

RESEARCH ARTICLE

Validated automatic speech biomarkers in primary progressive aphasia

Naomi Nevler¹ , Sharon Ash¹, David J Irwin¹, Mark Liberman² & Murray Grossman¹¹Penn Frontotemporal Degeneration Center, Department of Neurology, University of Pennsylvania, Philadelphia, Pennsylvania²Linguistic Data Consortium, Department of Linguistics, University of Pennsylvania, Philadelphia, Pennsylvania**Correspondence**

Naomi Nevler and Murray Grossman,
Department of Neurology – 3 West Gates,
Hospital of the University of Pennsylvania,
3400 Spruce St, Philadelphia, PA 19104-
4283.
Tel: 215-662-3361; Fax: 215-349-8464;
E-mails: mgrossma@pennmedicine.upenn.edu
and naomine@pennmedicine.upenn.edu

Funding Information

This work was supported in part by the
National Institutes of Health (AG017586,
AG038490, AG053940, AG052943,
NS088341, DC013063, AG054519), the
Institute on Aging, the Alzheimer's
Association (AACSF-18-567131), and the
Wyncote Foundation.

Received: 15 July 2018; Revised: 21 August
2018; Accepted: 22 August 2018

*Annals of Clinical and Translational
Neurology* 2019; 6(1): 4–14

doi: 10.1002/acn3.653

Abstract

Objective: To automatically extract and quantify specific disease biomarkers of prosody from the acoustic properties of speech in patients with primary progressive aphasia. **Methods:** We analyzed speech samples from 59 progressive aphasic patients (non-fluent/agrammatic = 15, semantic = 21, logopenic = 23; ages 50–85 years) and 31 matched healthy controls (ages 54–89 years). Using a novel, automated speech analysis protocol, we extracted acoustic measurements of prosody, including fundamental frequency and speech and silent pause durations, and compared these between groups. We then examined their relationships with clinical tests, gray matter atrophy, and cerebrospinal fluid analytes. **Results:** We found a narrowed range of fundamental frequency in patients with non-fluent/agrammatic variant aphasia (mean 3.86 ± 1.15 semitones) compared with healthy controls (6.06 ± 1.95 semitones; $P < 0.001$) and patients with semantic variant aphasia (6.12 ± 1.77 semitones; $P = 0.001$). Mean pause rate was significantly increased in the non-fluent/agrammatic group (mean 61.4 ± 20.8 pauses per minute) and the logopenic group (58.7 ± 16.4 pauses per minute) compared to controls. In an exploratory analysis, narrowed fundamental frequency range was associated with atrophy in the left inferior frontal cortex. Cerebrospinal level of phosphorylated tau was associated with an acoustic classifier combining fundamental frequency range and pause rate ($r = 0.58$, $P = 0.007$). Receiver operating characteristic analysis with this combined classifier distinguished non-fluent/agrammatic speakers from healthy controls (AUC = 0.94) and from semantic variant patients (AUC = 0.86). **Interpretation:** Restricted fundamental frequency range and increased pause rate are characteristic markers of speech in non-fluent/agrammatic primary progressive aphasia. These can be extracted with automated speech analysis and are associated with left inferior frontal atrophy and cerebrospinal phosphorylated tau level.

Introduction

Conversational speech is essential to our daily lives and allows us to vocalize thoughts and emotions in order to communicate a message to a listener. While language is often studied by analyses of segmental content such as words and sentences,^{1,2} speech involves the additional component of prosody. Prosody refers to suprasegmental aspects of speech, encompassing intonation, rhythm, and stress properties that are crucial for conveying linguistic and emotional information.³

Despite our natural sensitivity to prosodic features of speech, studies of its pathological form, dysprosody, are rare. This may stem from difficulties quantifying features of prosody in an objective manner. Most research on prosody has relied on subjective assessments, often focusing on the expression or comprehension of emotional speech.^{4–8} We developed an automated technique for speech analysis, based on a Speech Activity Detector (SAD),⁹ which we implemented to examine the prosodic characteristics of a semistructured speech sample in patients with variants of primary progressive aphasia

(PPA). We aimed to investigate the behavioral and neurobiologic basis for dysprosody in these patients while testing the implementation of our automated speech analysis method. We hypothesized distinct acoustic dysprosodic markers in variants of PPA. In particular, in the non-fluent/agrammatic variant of PPA (naPPA), there is impairment in constructing well-formed sentences,¹ which could impair these patients' ability to utilize appropriate prosody and limit their overall intonational range. Since this is the most dysfluent phenotype, we also expected to find more frequent pauses. We expected to relate these changes to specific biologic markers of pathology frequently associated with naPPA, including inferior frontal atrophy and a cerebrospinal fluid (CSF) surrogate of Frontotemporal Lobar Degeneration (FTLD) pathology involving the accumulation of misfolded tau (FTLD-tau).

Methods

Subjects

We examined digitized speech samples from 67 native English speakers who met published clinical consensus criteria for a specific PPA syndrome,¹⁰ including naPPA ($n = 18$), semantic variant PPA (svPPA, $n = 23$), logopenic variant PPA (lvPPA, $n = 26$), and 37 healthy controls (HC). All patients were assessed between April 1998 and September 2017 by experienced neurologists (MG, DJI) in the Department of Neurology at the Hospital of the University of Pennsylvania and were reviewed by a consensus conference according to published criteria,¹⁰ modified for lvPPA.¹¹ For this study, we excluded patients with a concurrent motor disorder such as progressive supranuclear palsy (PSP), corticobasal syndrome (CBS), or amyotrophic lateral sclerosis (ALS), to minimize potential motor confounds in our acoustic analyses. Nevertheless, seven naPPA patients had dysarthria or apraxia of speech (AoS), as determined by experienced neurologists (MG, DJI). Fifteen of the svPPA cases had concomitant behavioral symptoms, but their speech acoustic pattern did not differ from their counterparts with isolated svPPA. We reviewed all speech samples with a pitch range above or below 1.5 SD of their group mean and detected seven patients (three naPPA, two lvPPA, two svPPA) and six controls with extensive vocal-fry or "creaky voice". These vocal characteristics carry a high probability for pitch-tracking errors and so we excluded these 13 recordings from further analysis. Another lvPPA recording was excluded due to participation in an AD disease-modifying treatment trial. The final groups, totaling 59 PPAs and 31 HCs, were matched in all demographic characteristics except disease duration, which was shorter in naPPA compared to the other PPA groups

(Table 1). Additional neuropsychological test data and manually coded linguistic data¹ are presented in Table 1 to confirm typical characteristics of each patient group.

Speech samples

We used the Cookie Theft picture description task from the Boston Diagnostic Aphasia Examination¹² to elicit semistructured narrative speech samples.¹ Please refer to the supplement for details on speech sample collection. Characteristics of speech reported previously^{1,2} in these phenotypes include speech rate measured as words per minute (wpm), grammatical complexity reflected in dependent clauses per utterance (DC), mean length of utterance (MLU), and well-formed sentences per utterance (WFS), and we analyzed these manually.

Sound processing

We used a SAD developed at the University of Pennsylvania Linguistic Data Consortium (LDC)⁹ to time-segment the audio files and then pitch-tracked the segments of continuous speech, using a protocol described previously.¹³ We extracted the fundamental frequency (f_0 , see supporting information), the durations of speech, and silent pause segments. From these, we calculated the following measures: f_0 range (see supporting information), mean speech segment and pause segment durations, and pause rate, which was calculated as the number of pauses per minute (ppm) over total speech time. We validated our automatic measurements by comparing its results to a blinded assessment of restricted versus normal f_0 range performed by experienced human raters (NN and SA). Inter-rater agreement was substantial (Cohen's kappa = 0.81) and the cases of disagreement were reviewed and discussed until an agreement was reached. We compared these judgments to PPA subgroups formed by using a cutoff for normal f_0 range at 4.8 semitones (ST), based on an ROC analysis for all PPA patients versus controls. A chi-square test showed no difference in the distributions of the normal and restricted f_0 range categories when using the automated analysis compared to the subjective evaluation ($X^2=1.48$, $df = 1$, $P = 0.22$).

Analysis of likely pathology

Thirty-seven of our patients had a CSF sample collected within 5–39 months (mean 10.6) of cookie theft speech recording. Following a pathologically validated algorithm, we screened for a non-Alzheimer's disease (AD) CSF profile ($p\text{-Tau}/A\beta < 0.09$, available in 32 samples).^{14,15} This procedure identified 20 cases with a CSF

Table 1. Mean (SD) demographic characteristics of patients and controls.

	HC	naPPA	lvPPA	svPPA	P
<i>n</i>	31	15	23	21	
Age, y	69.29 (7.90)	69.67 (9.20)	65.91 (9.83)	64.48 (7.71)	0.14
Sex = Male (%)	11 (35.5)	6 (40.0)	7 (30.4)	10 (47.6)	0.68
Education, y	15.97 (2.58)	14.80 (3.12)	15.35 (3.19)	15.10 (2.81)	0.56
Disease duration, y	NA	2.60 (1.12)	4.00 (2.00)	4.05 (2.04)	0.04
MMSE total (0-30), <i>n</i> = 85	29.00 (1.07)	24.73 (5.24)	23.05 (5.72)	23.05 (6.11)	<0.001 ¹
PBAC Naming (0-6), <i>n</i> = 37	5.50 (0.71)	5.78 (0.67)	4.00 (1.83)	1.23 (1.48)	<0.001
F letter fluency, <i>n</i> = 41	17.75 (8.10)	6.33 (3.04)	6.36 (5.40)	8.21 (3.96)	0.001
Digit span forward, <i>n</i> = 69	7.00 (1.37)	5.61 (1.30)	4.45 (1.54)	6.06 (1.89)	<0.001
Digit span backward, <i>n</i> = 74	5.65 (1.31)	2.64 (1.11)	2.91 (1.08)	3.78 (1.70)	<0.001
Category fluency ⁴ , <i>n</i> = 54	19.67 (6.48)	10.11 (5.09)	9.85 (5.94)	5.36 (4.67)	<0.001
Total speech time ² (sec)	49.66 (18.23)	33.78 (18.53)	38.54 (15.02)	36.42 (13.85)	0.006
Total word count ³	166.32 (63.90)	65.80 (37.06)	114.13 (56.80)	134.05 (55.99)	<0.001
Speech rate, wpm	140.06 (36.74)	61.00 (24.85)	88.17 (36.09)	113.95 (40.76)	<0.001
MLU (words)	10.57 (1.98)	6.74 (2.38)	8.46 (2.45)	8.61 (2.75)	<0.001
DC/utterance ⁵	0.37 (0.23)	0.05 (0.09)	0.21 (0.21)	0.32 (0.27)	<0.001
WFS/utterance	0.91 (0.11)	0.72 (0.32)	0.71 (0.25)	0.78 (0.19)	0.003

wpm, words per minute; MLU, mean length of utterance; DC, dependent clauses; WFS, well-formed sentences; PBAC, Philadelphia Brief Assessment of Cognition.

¹MMSE total score did not differ between patient groups.

²This refers to the sum of all subject's speech segment durations, including verbal and nonverbal vocalizations, all available for pitch-tracking.

³This manual word count includes only verbal vocalizations that were comprehensible enough for transcription.

⁴Category = animals

⁵Data refer to the average number of clauses per utterance.

profile suggestive of non-AD FTLD underlying pathology.¹⁴ These included two autopsy-confirmed cases (one Tau, one TDP) and a third case with confirmed *MAPT* mutation.

To determine the association of speech features with *in vivo* measures of pathology, we examined the relationship between our acoustic variables and CSF biomarkers including beta-Amyloid (A β), total (t-Tau) and phosphorylated Tau (p-Tau) in this subset of high-probability FTLD pathology patients. We tested the effect of a combined acoustic parameter (see below) on each of these CSF analytes, applying multivariate regression analysis techniques (see below).

Statistical analysis

Demographic data were compared with analysis of variance (ANOVA) for continuous variables and the chi-square test for categorical variables. We used Kernel density and Q-Q plots to examine speech and cognitive variable distributions. Since these were normally distributed, we used ANOVA for between-group comparisons, covarying for disease duration, and post hoc tests with Tukey's Honest Significant Difference (HSD). Groups were compared for their *f*₀ range, speech segment duration, and pause rate. Because of the effect of sex on *f*₀, an additional *f*₀ analysis was conducted within male

and female subpopulations covarying for disease duration. MMSE total scores differed between our male patient groups, and so we also introduced MMSE as a covariate in their analysis.

Within the naPPA group, we compared patients with motor symptoms such as dysarthria or AoS (see below) to those without these speech features using a Student's *t* test. Simple correlations were performed with Pearson's method. Regression analyses included generalized linear models (GLMs) with log transformation for p-Tau levels as the outcome measure and a polynomial logistic regression for clinical phenotype as the outcome variable. GLM validation was based on residuals plots. A stepwise backward elimination approach was implemented in the p-Tau GLM in order to examine the effects of potential confounders (see results section) and find the best fit model.

We performed receiver operating characteristic (ROC) curve analyses on *f*₀ range and pause rate as acoustic classifiers for PPA phenotypes. These were tested individually and in combination (pause rate/*f*₀-range ratio, to control for opposite directionality) for patients versus controls and between patient groups. We used a bootstrap technique with 2000 permutations to compare ROC models of similar group-pairs. All calculations were conducted in RStudio¹⁶ with additional packages.^{17–25}

Gray matter (GM) density analysis

In an exploratory analysis, we assessed available high-resolution structural brain MRIs obtained on average within 1.5 ± 2.5 months of recording in 16 controls and 9 naPPA patients. The reasons for unavailability of an MRI scan included various contraindications for the test, absence of a T1 sequence, or a difficulty obtaining a good-quality scan. Clinical and demographic characteristics of these MRI subsets matched those of their original full sets. Details of data acquisition and preprocessing are reported in the supplement. We note our use of small, 2 mm isotropic voxels that are compatible with known cortical thickness of 2–3 mm, and that this biologically constrained technique results in many more statistical comparisons than traditional imaging studies and consequently less robust statistical results. We calculated GM density and then mapped naPPA atrophy compared to HC using voxel-wise comparisons in FSL^{26,27} with family-wise error correction and threshold-free cluster enhancement at a statistical threshold of $P \leq 0.01$ and cluster size threshold of $k \geq 50$ voxels. This stricter statistical threshold was selected in order to improve the likelihood that naPPA patients have an anatomic distribution of disease representative of naPPA. We then performed a regression analysis within the naPPA group's areas of cortical atrophy, covarying for age, and disease duration. We applied 10,000 permutations equivalent to statistical protection controlling for type I error and set a statistical threshold of $P \leq 0.05$ and cluster size threshold of $k \geq 10$ voxels.

Ethical considerations

All participants were enrolled in study protocols and participated in an informed consent procedure approved by the Institutional Review Board of the University of Pennsylvania. All personnel exposed to personal patient data, including voice samples, have been specifically trained in ethical handling of patient data.

Results

Speech parameter results

We found a significantly reduced f0 range only in naPPA (mean 3.86 ± 1.15 ST) compared with HC (mean 6.06 ± 1.95 ST; $P < 0.001$) and svPPA (mean 6.12 ± 1.77 ST; $P = 0.001$, Fig. 1A). Pause rate differed significantly between groups (Fig. 1B): each PPA group differed from HC (mean 32.24 ± 9.75 ppm; $P \leq 0.002$ per contrast). naPPA (mean 61.36 ± 20.8 ppm) differed from svPPA (mean 47.15 ± 14.34 ppm; $P = 0.02$), and lvPPA (58.74 ± 16.41 ppm) also differed from svPPA ($P = 0.04$).

Mean speech segment duration was reduced significantly in each patient group compared to HC ($P < 0.001$ for each contrast), but there were no significant differences among patient groups (Fig. 1C). Mean pause duration (overall mean 1.14 ± 0.7 sec) was similar in all groups (Fig. 1D).

Within naPPA, we compared patients with moderate to profound dysarthria ($n = 4$) or AoS ($n = 2$) or both ($n = 1$) to those without these features at the time of recording ($n = 8$). We found no differences in any measured acoustic marker between these subgroups (f0 range 4.06 ST with motor symptoms vs. 3.69 without them, $P = 0.6$; pause rate 67.4 ppm vs. 56.1 , respectively, $P = 0.3$; mean pause duration 1.2 sec vs. 1.6 , respectively, $P = 0.3$). Likewise, an analysis by sex revealed a comparable restriction of f0 range in the naPPA group within each gender (males: 3.19 ± 1.35 ST in naPPA vs. 5.96 ± 1.49 ST in HC, $P = 0.002$ and vs. 4.98 ± 1.18 ST in svPPA, $P = 0.06$; females: 4.31 ± 0.79 ST in naPPA vs. 6.11 ± 2.2 ST in HC, $P = 0.06$ and vs. 7.15 ± 1.58 ST in svPPA, $P = 0.004$).

We found significant correlations between pause rate and manually coded measures of fluency and grammaticality in all PPA patients (Fig. 2A–D), including: speech rate in wpm, WFS per utterance, DC per utterance, and MLU. We found a strong negative correlation between speech segment duration and pause rate across all patients ($r = -0.87$, $P < 0.001$). Concordantly, speech segment duration correlated with speech rate, WFS per utterance, DC per utterance, and MLU (Fig. 2E–H). f0 range correlated with speech rate in all PPA patients ($r = 0.29$, $P = 0.02$).

The narrow f0 range in naPPA did not correlate with any demographic or neuropsychological features (all P -values > 0.1). Similarly, pause rate and speech segment duration did not correlate with any demographic characteristics or cognitive measures. We found no cross-correlations between f0 range and our automatically derived duration measures of pause rate, pause and speech segment durations in any group (all P -values > 0.1). No correlation between the acoustic measures and manually coded features was found within the naPPA group.

In a regression model of f0 range and pause rate as main predictors of clinical phenotype, the likelihood of disease (naPPA) with a 1 ST increase in f0 range and constant pause rate was 0.35 (Table 2). Thus, a reduction of the measured f0 range by 1 ST, using the reciprocal of the regression, would increase the likelihood of the diagnosis of naPPA by a factor of 2.9 ($1/0.35$).

Finally, ROC curve analyses showed that f0 range as a single predictor of naPPA versus HC has an area under the curve (AUC) of 0.84 (95% CI: 0.72–0.96) and best threshold at 4.8 ST, while pause rate showed an AUC = 0.89 (95% CI: 0.75–1.00) and best threshold at 52.3 ppm, with no statistically significant difference

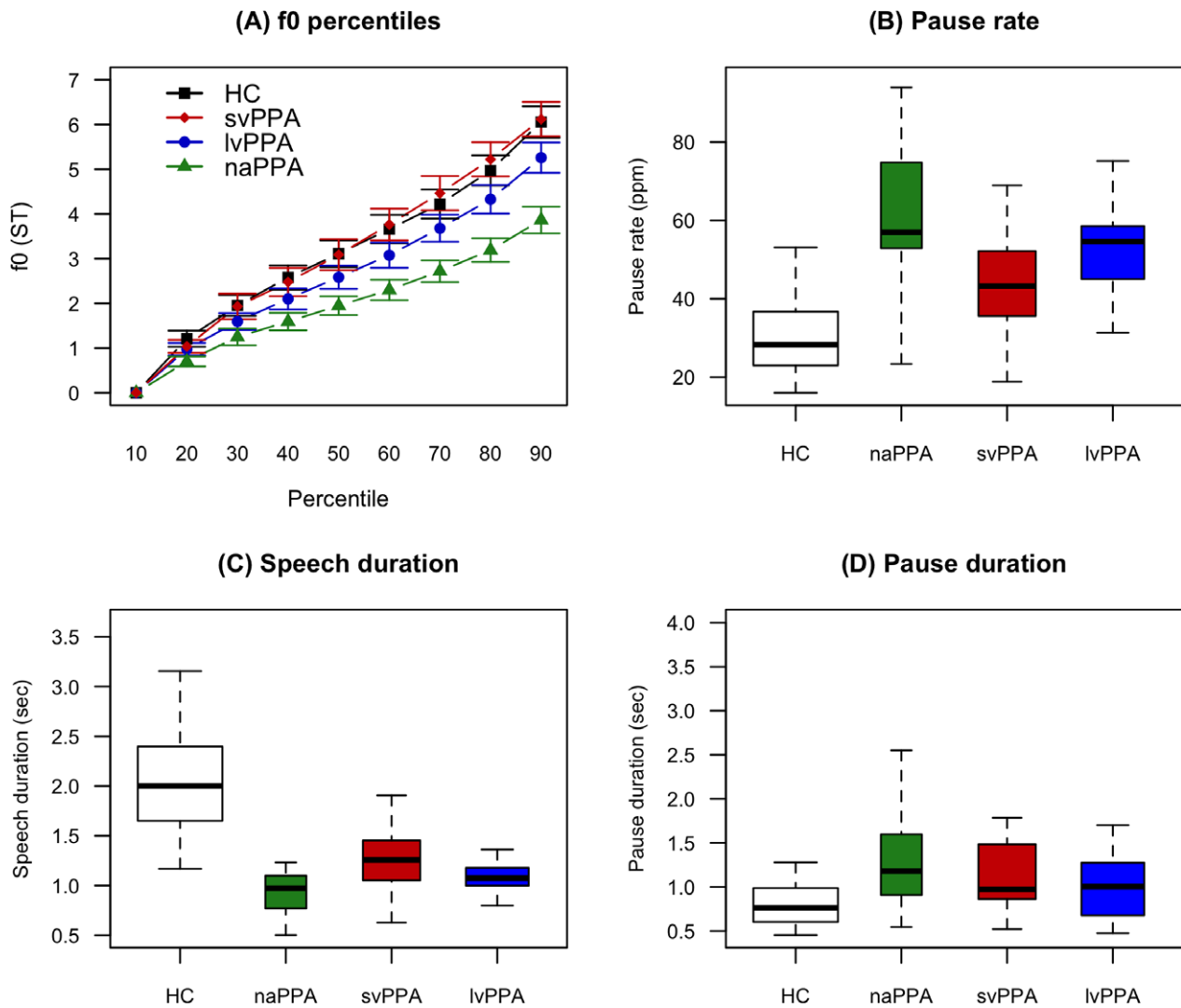


Figure 1. f0 and durations data. (A) f0 percentiles by clinical phenotype, expressed in semitones (ST). The 90th percentile represents the f0 range. (B) Pause rate, calculated as the number of pauses per minute of speech time. (C) Mean speech duration. (D) Silent pause mean duration. f0, fundamental frequency; ST, semitones; HC, healthy controls; lvPPA, logopenic variant Primary progressive aphasia; naPPA, non-fluent/agrammatic primary progressive aphasia; svPPA, semantic variant primary progressive aphasia; sec, seconds; ppm, pauses per minute.

between these two curves ($P = 0.6$, Fig. 3A). A combined acoustic parameter showed an AUC = 0.94 (95% CI: 0.87–1.00 at best threshold; sensitivity 87%, specificity 90%) distinguishing naPPA from HC (Fig. 3A–B). The same classifier distinguished naPPA from svPPA at an AUC = 0.86 (95% CI: 0.73–0.98 at best threshold; sensitivity 71%, specificity 87%), and distinguished naPPA from lvPPA with an AUC = 0.69 (95% CI: 0.50–0.87 at best threshold; sensitivity 87%, specificity 47%) (Fig. 3B).

Neuroimaging

The naPPA group showed bilateral frontotemporal atrophy, most prominently left frontal atrophy (Fig. S1A). In

an exploratory analysis, we associated f0 range within the naPPA group with GM atrophy in the left inferior frontal gyrus (IFG, Fig. S1B). MNI coordinates are provided in Table S1.

CSF results

We previously found CSF p-tau levels to correlate with the severity of postmortem tau pathology in FTLD.¹⁴ To determine if our automatically extracted speech variables were related to an in vivo marker of tau pathology, we examined the relationship between these speech features and CSF biomarkers in the subset of PPA patients with a CSF profile suggestive of FTLD pathology ($n = 20$).

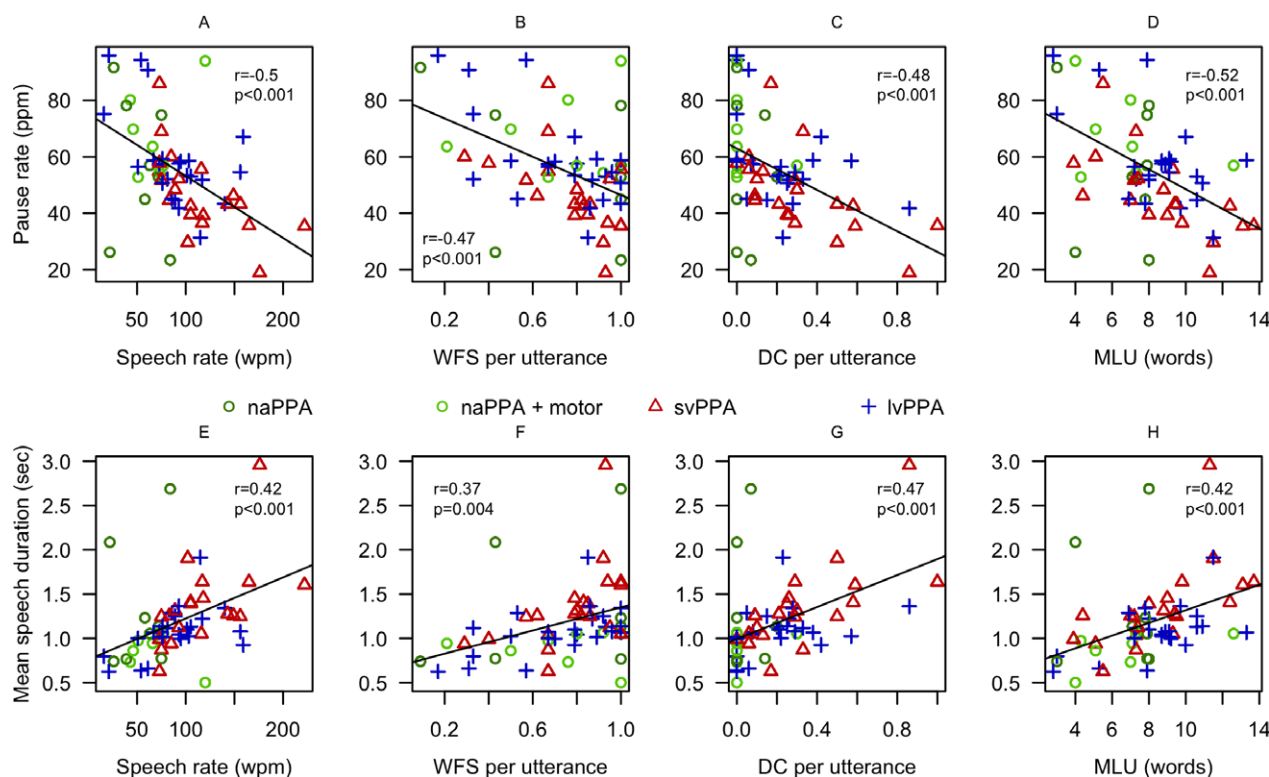


Figure 2. Correlations of automated measures with manual coding. A–D Correlations of automatically extracted pause rate with manually coded measures of fluency and grammaticality. E–H Correlations of automatically extracted mean speech segment duration with manual coding. The mirror image between the upper and lower panels coincides with the strong negative correlation between speech duration and pause rate (see text). lvPPA, logopenic variant primary progressive aphasia; naPPA, non-fluent/agrammatic primary progressive aphasia; svPPA, semantic variant primary progressive aphasia; sec, seconds; ppm, pauses per minute; wpm, words per minute; WFS, well-formed sentences; DC, dependent clauses.

Table 2. Results of polynomial logistic regression.

	f0 range			Pause rate		
	OR	95% CI	<i>P</i>	OR	95% CI	<i>P</i>
naPPA	0.35	0.18–0.69	0.002	1.18	1.10–1.26	<0.001
lvPPA	0.82	0.54–1.26	0.37	1.17	1.10–1.24	<0.001
svPPA	1.09	0.77–1.18	0.62	1.11	1.05–1.18	<0.001

Significant *P* values are in bold.

We found the natural logarithm of p-Tau levels was linearly associated with the natural logarithm of the combined acoustic parameter ($r = 0.58$, $P = 0.007$, Fig. 4). We tested the effect of potential confounding variables including age, disease duration, education, and the time interval between speech sample and CSF collection. These were found to have no significant effect (simple correlation was the best fit). We did not find an association of our prosodic marker with CSF biomarkers that were not directly associated with postmortem tau pathology (i.e., t-Tau and Abeta, data not shown).

Discussion

Our automated speech analysis protocol identified two basic acoustic markers that characterize patients with naPPA in a sensitive and specific way: f0 range, which correlates with perceived pitch; and pause rate, which is a measure of dysfluency. These were associated with left inferior frontal atrophy and CSF level of p-Tau. Speech analyses in naPPA thus may be an informative marker to screen for FTLD-Tau pathology.

Prosody is a distinct but integral element of spoken language, associated with neural networks supporting language.^{28–30} Although it plays a role in the phonological representation of some individual words in the auditory-aural system (e.g., stress provides the linguistic differentiation of “rècord” as a noun versus “recòrd” as a verb), prosody mainly contributes to suprasegmental aspects of sentence processing. For example, prosodic features mark the end of an utterance, distinguishing between a question and a statement (e.g., declining pitch for “you’re tired! versus ascending pitch for “you’re tired?”). Several aspects of

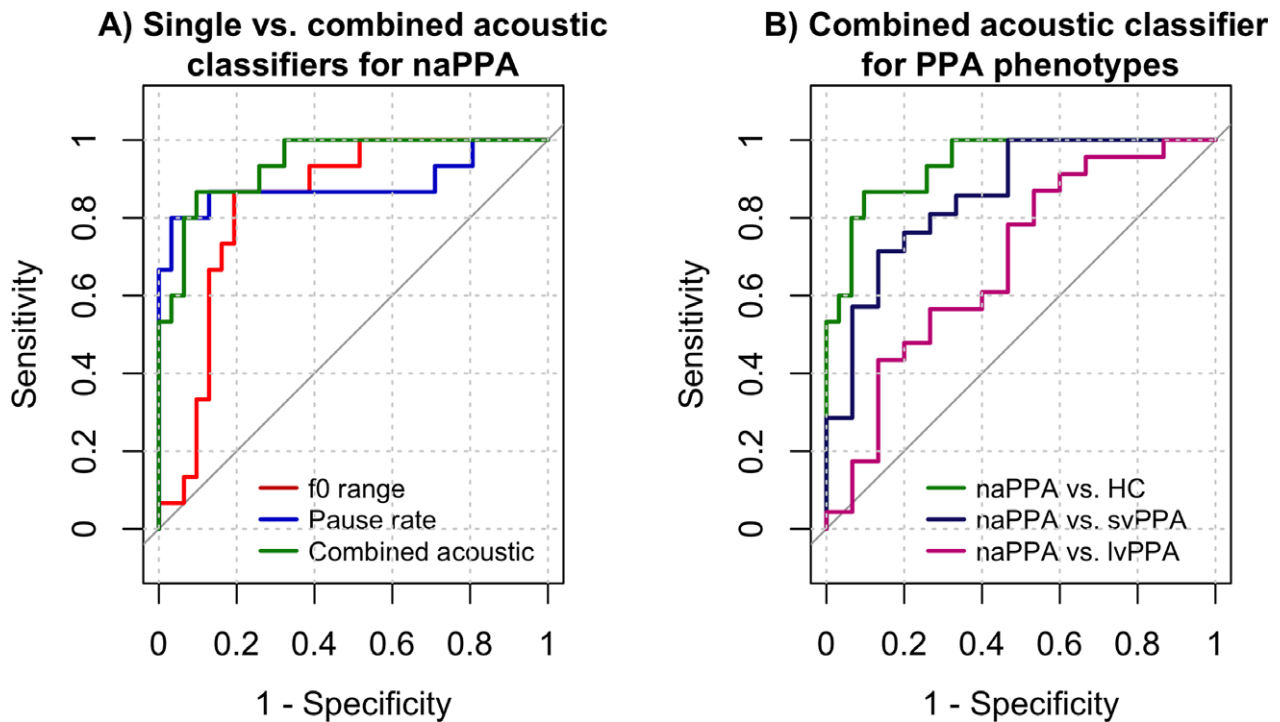


Figure 3. ROC analyses. (A) f0 range and pause rate as single classifiers for receiver operating characteristic (ROC) curve of naPPA vs. HC. A combined acoustic classifier (pause rate/f0 range) improves AUC (0.94). (B) Combined acoustic parameter (pause rate/f0 range) as classifier for naPPA vs. other phenotypes. AUC, area under the curve; HC, healthy controls; lvPPA, logopenic variant primary progressive aphasia; naPPA, non-fluent/agrammatic primary progressive aphasia; svPPA, semantic variant primary progressive aphasia; ROC, receiver operating curves.

sentence-level processing are disturbed in naPPA, including grammatical expression.^{1,2} Here we show that impaired intonation, expressed as reduced f0 range, and dysfluency as manifested in frequent pauses, are also observed in patients with naPPA. Indeed, we were able to associate our findings of limited prosodic expression with previously validated manual coding of grammaticality.^{1,2} Although naPPA patients use shorter sentences, this potential confounding variable does not apparently limit their intonational ability since there was no correlation between f0 range and mean speech duration or mean length of utterance. The correlation found between f0 range and speech rate may reflect a common expressive prosodic impairment affecting both fluency and intonation. This hypothesis requires further investigation. Impaired prosodic comprehension has also been reported in naPPA,^{31,32} and more specific studies are needed to relate expressive dysprosody to this observation.

Our finding of highest pause rate in the naPPA group coincides with previous speech analyses.^{33,34} We correlated pause rate with a manually coded measure of reduced speech rate (words/minute) that is associated with the characteristic effortfulness heard in naPPA speech. This validates the use of our automated algorithm, which does not depend on the time-consuming

generation of transcripts and makes use of natural breath-group boundaries (see below). While pause rate is increased in naPPA, pause duration and speech segment duration do not differ between groups. Thus, these duration measures cannot easily explain the impression of effortful, non-fluent speech in naPPA.

In our cohort, seven patients with naPPA had either dysarthria or AoS as part of their clinical presentation. AoS has been reported to affect some acoustic measures of patients' speech, specifically prolonged duration of stressed syllables.³⁵ This was unlikely to be a confound in our study. Speech duration was measured over entire breath-groups (see below), not within words, and f0 originates at the level of the vocal folds and is mainly a function of subglottal air pressure.³ Thus, impaired articulation due to difficulty coordinating the motor speech apparatus should minimally affect f0.^{3,36} f0 can also be affected by tension of the vocal folds.³ Possible involvement of the vocal folds may be manifested as a coarse or hypophonic voice. Such voice quality is also highly susceptible to pitch-tracking errors, and, as mentioned above, we excluded samples with these voice characteristics for that reason.

naPPA was characterized by reduced f0 range and increased pause rate. There was no colinearity between

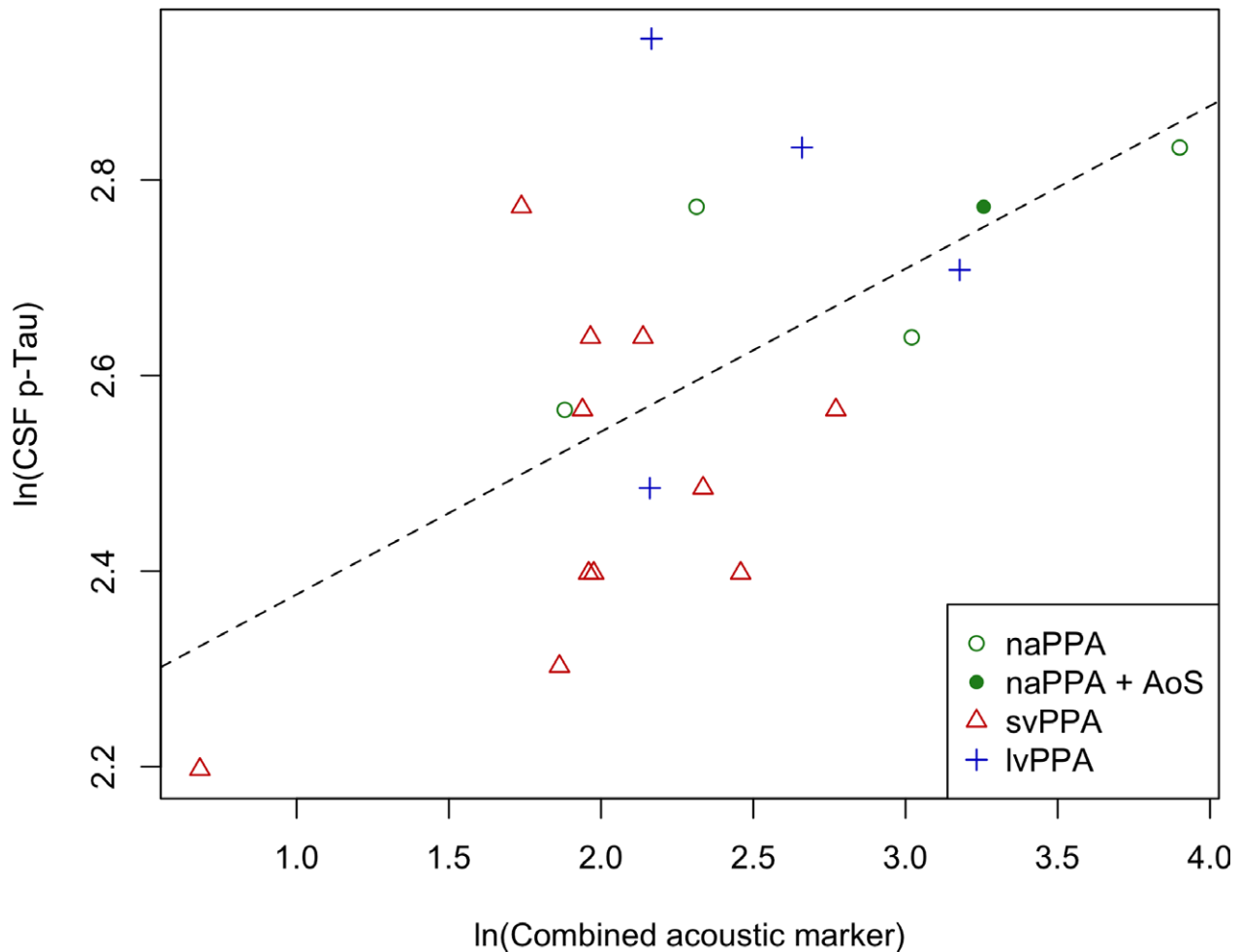


Figure 4. CSF p-Tau correlation with combined acoustic marker. Pearson correlation showing linear association between the natural logarithm of CSF p-Tau levels and the natural logarithm of the combined acoustic marker ($r = 0.58$, $P = 0.007$). lvPPA, logopenic variant primary progressive aphasia; naPPA, non-fluent/agrammatic primary progressive aphasia; svPPA, semantic variant primary progressive aphasia.

these two acoustic variables, even though f_0 range correlated with speech rate in the full set of PPA subjects. This suggests that the two acoustic parameters are relatively independent characteristics of the non-fluent variant of PPA. Thus, we sought to investigate whether these together can distinguish naPPA from healthy controls and from other PPA phenotypes. In addition to robustly discriminating naPPA from healthy controls, we found that the combination of these acoustic features can reasonably distinguish naPPA from svPPA, suggesting that these speech deficits are not nonspecific impairments found in any aphasic patient, but instead may be specific to a particular PPA phenotype. Others have used lengthy neuropsychological measures to distinguish between naPPA and svPPA.^{37,38} Our observations are consistent with the clinical impression that svPPA is associated with a relatively intact suprasegmental speech pattern and emphasize that the disorder in svPPA is most prominently at the

level of the representation of single word and object meaning.^{10,39,40}

We found previously that gender impacts f_0 range in bvFTD.¹³ However, we found that gender had no effect in PPA. This suggests that pitch range may interact with gender selectively in bvFTD as a component of their social disorder. While this observation emphasizes the importance of assessing f_0 in both males and females with neurodegenerative disorders, it does not appear that sex per se has a significant effect on prosody in PPA.

The acoustic characteristics of lvPPA were intermediate between those of naPPA and svPPA. The lvPPA phenotype has proven to be the least amenable to clinical identification,^{38,41} although recent studies have begun to characterize it more reliably.^{11,42} It is possible that all lvPPA patients have an attenuated version of the speech disorder found in naPPA. Alternatively, a subset of patients with lvPPA may exhibit some of the speech

characteristics of naPPA.³⁸ Additional work is needed to test these hypotheses in a larger cohort.

We associated f0 range impairment in naPPA with atrophy in left inferior frontal cortex. These results should be considered preliminary and interpreted with caution considering the small sample of only nine naPPA patients in the MRI subset. However, this preliminary observation coincides with our previous analysis in bvFTD,¹³ in which bilateral IFG involvement was established with a similar acoustic analysis. In the current study, we show the prosodic impairment in a group with non-fluent speech and grammatical deficits but no apparent social-behavioral impairment. Thus, a potential association of f0 range restriction with the left IFG would be most consistent with the hypothesis that this acoustic marker may reflect derangement of a system of linguistic expression. This hypothesis is supported by other published reports, such as in Wildgruber et al.,³⁰ where functional MRI studies associated linguistic prosody with the left IFG, while emotional prosodic processing was represented in orbitofrontal areas bilaterally.

We found a correlation of acoustic markers with CSF p-Tau levels in the non-AD subset of PPA patients. We recently found a linear association of antemortem CSF p-Tau, but not t-Tau, with postmortem density of tau pathology in the brain of FTLD patients.¹⁴ Patients with confirmed FTLD-Tau pathology had higher CSF p-tau levels than their counterparts with confirmed FTLD-TDP pathology. Thus, the link we report here between acoustic speech markers and CSF p-Tau levels is consistent with the hypothesis connecting markers of dysprosody to the diagnosis of FTLD-Tau pathology,^{43–45} which is the most prominent pathology underlying naPPA.^{46,47} This finding remains to be confirmed in a larger autopsy sample or in future studies with in vivo PET tau molecular imaging that can detect FTLD-Tau pathology. Because this speech analysis is highly repeatable with minimal learning effects, it can potentially serve as a surrogate endpoint in treatment trials targeting tau pathology in patients with FTLD-Tau pathology.

Strengths and limitations

Strengths of our study include the objective and reliable measurement of speech intonation and rhythm without the use of subjective ratings. The use of the SAD enabled automatic analyses directly from digitized audio recordings, freeing us from the time-consuming and laborious work of transcription. The SAD is also independent of a specific language and thus can theoretically be applied cross-linguistically without preprogramming. The use of a natural speech sample is an advantage when considering its possible implications in clinical research requiring repeated and

frequent evaluations with minimal learning effects. Nevertheless, there are some shortcomings in our study. We were able to examine only a small group of PPA patients. Pitch-tracking involves complicated computational algorithms that estimate f0⁴⁸ and are subject to errors caused by background noise, unfavorable voice quality, octave jumps in pitch and overlapping speech. We applied multiple quality control measures to minimize pitch-tracking inaccuracies and confounds both at the tracking level and in our statistical analyses. Due to the nature of the SAD we can only relate these acoustic data to the prosodic “breath-group”. For a more detailed analysis at the sentence, word, syllable, or phoneme level, a complete alignment of the sound to its transcript is needed. While we used autopsy-verified levels of CSF analytes to characterize participants in this study, we did not have autopsy evidence of specific pathology in all participants.

With these caveats in mind, this work reports implementation of the SAD as a novel automated speech analysis tool in the study of PPA. We identified characteristics of speech that distinguish PPA phenotypes, linking these to other language characteristics of naPPA, left frontal cortical atrophy and CSF levels of p-Tau. Speech analyses in naPPA thus may be an informative marker to screen for FTLD-Tau pathology. These findings support the potential use of the SAD in the study of the cognitive processes underlying speech and for the measuring of a naturalistic, repeatable endpoint in clinical treatment trials.

Acknowledgments

This work was supported in part by the National Institutes of Health (AG017586, AG038490, AG053940, AG052943, NS088341, DC013063, AG054519), the Institute on Aging, the Alzheimer’s Association (AACSF-18-567131), and the Wyncote Foundation. We thank our patients and their caregivers and family members for their continuous effort and contribution to our clinical research.

Author Contributions

Study concept and design: N.N, M.G; data acquisition, analysis and interpretation: all authors; drafting of the manuscript: N.N; critical review of manuscript: all authors.

Conflicts of Interest

Dr. Nevler reports grants from Institute on Aging (IOA), the National Institutes of Health (NIH) and the Alzheimer’s Association; Dr. Ash reports grants from National Institutes of Health the Wyncote Foundation; Dr. Irwin

reports grants from the National Institutes of Health (NIH); Dr. Liberman has nothing to disclose; Dr. Grossman reports grants from the National Institutes of Health, the Wyncote Foundation and Biogen, nonfinancial support from Avid Radiopharmaceuticals and Piramal, personal fees from UCB not related to the submitted work.

References

- Ash S, Evans E, O'Shea J, et al. Differentiating primary progressive aphasia in a brief sample of connected speech. *Neurology* 2013;81:329–336.
- Gunawardena D, Ash S, McMillan C, et al. Why are patients with progressive nonfluent aphasia nonfluent? *Neurology* 2010;75:588–594.
- Liberman P. Intonation, perception, and language. Cambridge, MA: M.I.T Press, 1968.
- Ethofer T, Anders S, Erb M, et al. Cerebral pathways in processing of affective prosody: a dynamic causal modeling study. *NeuroImage* 2006;30:580–587.
- Leitman DI, Wolf DH, Ragland JD, et al. “It’s not what you say, but how you say it”: a reciprocal temporo-frontal network for affective prosody. *Front Hum Neurosci* 2010;4:1–13.
- Pell MD. Fundamental frequency encoding of linguistic and emotional prosody by right hemisphere-damaged speakers. *Brain Lang* 1999;69:161–192.
- Pichon S, Kell CA. Affective and sensorimotor components of emotional prosody generation. *J Neurosci* 2013;33:1640–1650.
- Ross ED, Monnot M. Neurology of affective prosody and its functional-anatomic organization in right hemisphere. *Brain Lang* 2008;104:51–74.
- Ryant N. LDC HMM Speech Activity Detector (v.1.0.4). University of Pennsylvania, 2013.
- Gorno-Tempini ML, Hillis AE, Weintraub S, et al. Classification of primary progressive aphasia and its variants. *Neurology* 2011;76:1006–1014.
- Giannini LAA, Irwin DJ, McMillan CT, et al. Clinical marker for Alzheimer disease pathology in logopenic primary progressive aphasia. *Neurology* 2017;88:2276–2284.
- Goodglass H, Kaplan E, Weintraub S. Boston diagnostic aphasia examination. Philadelphia, PA: Lea & Febiger, 1983.
- Nevler N, Ash S, Jester C, et al. Automatic measurement of prosody in behavioral variant FTD. *Neurology* 2017;2017:650–656.
- Irwin DJ, Lleo A, Xie SX, et al. Ante mortem cerebrospinal fluid tau levels correlate with postmortem tau pathology in frontotemporal lobar degeneration. *Ann Neurol* 2017;82:247–258.
- Lleó A, Irwin D, Illán-Gala I, et al. A CSF algorithm for the selection of FTLD subtypes in an autopsy cohort. *JAMA Neurol* 2018;75:738–745.
- Team R. RStudio: integrated development for R. 0.99.879 ed. Boston, MA: RStudio, Inc., 2015.
- Marchetti GM, Drton M, Sadeghi K. ggm: Functions for graphical Markov models. 2.3 ed 2015. p. R package.
- Revelle W. psych: procedures for personality and psychological research. 1.7.5 ed. Evanston, IL: Northwestern University, 2017.
- Wickham H. Reshaping data with the reshape package. *J Stat Softw* 2007;21:1–20.
- Xavier R, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;12:77.
- Yoshida K, Bohn J. Tableone: Create “Table 1” to Describe Baseline Characteristics, 2015.
- Morales M. sciplot: Scientific Graphing Functions for Factorial Designs. 1.1-1 ed, 2017.
- Fox J, Weisberg S. An R companion to applied regression. Thousand Oaks, CA: Sage, 2011.
- Wickham H. The split-apply-combine strategy for data analysis. *J Stat Softw* 2011;40:1–29.
- Venables WN, Ripley BD. Modern applied statistics with S, 4th ed. New York: Springer, 2002.
- Jenkinson M, Beckmann CF, Behrens TE, et al. FSL. *NeuroImage* 2012;62:782–790.
- Winkler AM, Ridgway GR, Webster MA, et al. Permutation inference for the general linear model. *NeuroImage* 2014;92:381–397.
- Aziz-Zadeh L, Sheng T, Gheyntchi A. Common premotor regions for the perception and production of prosody and correlations with empathy and prosodic ability. *PLoS ONE* 2010;5:e8759.
- Pell MD. The temporal organization of affective and non-affective speech in patients with right-hemisphere infarcts. *Cortex* 1999;35:455–477.
- Wildgruber D, Ackermann H, Kreifelts B, Ethofer T. Cerebral processing of linguistic and emotional prosody: fMRI studies. *Prog Brain Res* 2006;156:249–268.
- Rohrer JD, Sauter D, Scott S, et al. Receptive prosody in nonfluent primary progressive aphasia. *Cortex* 2012;48:308–316.
- Charles D, Olm C, Powers J, et al. Grammatical comprehension deficits in non-fluent/agrammatic primary progressive aphasia. *J Neurol Neurosurg Psychiatry* 2014;85:249–256.
- Mack JE, Chandler SD, Meltzer-Asscher A, et al. What do pauses in narrative production reveal about the nature of word retrieval deficits in PPA? *Neuropsychologia* 2015;77:211–222.
- Pakhomov SV, Smith GE, Chacon D, et al. Computerized analysis of speech and language to identify psycholinguistic correlates of frontotemporal lobar degeneration. *Cogn Behav Neurol* 2010;23:165–177.
- Duffy JR, Hanley H, Utianski R, et al. Temporal acoustic measures distinguish primary progressive apraxia of speech

- from primary progressive aphasia. *Brain Lang* 2017;07:84–94.
36. Ladefoged P, Disner SF. *Vowels and consonants*. West Sussex, UK: Wiley - Blackwell, 2012.
 37. Hodges JR, Patterson K. Nonfluent progressive aphasia and semantic dementia: a comparative neuropsychological study. *J Int Neuropsychol Soc* 1996;2:511–524.
 38. Mesulam MM, Wieneke C, Thompson C, et al. Quantitative classification of primary progressive aphasia at early and mild impairment stages. *Brain* 2012;135(Pt 5):1537–1553.
 39. Cousins KA, York C, Bauer L, Grossman M. Cognitive and anatomic double dissociation in the representation of concrete and abstract words in semantic variant and behavioral variant frontotemporal degeneration. *Neuropsychologia* 2016;84:244–251.
 40. Bonner MF, Price AR, Peelle JE, Grossman M. Semantics of the visual environment encoded in parahippocampal cortex. *J Cogn Neurosci* 2016;28:361–378.
 41. Sajjadi SA, Patterson K, Arnold RJ, et al. Primary progressive aphasia: a tale of two syndromes and the rest. *Neurology* 2012;78:1670–1677.
 42. Mesulam MM, Weintraub S, Rogalski EJ, et al. Asymmetry and heterogeneity of Alzheimer's and frontotemporal pathology in primary progressive aphasia. *Brain* 2014;137 (Pt 4):1176–1192.
 43. Kuiperij HB, Versleijen AA, Beenes M, et al. Tau rather than TDP-43 proteins are potential cerebrospinal fluid biomarkers for frontotemporal lobar degeneration subtypes: a pilot study. *J Alzheimers Dis* 2017;55:585–595.
 44. Irwin DJ, Trojanowski JQ, Grossman M. Cerebrospinal fluid biomarkers for differentiation of frontotemporal lobar degeneration from Alzheimer's disease. *Front Aging Neurosci* 2013;5:1–11.
 45. Irwin DJ, Lleó A, McMillan CT, et al. Ante mortem CSF tau levels correlate with post mortem tau pathology in FTL. *Ann Neurol* 2017;82:247–258.
 46. Grossman M. The non-fluent/agrammatic variant of primary progressive aphasia. *Lancet Neurol* 2012;11:545–555.
 47. Josephs KA, Hodges JR, Snowden JS, et al. Neuropathological background of phenotypical variability in frontotemporal dementia. *Acta Neuropathol* 2011;122:137–153.
 48. Boersma P. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences*. 1993;17(University of Amsterdam):97–110.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. Gray matter (GM) density analysis.

Table S1. GM atrophy and regression results for f0 range in nPPA.