TOWARDS DEVELOPMENT OF AN AUTOMATIC ANALYSIS FOR ASSESSMENT OF DYSPHONIC SPEECH

KEIKO ISHIKAWA, MM, MA, CCC-SLP

MARISHA SPEIGHTS, MA, CCC-SLP,

JOEL MACAUSLAN, PHD,

SUZANNE BOYCE, PHD, CCC-SLP

SpeechMark®

UNIVERSITY OF Cincinnati

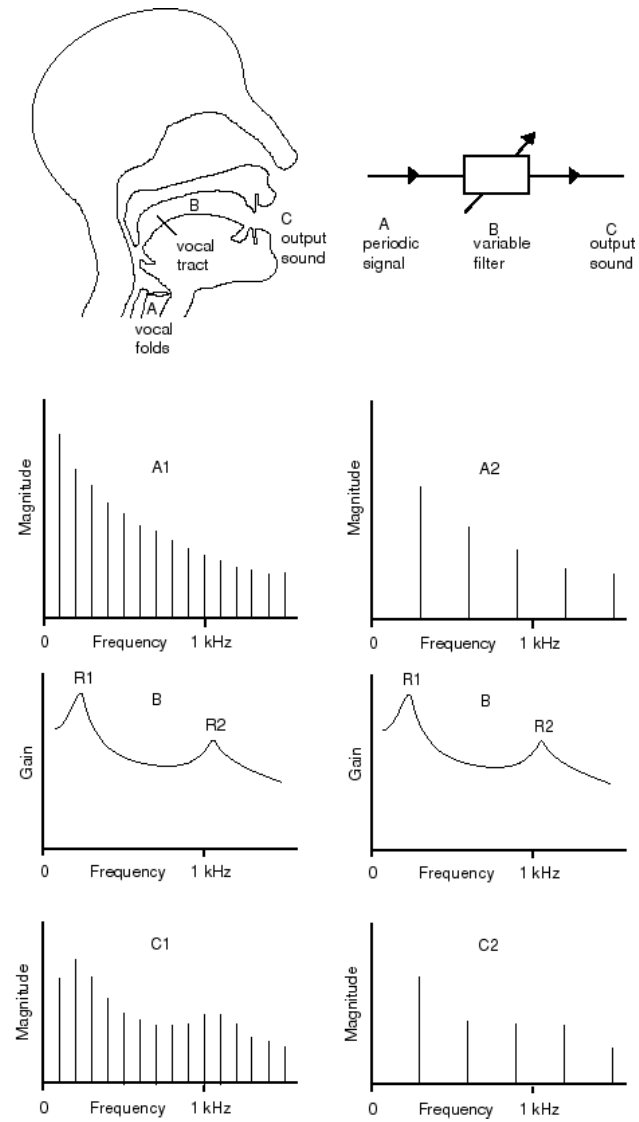
DYSPHONIA & INTELLIGIBILITY

- **Lack of intelligibility in noisy environment is the primary complaint of many patients with dysphonia.** (Jacobson et al., 1997)
- **What aspects of dysphonic speech causes the reduction in intelligibility?**

WHAT IS SPEECH?

Source-Filter model

- Sound energy source (e.g. the vocal fold vibration)
- Filtered according to a resonance characteristics determined by the shape of the supralaryngeal vocal tract.



From:
http://www.animations.physics.unsw.edu.au/jw/speech.html

WHAT DO WE NEED IN SIGNAL FOR DECODING SPEECH?

How do we find “information” in a speech signal?

- The human perception system attends and responds to a change in a signal.
- The auditory system needs a change in signal that is strong enough for it to respond.
- The decision on whether the change in signal contains linguistically meaningful information (i.e. phonetic feature) is made by the subsequent systems.

LARYNGEAL PATHOLOGY & SOURCE PRODUCTION

Change in vibratory characteristics of the vocal folds and glottic closure

- Increase in noise
- Decrease in harmonic power

Information important for speech perception is obscured

ACOUSTIC TOOLS FOR DYSPHONIA EVALUATION

- **Acoustic studies of voice have focused on determining acoustic correlates of :**
  - perceptual judgment of voice quality (Kreiman et al., 1993; Yumoto, Sasaki, & Okamura, 1984; Wolfe & Martin, 1997, etc.)
  - vibratory nature of the vocal folds (Mehta, Zañartu, Quatieri, Deliyski, & Hillman, 2011)

THE CLINICAL NEEDS

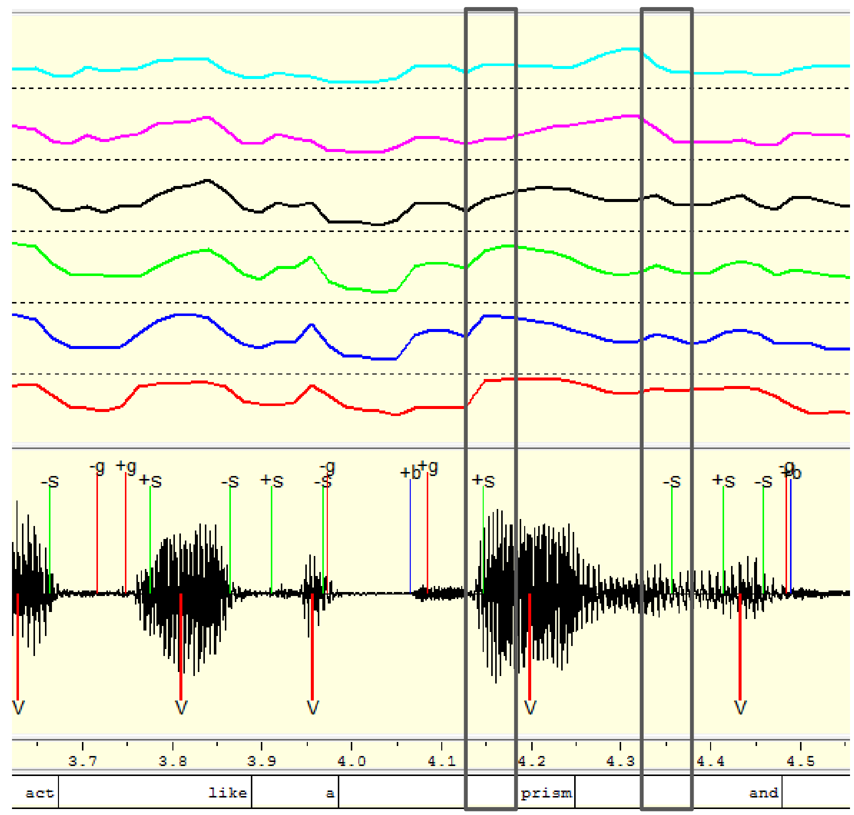
- Current algorithms are not designed to study aspects of signal that are relevant to intelligibility.
- Such speech analysis can be highly laborious and time-consuming – implementation of automatic tool would be helpful.

LANDMARK ANALYSIS PROVIDES NEW INSIGHTS

- The acoustic region with the abrupt change contains information that is particularly salient to listeners for making a decision about speech sound. → LANDMARKS

  (Stevens, 2000, 2002)

- Designed for automatic speech analysis



Landmark analysis by SpeechMark®

LANDMARK ANALYSIS PROVIDES NEW INSIGHTS

Landmarks

- Follows the tradition of distinctive features (Jakobson, 1928; Chomsky & Halle, 1968)
- SpeechMark®: 6 consonantal and 1 vowel landmarks
    - “glottis” ([+g] and [-g])
    - “burst” ([+b] and [-b])
    - “syllabicity” ([+s] and [-s])
    - “voiced frication” ([+v] and [-v])
    - “frication” ([+f] and [-f])
    - vowel landmark ([V]).

[+]: onset
[-]: offset

LANDMARK ANALYSIS PROVIDES NEW INSIGHTS

Past LM studies:

- **“Clear” vs. “casual” speech: Greater number of LMs found in clear speech.** (Boyce et al., 2013)
- **“Dysarthric” vs. “normal” speech: Deletion of expected LMs and insertion of unexpected LMs found in dysarthric speech.** (Dicicco & Patel, 2008)
- **“Parkinson’s” vs. “normal” speech: Parkinson group had a reduced LM cluster rate** (Boyce, Fell, Wilde & MacAuslan, 2011)

CAN LM ANALYSIS DIFFERENTIATE DYSPHONIC SPEECH FROM NORMAL SPEECH?

METHODS

Materials:

- 33 dysphonic & 36 normal speakers
- Moderate to severe dysphonia
- The 1st sentence of the Rainbow passage (Kay Disordered Voice Database)

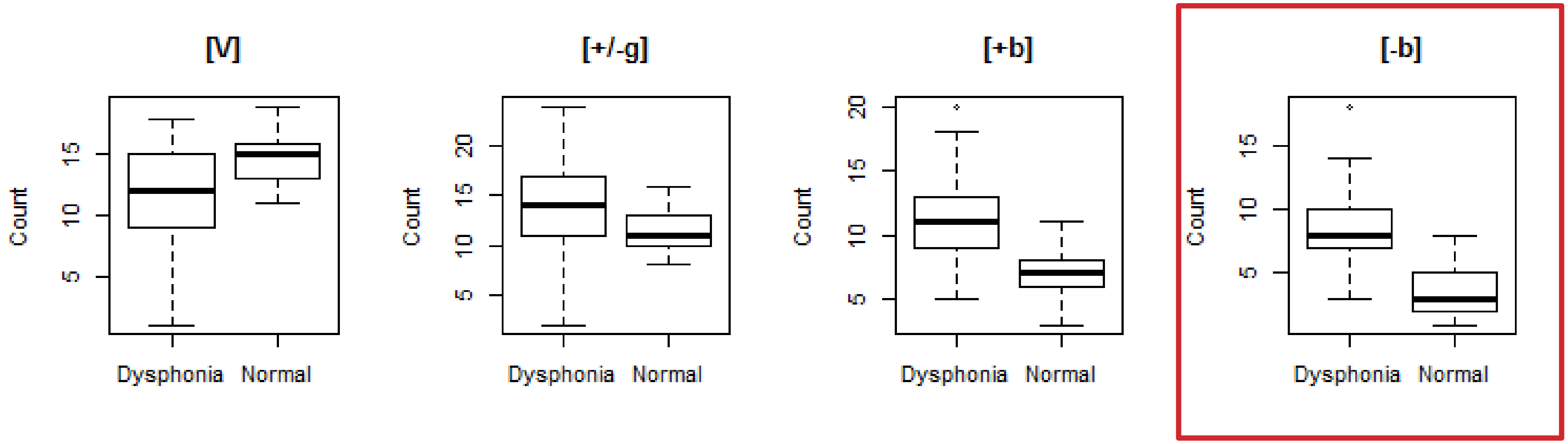
Measure:

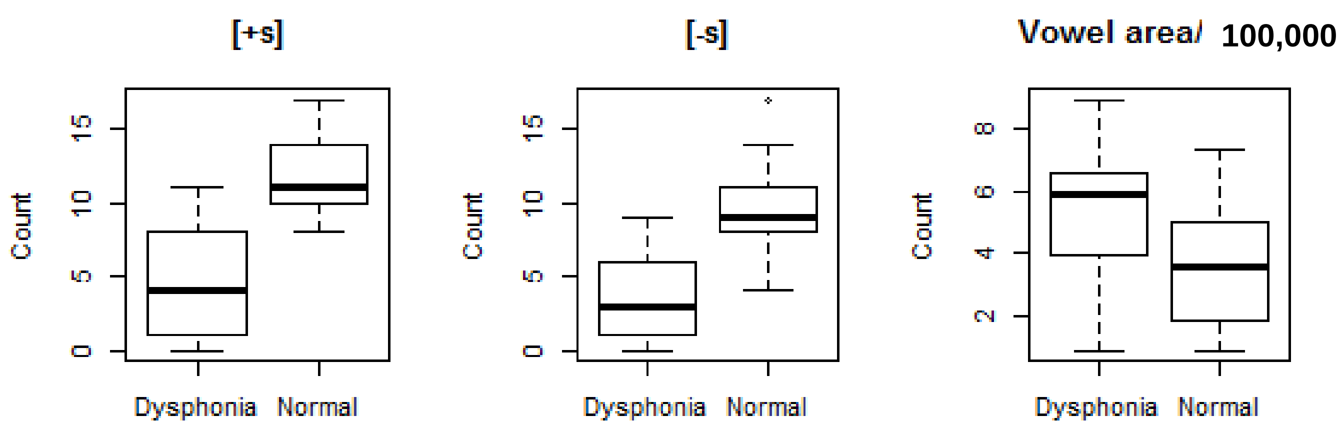
- Count of LMs

Statistical methods:

- **Logistic regression**
- **Classification tree**
- **Response variables:**
  - Vocal status (normal vs. disordered)
- **Predictor variables:**
  - Count of each LM, vowel area

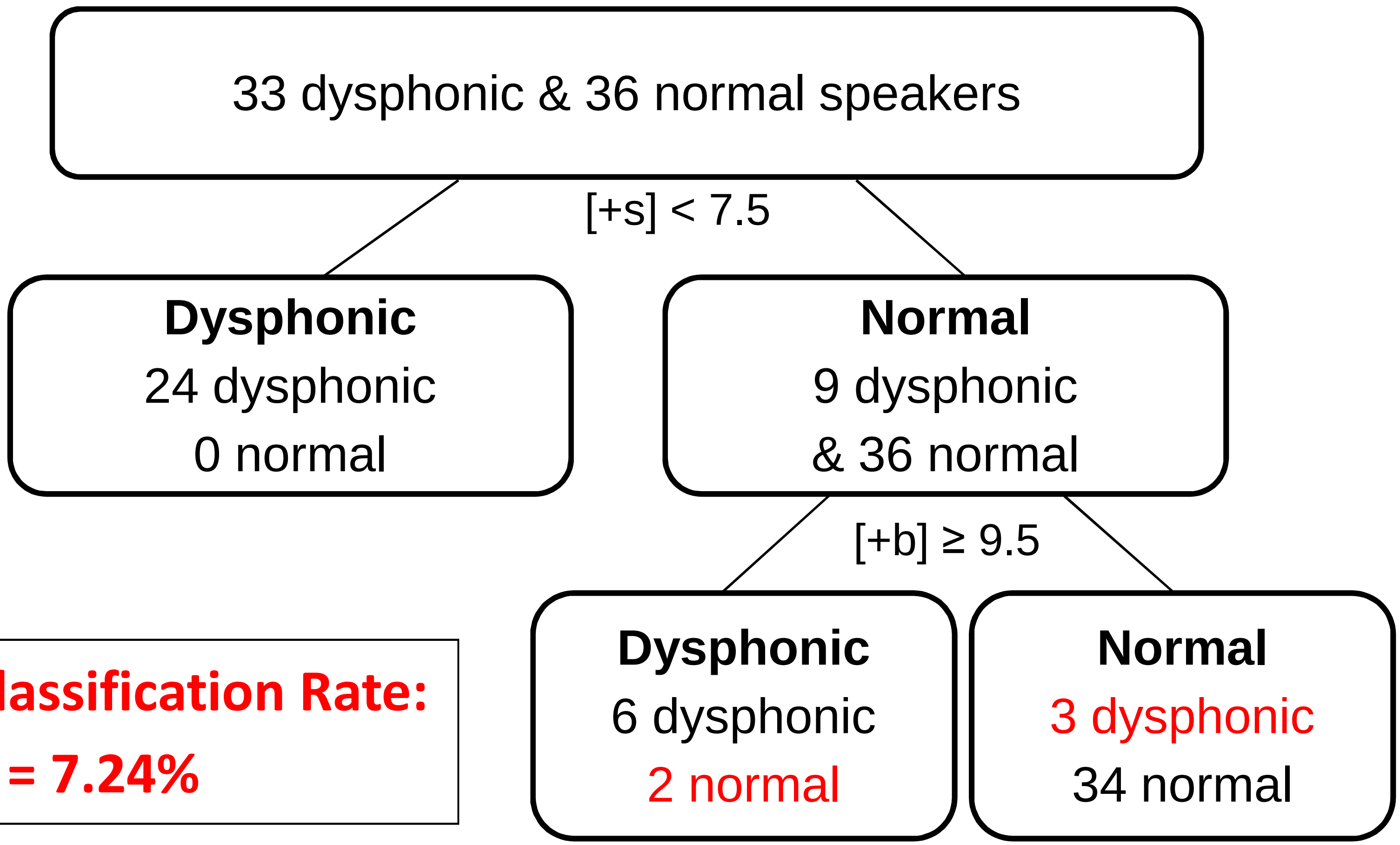
RESULTS





Logistic regression model [-b] is a significant predictor for dysphonia ($p = 0.045$)

RESULTS



Misclassification Rate:
5/69 = 7.24%

CAN LM ANALYSIS DETECT UNDERLYING PATHOLOGIES?

Difference in underlying physiology may be reflected in the speech signal.

Participants

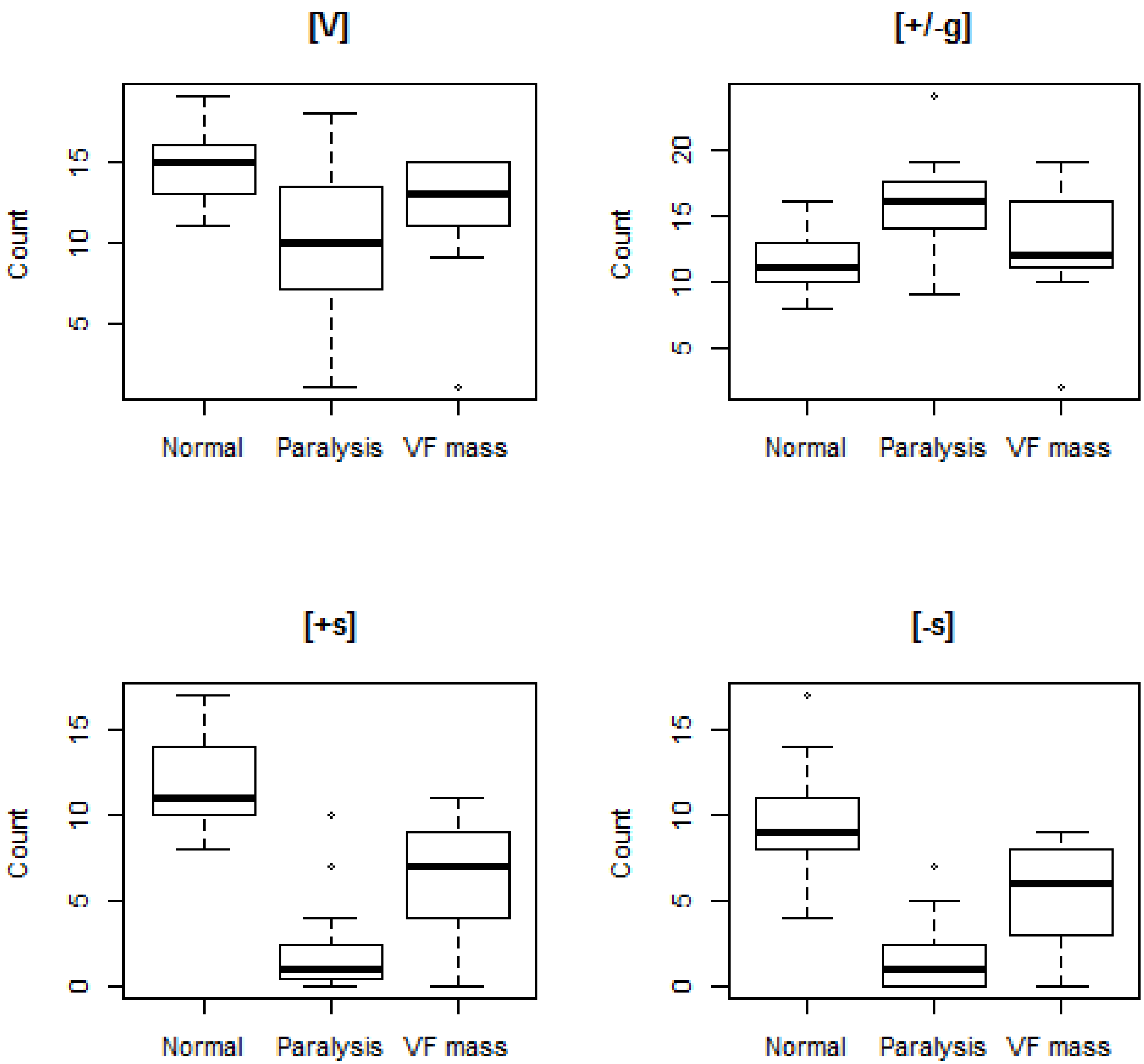
- 36 normal speakers (15 females, 21 males)
- 33 dysphonic speakers
  - VF paralysis (8 females, 8 males)
  - VF mass: VF polyp (5 females, 4 males); VF nodules (8 female)

Response variables

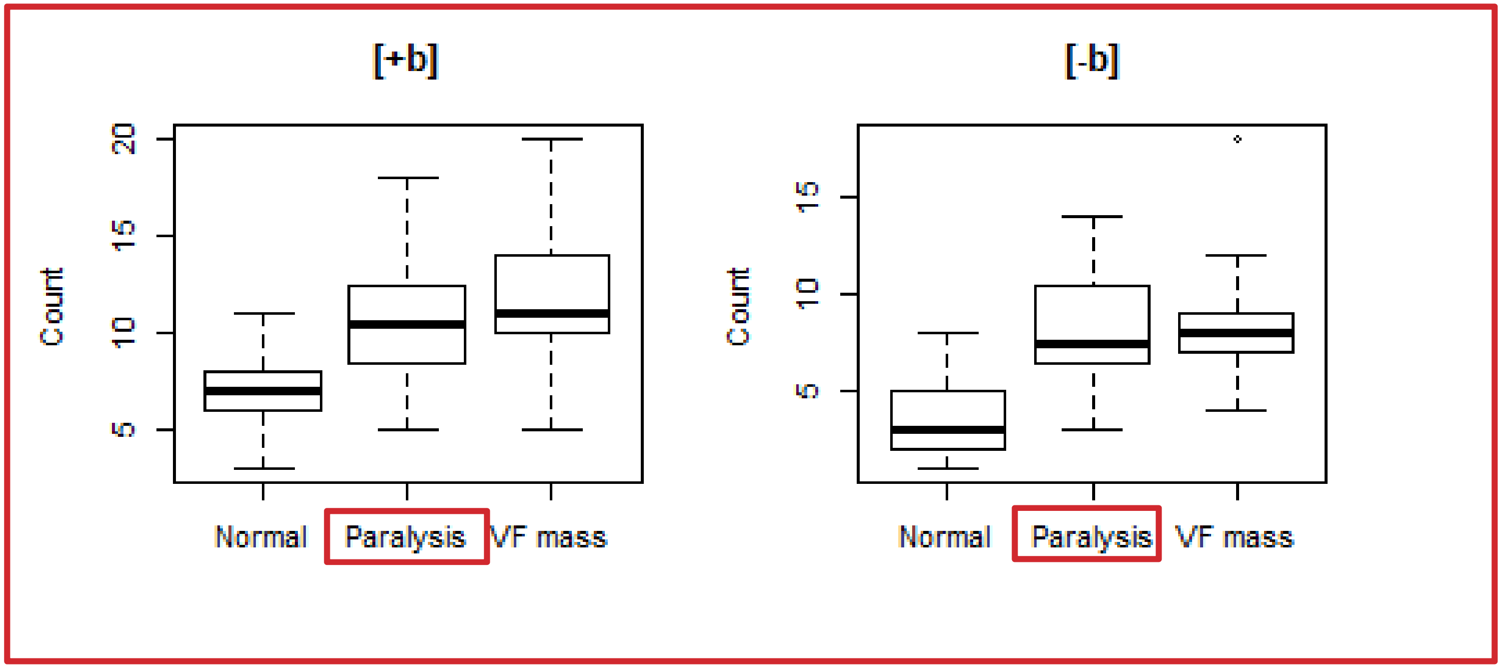
- Laryngeal diagnosis (normal vs. VF paralysis vs. VF mass)

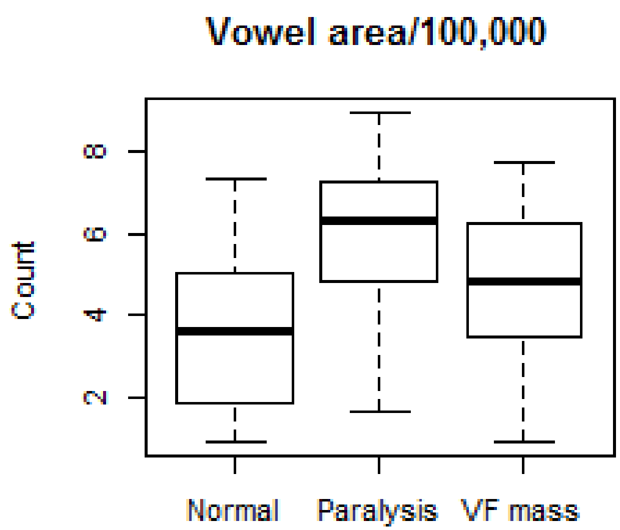
Predictors/Independent variables

- Count of each LM, vowel area



[V]
[+/-g]
[+s]
[-s]
Count
5
10
15
20
0
Normal
Paralysis
VF mass

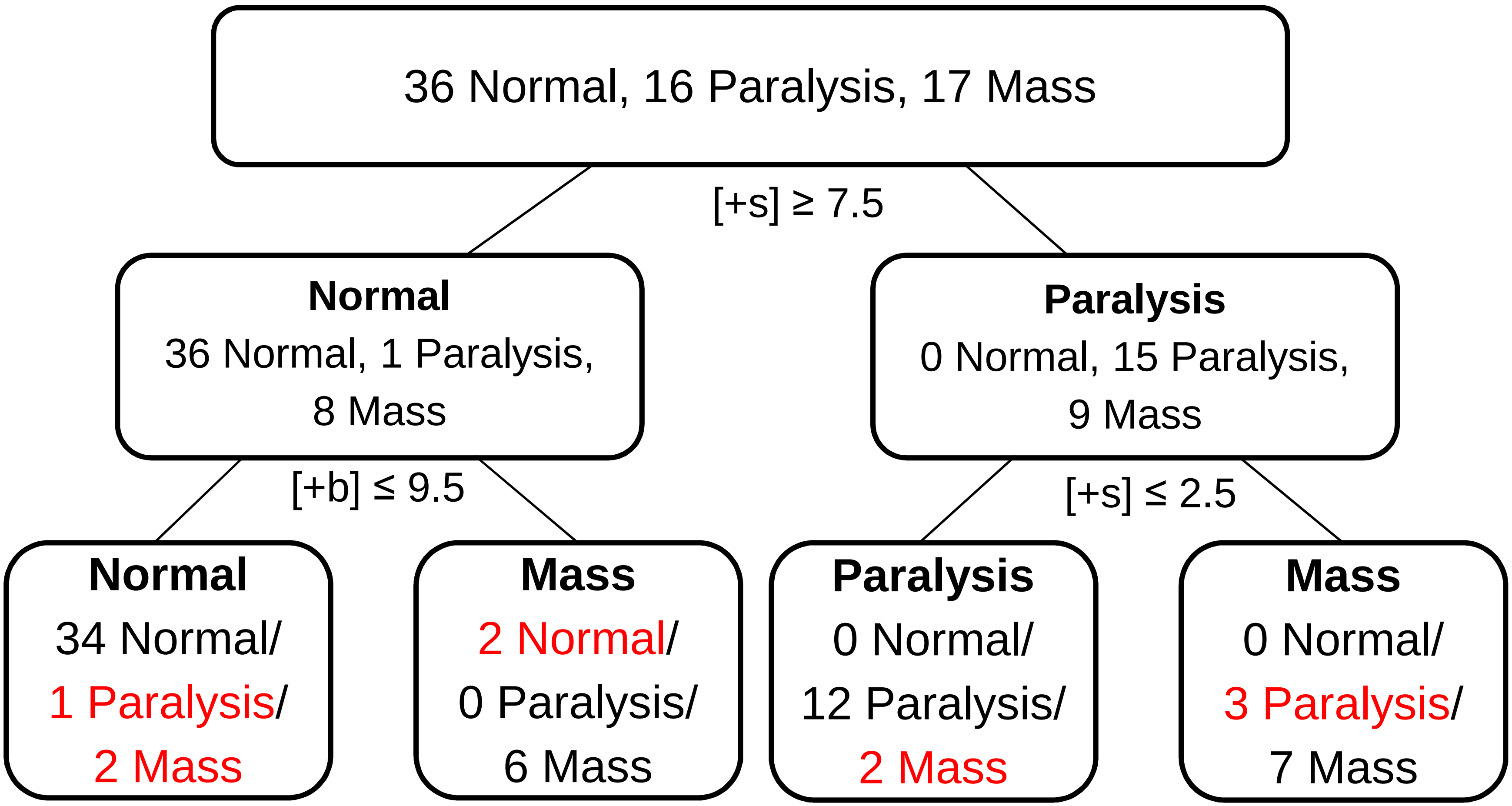




Multinomial regression model

- [+b] and [-b] are significant predictors for VF paralysis ($p < 0.05$)
- None of the predictors were significant for VF mass

RESULTS



Misclassification Rate: 10/69 = 14.49%

CONCLUSIONS

- **Based on LM count from 1 sentence, LM analysis is a useful method to detect acoustical differences between normal and dysphonic speech.**
- **LM analysis is less effective for detecting difference between underlying pathologies → Not surprising**
- **Burst and syllabicity LMs are significant predictors of dysphonic speech**
  - Greater number of [+/-b] : more frequent onset/offset in VF vibration?
  - Reduced number of [+/-s]: inability to create abrupt change in voiced regions (lack of sonorancy)

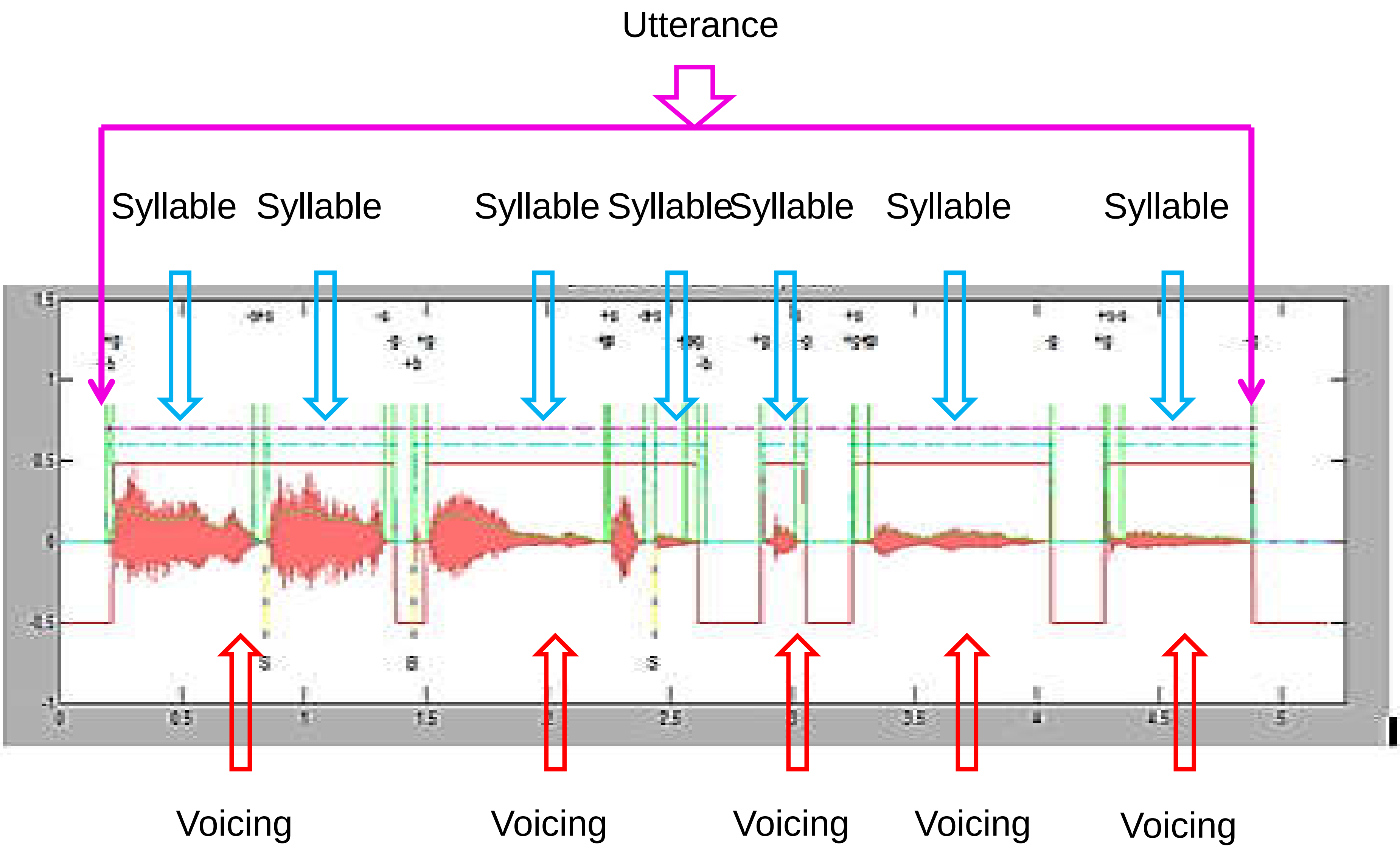
SYLLABIC CLUSTER ANALYSIS

- **Detection of syllabic structure is important for speech perception.** (Healy, 1976; Mehler et al., 1981)
- **Is a signal of dysphonic speech degraded enough to disturb detection of syllable structure by listeners?**
- **SpeechMark® summarizes LMs based on sequences of 6 LM types ([+/-g], [+/-b], [+/-s])**
- **Syllabic cluster – Group of LMs defined by a few acoustic rules that are based on physiological nature of human speech production (i.e. phonetically possible patterns in English).**

SYLLABIC CLUSTER ANALYSIS

- **“Syllable” in LM analysis**
  - Acoustic definition of what syllable looks like
  - What was uttered ≠ what was supposed to be uttered

    Example: “interesting”
    - /ɪntərɛstɪŋ/ – 4 syllables in canonical form
    - /ɪnrɛstɪŋ/ – 3 syllables, reduced complexity



Utterance
Syllable
Syllable
Syllable
Syllable
Syllable
Syllable
Syllable
Voicing
Voicing
Voicing
Voicing
Voicing

METHODS

Response variables:

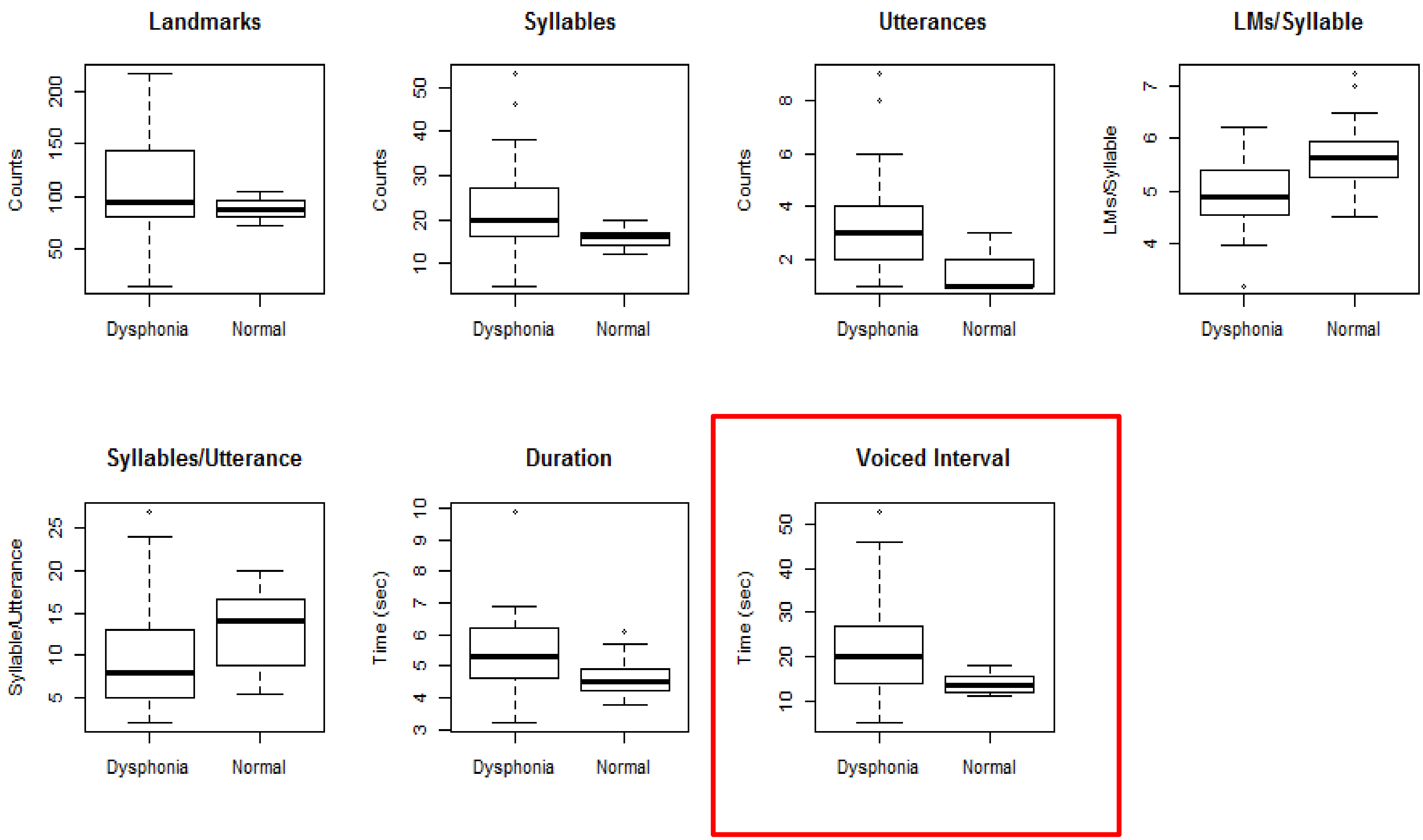
- **Vocal status (Normal vs. Dysphonia)**
- **Diagnoses (Normal vs. VF mass vs. VF paralysis)**

Predictors:

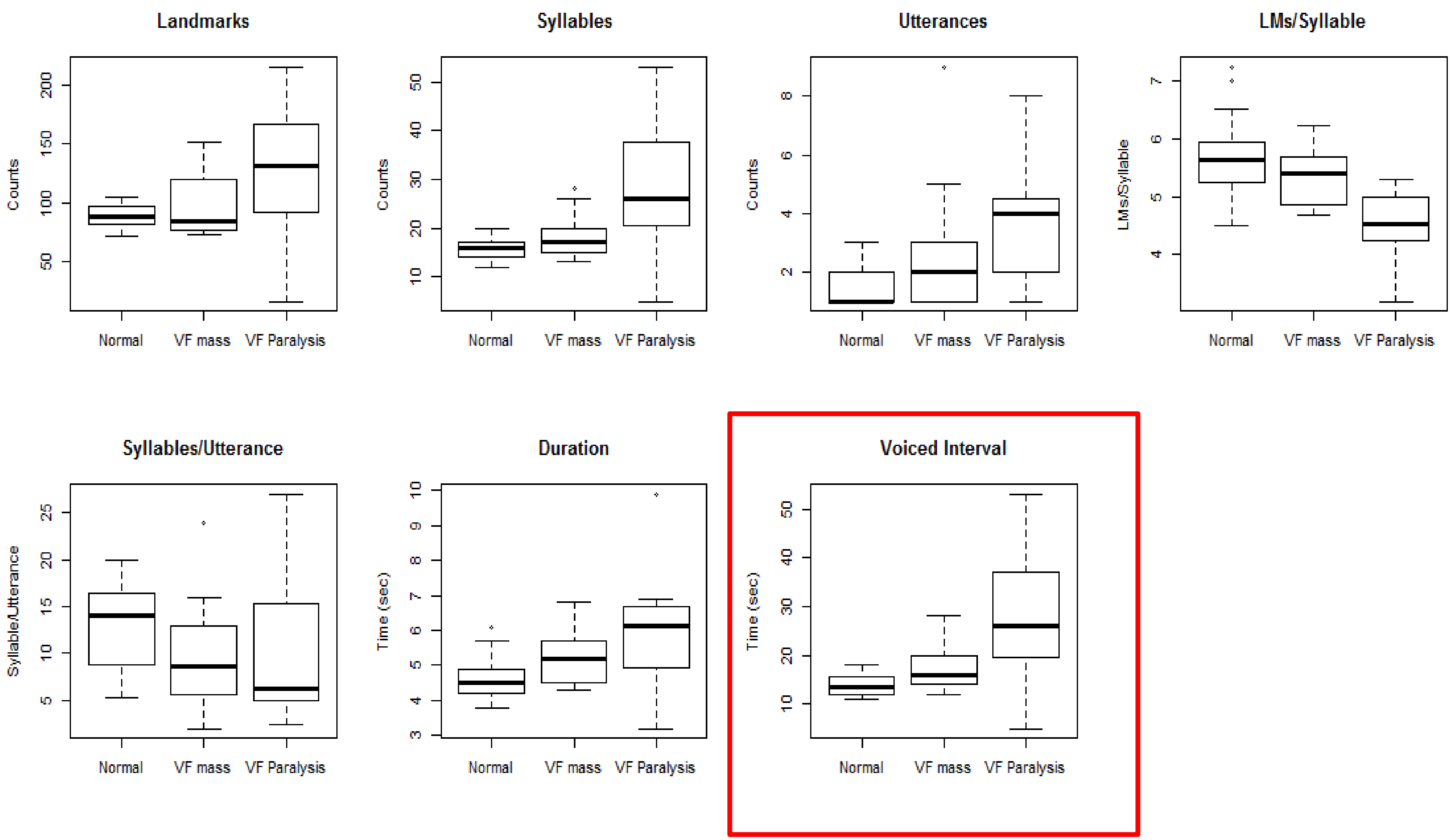
- **Count of LMs**
- **# of “syllables” and “utterances”**
- **# of LMs/syllable, # of syllables/utterance**
- **duration of utterances**
- **duration of voiced segments**

Statistical models: Logistic regression, multinomial regression

RESULTS: NORMAL VS DYSPHONIA



RESULTS: NORMAL VS VF MASS VS VF PARALYSIS



CONCLUSIONS

- Voiced interval was greater in dysphonic speech – more frequent instance of unexpected vocal fold onset and offset occurred in unexpected moments
- Other parameters for syllabic cluster measure were not significant predictors for dysphonic speech. The analysis may have been affected by the wide variability seen in dysphonic group (especially VF paralysis group).

LIMITATIONS

- Variabilities among dysphonic speakers were large → larger sample size may be helpful.
- The samples used for these studies are phonetically limited. May need more comprehensive speech samples for generalizable outcomes.
- A study with more linguistically comprehensive sample and larger number of speakers is underway.

ACKNOWLEDGEMENTS

The United States National Institutes of Health (NIDCD R43 and R44 DC010104, R42 AG033523, R21 HL086689, and R42 HD34686 to Joel MacAuslan, Harriet Fell and Suzanne Boyce)

The SpeechMark™ system was developed under a number of previous NIH grants to the Speech Technology and Applied Research (S.T.A.R.) Corporation

The American Speech-Language Hearing Association Kala-Singh Memorial Fund Graduate Student Scholarship

The Council of Academic Programs in Communication Sciences and Disorders PhD Scholarship

National Institutes of Health

THANK YOU!