Modern data analysis and pathological speech

Stefan Werner

Linguistics & Language Technology
School of Humanities
University of Eastern Finland

BAULT – Final Symposium, Univ. of Helsinki
2016-12-01

UNIVERSITY OF
EASTERN FINLAND

Background

- Acoustic assessment of speech pathologies
- Automatic *black-box* classification vs. laborious manual analysis

The beginnings

- Rule-based evaluation of manual acoustic measurements
- Focus on one acoustic parameter at a time

Example: *Speech Examination* (Keller et al. 1991)

EXAMEN DE LA PAROLE

("Speech Examination", version française [French])
août 1990

PROTOCOLE A

CONSIGNES GÉNÉRALES:

1. Chaque stimulus n'est présenté qu'une seule fois (après avoir vérifié que le Sujet soit pret à l'entendre). Le sujet initie la répétition immédiatement après la présentation du stimulus. Le sujet ne doit ni lire ni voir les stimuli écrits.
2. Permettre au sujet au maximum **deux** tentatives de production du stimulus.
3. Inciter le sujet à prononcer les stimuli **aussi clairement que possible**.
4. Afin d'assurer la comparabilité des résultats entre différents laboratoires, il faut résolument éviter de modifier les instructions de ce protocole.
5. Les instructions pour le sujet sont fournies en italiques. Les instructions pour l'expérimentateur se trouvent entre crochets.

S.V.P. enregistrer sur la cassette:

1. Nom du sujet et initiales
2. Age du sujet
3. Niveau d'éducation (dernière année d'école complétée)
4. Statut du français (langue maternelle?)
5. Caractérisation clinique de la maladie
6. Temps du début de la maladie
7. Date et temps de départ du protocole

A-1: PAROLE SPONTANÉE

Parlez brièvement sur n'importe quel sujet [Suggérer un sujet non menaçant si nécessaire. Une minute.]

A-2: SONS PROLONGÉS (1)

Prenez un bon souffle et tenez le son pendant au moins 5 secondes. [Après 6 secondes, arrêter le sujet (lever la main). Si le sujet arrête avant 5 secondes, demander une deuxième production du son.]

A-2.1 /i/ comme dans l**i**t
A-2.2 /A/ comme dans ch**a**t
A-2.3 /o/ comme dans b**eau**
A-2.4 /s/ comme dans s**c**ène
A-2.5 /z/ comme dans **z**èle
A-2.6 /S/ comme dans **ch**er
A-2.7 /Z/ comme dans **j**our

A-3: SYLLABES REPETES

Répétez les syllabes suivantes à un débit normal pour une période d'au moins 5 secondes. Utilisez un rythme régulier ("comme un métronome"). [Montrer

A-5.6 chanson /S/
A-5.7 plastique /st/
A-5.8 chasseur /S/
A-5.9 dicter /kt/
A-5.10 chalet /S/

A-5.11 recteur /kt/
A-5.12 rester /st/
A-5.13 sachet /S/
A-5.14 station /st/
A-5.15 actif /kt/

A-6: SONS PROLONGÉS (2)

[Reprendre la section A-2 afin d'augmenter le nombre d'observations.]

A-7: RYTHME

Répéter les syllabes suivantes pendant cinq secondes à un rythme normal. [Montrer au sujet le patron d'accentuation à prendre.]

A-7.1 /pá pa pá pa pá pa pá pa.../
[accentuation alternante]
A-7.2 /pá pá pa pa pá pá pa pa.../
[2 accentuées, 2 non accentuées]

Example: *Speech Examination* (Keller et al. 1991)

WORKSHEET FOR HELD VOWELS - FUNDAMENTAL FREQUENCY					
	Vowels			Fricatives	
	/i/	/A/	/o/	/z/	/Z/
1. f0 uneven - in 1/3 of 5 sec, after 200 ms - in 2/3 of 5 sec, after 200 ms - in 3/3 of 5 sec, after 200 ms					
2. Bitonal voice (two f0s)					
3. Abrupt changes in f0 (voice break)					
4. f0 excessively high (strangled voice)					
5. f0 low and uneven (raspy voice)					
6. Harmonics affected					
7. Additional problems					
Errors for each sound:	Vowels:			Fricatives:	
Total analyzed:					
Total errors:					
Success% [1]:					
	Mean success% ($x_{V\%,F\%}$):				

Example: *Speech Examination* (Keller et al. 1991)

Ricardo Bion's semi-automatization of SE as a *Praat* plugin (2006):

Speech Examination Database Manager

This website was developed as a final project for the course course "Methods in Speech Corpora," and "Acoustic Analysis of Speech," taught in the Spring term of 2006 at the University of Joensuu (FI).

The Project

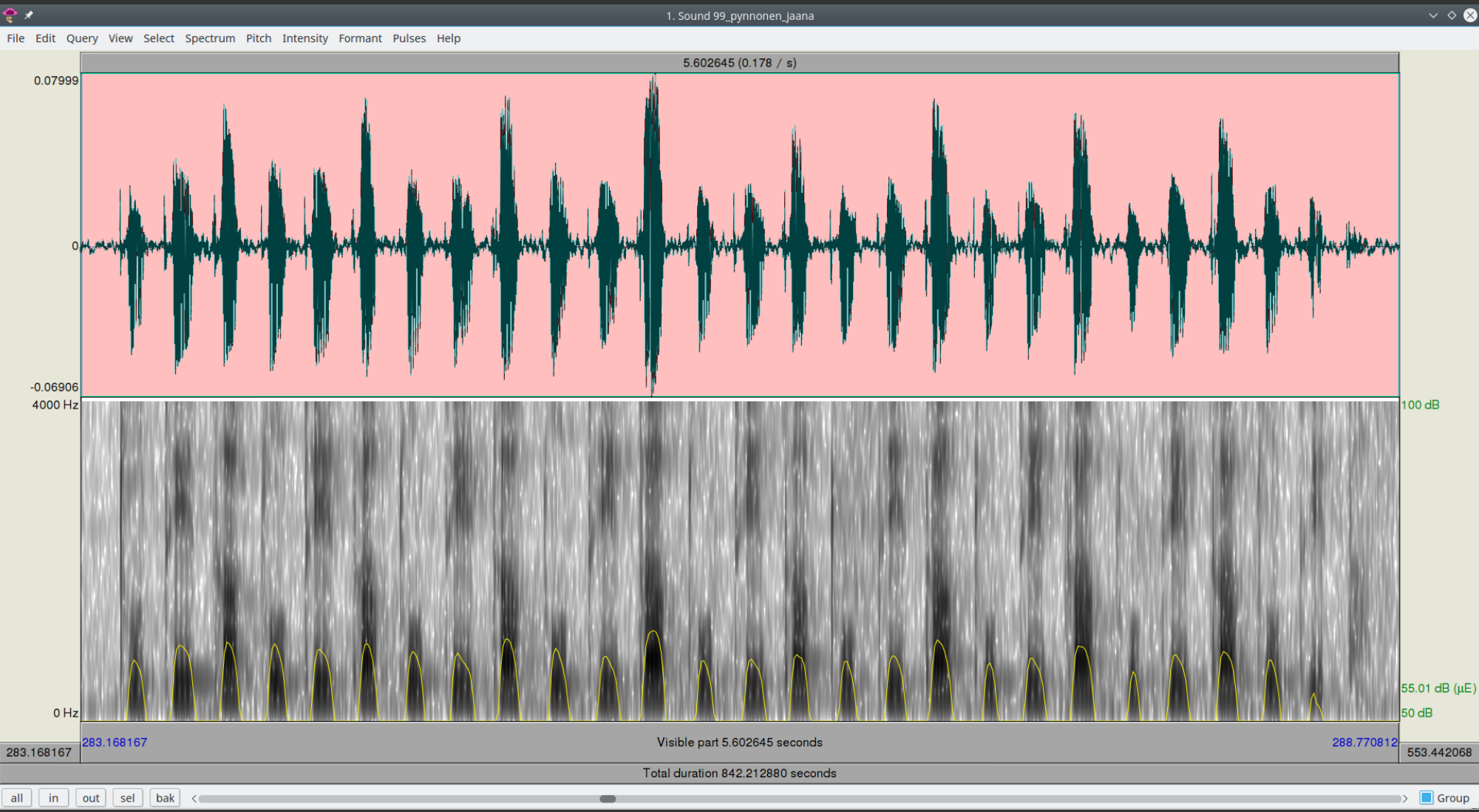
This project aims at implementing an interactive environment in which speech therapists can record, organize, annotate, retrieve, and analyze their speech data. This environment is developed as a plug-in for the Praat program, and, for illustration purposes, follows the diagnostic framework outlined by Keller and colleagues (1991). This framework focuses on acoustic criteria for the evaluation of neurogenically disordered speech, and it is composed of a a brief protocol and an analysis procedure. Our interactive environment automatizes the recording and analysis procedure, generating automatic annotation, and a large number of measurements with little intervention by the therapist. Additionally, functionality is added by allowing the therapist to organize patients' data and personal information without the need of additional programs.

The files necessary for the installation of the *SE Database Manager* are available in a single compressed file, which should be decompressed and moved to the Praat preferences directory - for more detailed instructions, check the Praat plug-in mechanism manual pages. A description of the files contained in this folder is also available.

A video tutorial was chosen as the presentation form for this project. This tutorial is divided into three parts, installing the database, recording the data, and analyzing the data. This is an uncommon way to present a final project for a graduate course, but it appears as a clear and objective way to introduce the basic functions of this database manager, without relying on extensive written descriptions. I also assume this to be the best way to introduce this sort of program to speech therapists, in case it is ever to be implemented. Several video tutorials could come with such a program, discussing all the steps necessary for the analysis. The advantage of movies in the Flash format reside in their small size and compatibility with different operational systems.

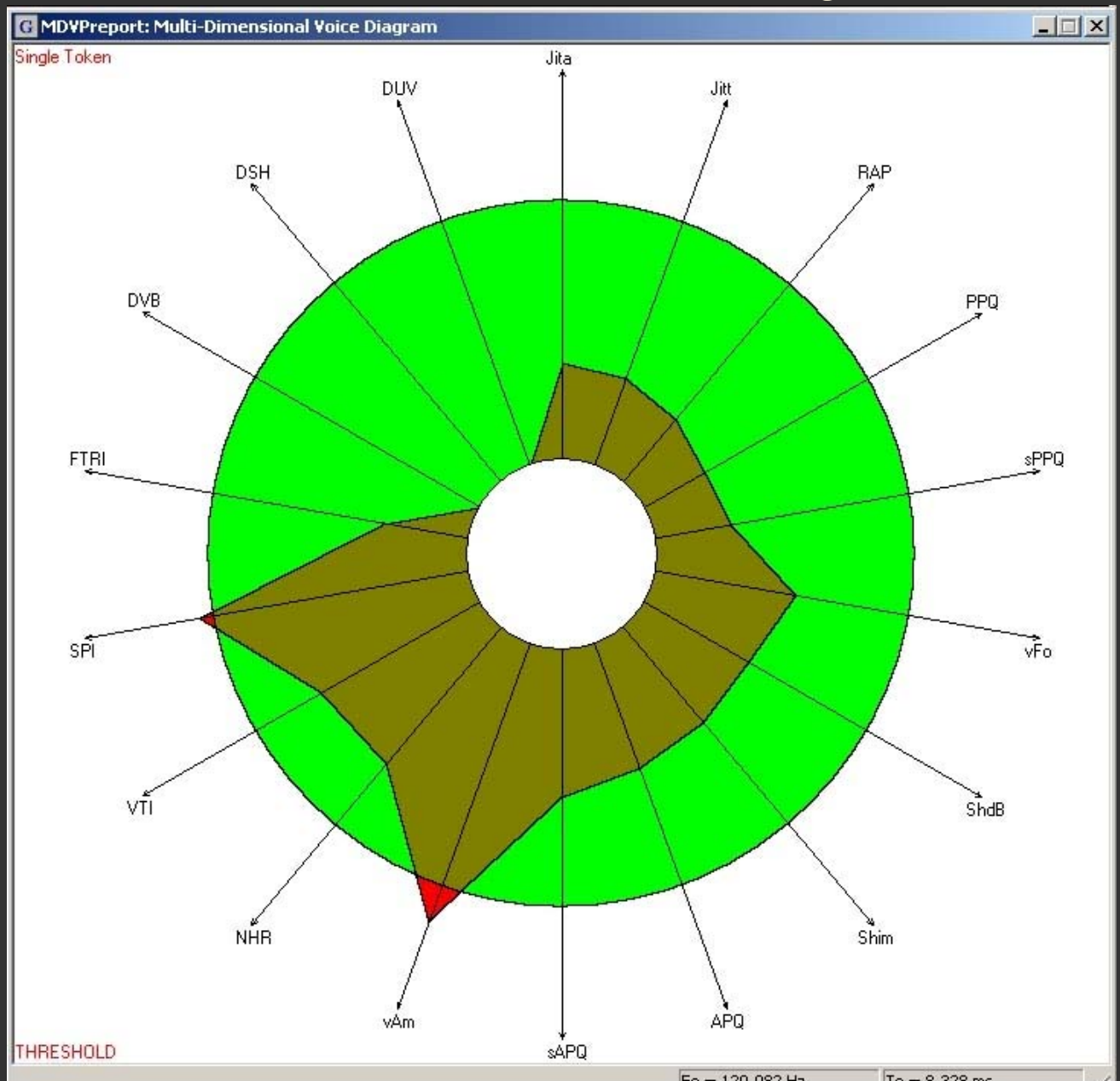
As an alternative to the movie tutorial, users can also read the written version of the manual of the *SE Database Manager*, which can be accessed online or on the Help option of the database manager. This manual was written as Praat ManPages, and it is very similar to the manual pages of the Praat program itself.

Example: Diadochokinetic task



The present: clinical assessment

Multi-Dimensional Voice Program



The present – outside of traditional speech pathology research

- Fully automatic acoustic assessment:
  1. Automatic auditory feature extraction, e.g. *openSMILE*
  2. Features used as input to machine learning system, e.g. *WEKA*
  3. After supervised learning, the system is able to generalize and classify new data.

Problem

The number of extracted features is very high — which makes it hard to draw conclusions about relevant factors.

Goal

Identify most relevant parameter combinations and correlations.

Method

Extensive feature extraction and subsequent dimensionality reduction.

Typical options for dimensionality reduction

- Principal Components Analysis
- Feature-Vector Cluster Analysis
- Various kinds of neural networks
- Self-Organized Maps
- Support Vector Machines
- ...
- Functional Data Analysis

Or variations on old themes:

- Try out (ordered logistic) regressions on heuristically selected feature subsets
- Look systematically for interactions between parameters

Related research

E.g. Asgari and colleagues found that reading tasks provide more relevant data than more constrained tasks and that covariance features tend to carry more crucial information than single measurement parameters.

But their models still contain hundreds of parameters.

Automatic auditory feature extraction: *openSMILE*

- Extracts low-level audio features and their statistics (*functionals*)
- Typically used to produce input to some machine learning system for classification

Example parameters

- Low-level audio features:
  - spectra
  - intensity
  - fundamental frequency
  - voicing probability
  - jitter, shimmer
  - ...
- “Functionals”:
  - extreme values
  - means
  - moments
  - durations
  - ...

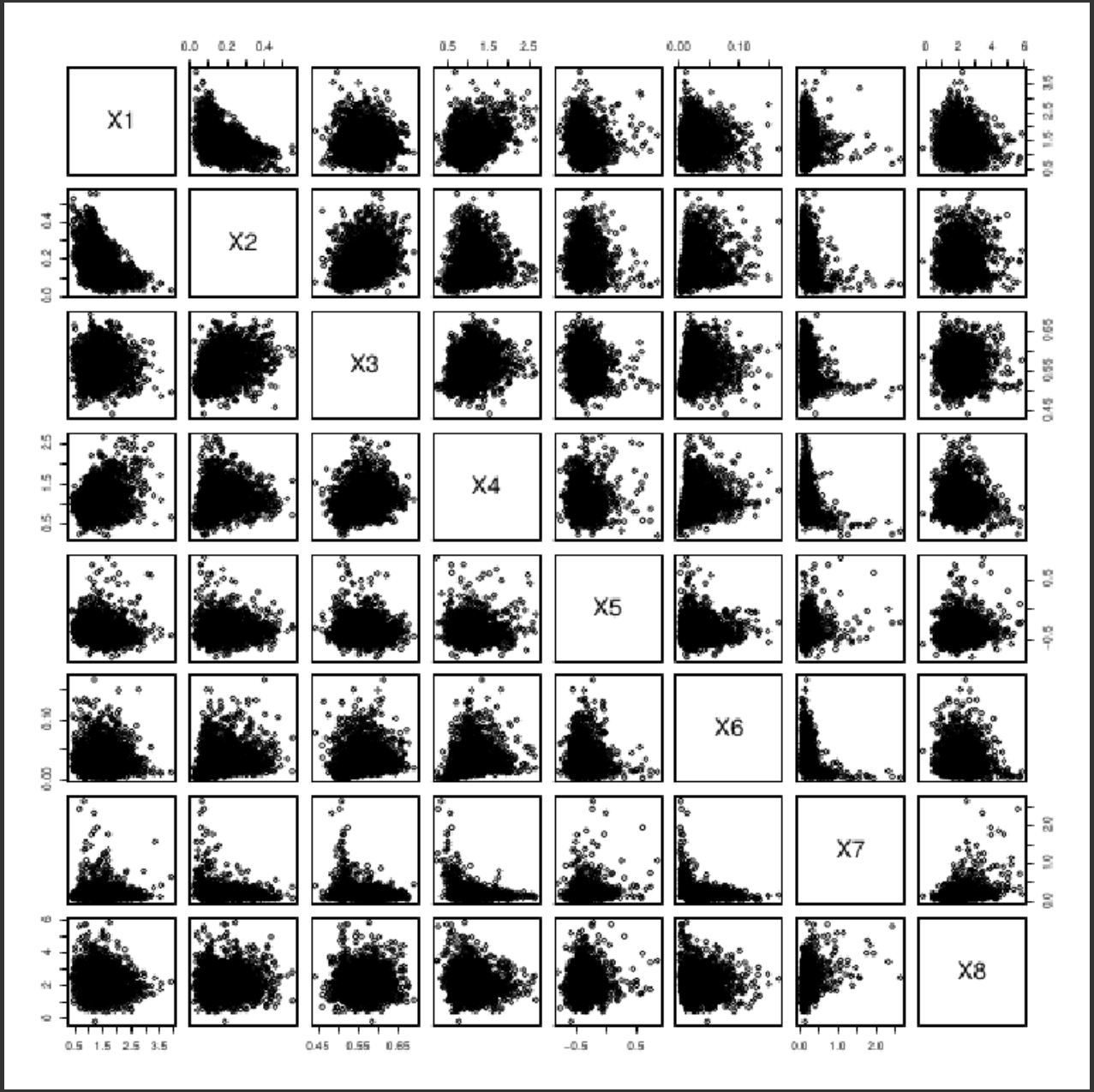
Example *openSMILE* feature sets

- Interspeech 2009 Emotion Challenge:
  384 features derived from 16 low-level descriptors
- Interspeech 2010 Paralinguistic Challenge:
  1428 features derived from 34 low-level descriptors (and their delta coefficients) plus 152 f_0-based features and two duration measures

Parameters...


rating;name;audspec_lengthL1norm_sma_range;audspec_lengthL1norm_sma_maxPos;audspec_lengthL1norm_sma_minPos;audspec_lengthL1norm_sma_quartile1;audspec_lengthL1norm_sma_quartile2;audspec_lengthL1norm_sma_quartile3;audspec_lengthL1norm_sma_iqr1-2;audspec_lengthL1norm_sma_iqr2-3;audspec_lengthL1norm_sma_iqr1-3;audspec_lengthL1norm_sma_percentile1.0;audspec_lengthL1norm_sma_percentile99.0;audspec_lengthL1norm_sma_pctlrange0-1;audspec_lengthL1norm_sma_stddev;audspec_lengthL1norm_sma_skewness;audspec_lengthL1norm_sma_kurtosis;audspec_lengthL1norm_sma_meanSegLen;audspec_lengthL1norm_sma_maxSegLen;audspec_lengthL1norm_sma_minSegLen;audspec_lengthL1norm_sma_segLenStddev;audspec_lengthL1norm_sma_upleveltime25;audspec_lengthL1norm_sma_downleveltime25;audspec_lengthL1norm_sma_upleveltime50;audspec_lengthL1norm_sma_downleveltime50;audspec_lengthL1norm_sma_upleveltime75;audspec_lengthL1norm_sma_downleveltime75;audspec_lengthL1norm_sma_upleveltime90;audspec_lengthL1norm_sma_downleveltime90;audspec_lengthL1norm_sma_risetime;audspec_lengthL1norm_sma_falltime;audspec_lengthL1norm_sma_lpgain;audspec_lengthL1norm_sma_lpc0;audspec_lengthL1norm_sma_lpc1;audspec_lengthL1norm_sma_lpc2;audspec_lengthL1norm_sma_lpc3;audspec_lengthL1norm_sma_lpc4;audspecRasta_lengthL1norm_sma_range;audspecRasta_lengthL1norm_sma_maxPos;audspecRasta_lengthL1norm_sma_minPos;audspecRasta_lengthL1norm_sma_quartile1;audspecRasta_lengthL1norm_sma_quartile2;audspecRasta_lengthL1norm_sma_quartile3;audspecRasta_lengthL1norm_sma_iqr1-2;audspecRasta_lengthL1norm_sma_iqr2-3;audspecRasta_lengthL1norm_sma_iqr1-3;audspecRasta_lengthL1norm_sma_percentile1.0;audspecRasta_lengthL1norm_sma_percentile99.0;audspecRasta_lengthL1norm_sma_pctlrange0-1;audspecRasta_lengthL1norm_sma_stddev;audspecRasta_lengthL1norm_sma_skewness;audspecRasta_lengthL1norm_sma_kurtosis;audspecRasta_lengthL1norm_sma_meanSegLen;audspecRasta_lengthL1norm_sma_maxSegLen;audspecRasta_lengthL1norm_sma_minSegLen;audspecRasta_lengthL1norm_sma_segLenStddev;audspecRasta_lengthL1norm_sma_upleveltime25;audspecRasta_lengthL1norm_sma_downleveltime25;audspecRasta_lengthL1norm_sma_upleveltime50;audspecRasta_lengthL1norm_sma_downleveltime50;audspecRasta_lengthL1norm_sma_upleveltime75;audspecRasta_lengthL1norm_sma_downleveltime75;audspecRasta_lengthL1norm_sma_upleveltime90;audspecRasta_lengthL1norm_sma_downleveltime90;audspecRasta_lengthL1norm_sma_risetime;audspecRasta_lengthL1norm_sma_falltime;audspecRasta_lengthL1norm_sma_lpgain;audspecRasta_lengthL1norm_sma_lpc0;audspecRasta_lengthL1norm_sma_lpc1;audspecRasta_lengthL1norm_sma_lpc2;audspecRasta_lengthL1norm_sma_lpc3;audspecRasta_lengthL1norm_sma_lpc4;pcm_RMSenergy_sma_range;pcm_RMSenergy_sma_maxPos;pcm_RMSenergy_sma_minPos;pcm_RMSenergy_sma_quartile1;pcm_RMSenergy_sma_quartile2;pcm_RMSenergy_sma_quartile3;pcm_RMSenergy_sma_iqr1-2;pcm_RMSenergy_sma_iqr2-3;pcm_RMSenergy_sma_iqr1-3;pcm_RMSenergy_sma_percentile1.0;pcm_RMSenergy_sma_percentile99.0;pcm_RMSenergy_sma_pctlrange0-1;pcm_RMSenergy_sma_stddev;pcm_RMSenergy_sma_skewness;pcm_RMSenergy_sma_kurtosis;pcm_RMSenergy_sma_meanSegLen;pcm_RMSenergy_sma_maxSegLen;pcm_RMSenergy_sma_minSegLen;pcm_RMSenergy_sma_segLenStddev;pcm_RMSenergy_sma_upleveltime25;pcm_RMSenergy_sma_downleveltime25;pcm_RMSenergy_sma_upleveltime50;pcm_RMSenergy_sma_downleveltime50;pcm_RMSenergy_sma_upleveltime75;pcm_RMSenergy_sma_downleveltime75;pcm_RMSenergy_sma_upleveltime90;pcm_RMSenergy_sma_downleveltime90;pcm_RMSenergy_sma_risetime;pcm_RMSenergy_sma_falltime;pcm_RMSenergy_sma_lpgain;pcm_RMSenergy_sma_lpc0;pcm_RMSenergy_sma_lpc1;pcm_RMSenergy_sma_lpc2;pcm_RMSenergy_sma_lpc3;pcm_RMSenergy_sma_lpc4;pcm_zcr_sma_range;pcm_zcr_sma_maxPos;pcm_zcr_sma_minPos;pcm_zcr_sma_quartile1;pcm_zcr_sma_quartile2;pcm_zcr_sma_quartile3;pcm_zcr_sma_iqr1-2;pcm_zcr_sma_iqr2-3;pcm_zcr_sma_iqr1-3;pcm_zcr_sma_percentile1.0;pcm_zcr_sma_percentile99.0;pcm_zcr_sma_pct

Correlations...



Analysis of feature values

Derived features, capturing dynamics of measurements (changes in spectral parameters, f_0 and duration) seem particularly promising.

Analysis of feature values

Repeated regression analyses on hand-picked feature subsets:
focus on checking **interactions** between parameters

Results so far

1. When done in a systematic way, the regression optimization is **extremely** time-consuming.
2. Very short recordings do not provide enough data for successful prediction.
3. Regression models using only eight to ten features predicted the *Nemours* assessment score rank order correctly.
4. As for Asgari et al., speech material elicited through conventional methods like sustained phonation and diadochokinetic tasks tended to be less useful than more natural speech.
5. Much of the relevant information seems to be in the **interactions** between parameters, not in single parameters themselves.

Outlook

1. More sophisticated automatic assessment systems for clinicians
2. Explicit links back from the acoustic to the articulatory domain