Schiff GD, Fung S, Speroff T, McNutt RA. Decompensated heart failure: symptoms, patterns of onset, and contributing factors. Am J Med. 2003;114:625–630. doi: 10.1016/S0002-9343(03)00132-3

25. 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. doi: 10.3109/0886022X.2014.986108

26. Amir O, Anker SD, Gork I, Abraham WT, Pinney SP, Burkhoff D, Shalmon ID, Haviv R, Edelman ER, Lotan C. Feasibility of remote speech analysis in evaluation of dynamic fluid overload in heart failure patients undergoing haemodialysis treatment. ESC Heart Fail. 2021;8:2467–2472. doi: 10.1002/ehf2.13367

27. Sara JDS, Maor E, Orbelo D, Gulati R, Lerman LO, Lerman A. Noninvasive voice biomarker is associated with incident coronary artery disease events at follow-up. Mayo Clin Proc. 2022;97:835–846. doi: 10.1016/j.mayocp.2021.10.024

28. Metzger FA, Denney DR. Perception of cognitive performance in patients with chronic fatigue syndrome. Ann Behav Med. 2002;24:106–112. doi: 10.1207/s15324796abm2402_07

29. Van der Elst W, Van Boxtel MP, Van Breukelen GJ, Jolles J. The Stroop color-word test: influence of age, sex, and education; and normative data for a large sample across the adult age range. Assessment. 2006;13:62–79. doi: 10.1177/1073191105283427