This study was done in collaboration with the University of Virginia and before I start I want to acknowledge the help of the people who are mentioned above.

First let’s define what we are studying. We are focusing on dysfluency types. A lot of jargon is used in stuttering literature.
This list of dysfluencies comes from a review by Einarsdottir and Ingham. Clearly, there is a plethora of labels, which does not help students, nor researchers for that matter. Some terms seem synonymous, but they may convey a different nuance each (tense pause & blocks, broken words & part-word repetition). Some refer to a unit of meaning or a unit of syntax, suggesting that semantics or syntactics are implicated in the disorder and confounding cross-linguistic evaluation (incomplete phrases, linguistic nonfluency, part-word repetition, broken words), in contrast to labels referring to syllables, which are motor speech units.
Moreover, Ray Kent points to the drawbacks of auditory-perceptual evaluations.

Dysfluency taxonomies

- valid
  i.e. genuine correlate of severity
- reliable + interjudge agreement
  i.e. unequivocally defined and observable

The basic requirements for any testing tool, whatever it measures, are validity and reliability. A test is valid when it actually measures what it promises to measure. In the case of a taxonomy of stuttering dysfluencies this means...
Dysfluencies according to Packman & Onslow

Repeated movements
- Syllable repetition
- Incomplete syllable repetition
- Multisyllable repetition

with audible airflow
- without audible airflow

Fixed postures
- hear it (vowel or fricative)-sound that you can prolongate
- don't hear it- block (stutter suffering to continue)

3 variations of Repeated Measures:
- **Syllable repetition** (Rs): 1 sound and more than one sound, vowel and consonant or one sound that is a word on its own (ex: I and an)
- **Incomplete syllable repetition** (Ri): 1 sound recurring (s.s.s.s...sick ex: fricatives)
- **Multisyllable repetition** (Rm): more than one sound recurring (antwe.antwe.antwe..ex: antwerp)
Dysfluencies according to Packman & Onslow

Superfluous behaviors:
- **nonverbal**- movement...squirm...CAN'T ANNOTATE! (only use audible things) - note this in presentation, smacking, clearing throat
- **verbal**- "uh," revisions-start and stop sentence and start over with new syntax, um-verbal interjection
using the annotation tool in Praat software, it is possible to implement the taxonomy of your choice and to extract the number of instances as well as the duration of dysfluencies. The goal of this study was to check the predictive value of tallied and timed parameters when it comes to evaluating stuttering severity.

Questions & Hypotheses

- Is there a **relationship** (strong enough to make reliable predictions) between perceived severity and tallied/timed behaviors?

- What type of **predictors have more predictive power** (timing vs. tallying behaviors)

- What is the **handiest equation to predict** perceived stuttering severity, starting from a one minute recording

- Relation is condition to establish regressions & prediction
- Behaviors with strong influence on listeners' rating are relevant for diagnosis (severity, type or profile), for therapy (priorities, monitoring progress)
Speech samples

- 36 speech samples (University College London’s Archive of Stuttered Speech, UCLASS) (funded by The Wellcome Trust) [http://www.uclass.psychol.ucl.ac.uk/Audio/mp3/]
- Speech samples were in English; participants were all English-speakers
- Each sample was played for approximately one minute
- Samples were played in different sequence

Participant Details

- Total Participants: 120 (3 subgroups of 40)
- Age (range): 14-62 yrs
- Selected by convenience sampling
- 36 speech samples: (3 subsets of 12); each subset was rated by a subgroup of 40 participants
- Each subgroup of 40 heard the same 12 samples, in different sequence
Participant Instructions

- Listen to speech samples collected from University College London site
- Rate 12 speech samples on a Likert Scale (1-5) for severity
  - 1 - mild
  - 3 - moderate
  - 5 - severe
- To calibrate responses, participants were asked to listen to a sample of a "moderate" (3) stutter prior to beginning the rating session.

Mean Likert scale ratings

- Almost the complete continuum of severity was represented in the database.
Annotations: mnemonics

Annotating was done in terms of short mnemonics

Praat software: annotations

“1” : our code for fluent stretch
Superfluous behavior, if nonverbal and inaudible, does not appear on a Praat editor screen. If we want to time it exactly, we could add a separate TextGrid for inaudible nonverbal superfluous behavior.

We used a custom made script to extract and process annotated details.
Dia 25

Praat script outcome

Dia 28

Is there a relationship....

Parameters pertaining to fluent fragments
Is there a relationship....

Parameters pertaining to fixed postures

<table>
<thead>
<tr>
<th>Parameters</th>
<th>MEAN/RATING</th>
<th>Pearson Correlation (Sig. 2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of standing/walking (Fixed)</td>
<td>M</td>
<td>0.11</td>
</tr>
<tr>
<td>Number of standing/walking (Fixed)</td>
<td>N</td>
<td>0.11</td>
</tr>
<tr>
<td>Percentage time standing/walking (Fixed)</td>
<td>M</td>
<td>0.11</td>
</tr>
<tr>
<td>Percentage time standing/walking (Fixed)</td>
<td>N</td>
<td>0.11</td>
</tr>
<tr>
<td>Duration of standing/walking (Fixed)</td>
<td>M</td>
<td>0.11</td>
</tr>
<tr>
<td>Duration of standing/walking (Fixed)</td>
<td>N</td>
<td>0.11</td>
</tr>
</tbody>
</table>

*Correlation is significant at the 0.05 level (2-tailed).  **Correlation is significant at the 0.01 level (2-tailed).
Is there a relationship....

Parameters pertaining to repeated movements

Dia 32
Is there a relationship....

Note that basic dysfluency types (all subtypes combined, e.g. Repetitions = Rs+Ri+Rm) correlate very well
Is there a relationship....

Parameters pertaining to superfluous behaviors, cont'd

### Dia 35

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Pearson Correlation Sig. (2-tailed)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of interactions (superfluous behavior)</td>
<td>-0.166</td>
<td>37</td>
</tr>
<tr>
<td>Mean duration of interactions (superfluous behavior)</td>
<td>-0.254</td>
<td>37</td>
</tr>
<tr>
<td>Percentage of time of interactions (superfluous behavior)</td>
<td>0.167</td>
<td>37</td>
</tr>
<tr>
<td>Number of requests (superfluous behavior)</td>
<td>0.085</td>
<td>37</td>
</tr>
<tr>
<td>Mean duration of requests (superfluous behavior)</td>
<td>0.096</td>
<td>37</td>
</tr>
<tr>
<td>Number of lack of requests (superfluous behavior)</td>
<td>-0.095</td>
<td>37</td>
</tr>
<tr>
<td>Mean duration of lack of requests (superfluous behavior)</td>
<td>0.079</td>
<td>37</td>
</tr>
</tbody>
</table>

* Correlation is significant at the 0.05 level (2-tailed).

### Dia 36

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Pearson Correlation Sig. (2-tailed)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack of interactions</td>
<td>-0.006</td>
<td>37</td>
</tr>
<tr>
<td>Lack of mean duration</td>
<td>-0.002</td>
<td>37</td>
</tr>
<tr>
<td>Lack of percentage of time</td>
<td>0.007</td>
<td>37</td>
</tr>
<tr>
<td>Lack of number of requests</td>
<td>0.006</td>
<td>37</td>
</tr>
<tr>
<td>Lack of mean duration of requests</td>
<td>0.003</td>
<td>37</td>
</tr>
<tr>
<td>Lack of number of lack of requests</td>
<td>0.004</td>
<td>37</td>
</tr>
<tr>
<td>Mean duration of lack of requests</td>
<td>0.001</td>
<td>37</td>
</tr>
</tbody>
</table>

* Correlation is significant at the 0.05 level (2-tailed).
Is there a relationship....

<table>
<thead>
<tr>
<th>Significant correlation</th>
<th>Tallying (number of)</th>
<th>Timing (mean duration and/or % of time)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fluent fragments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixed postures (Fp, Fb)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repeated movements (Rs, Ri, Rm)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Superfluous behavior (Si, Sr, Sn)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Clearly more significant correlations between perceived severity and timed indices (compared to tallied indices)

Dia 38

Predictors with more weight ....

top five based on stepwise regression: tallied predictors: 0, timed predictors: 5
This equation is based on a high correlation, it has face validity and it only requires annotating fluent stretches. Of course, annotating all (sub)types of stuttering behaviors provides more predictors, which could be used to combine in a multiple regression (this is a single predictor regression). Moreover: annotating all stuttering behaviors from the taxonomy may yield a profile of each particular patient and reveal priorities for therapy.
Back to “Questions & Hypotheses”

• Is there a relationship (strong enough to make reliable predictions) between perceived severity and tallied/timed behaviors?

• Timing behaviors yields predictors with more weight than tallying behaviors.

• What is the handiest equation to predict perceived stuttering severity, starting from a one minute recording?