Prosodic Control in Severe Dysarthria: Preserved Ability to Mark the Question-Statement Contrast

Rupal Patel

Presented by Aryan Arbabi

Outline

- Introduction
  - Communication problems due to dysarthria
  - Prosody as an Information Carrier

- Current Study
  - Experiment method
  - Results
  - Discussion

Dysarthria: communication problems

- Study’s focus is on cerebral palsy

- Slow and imprecise speech motor control
  - Reduced vowels
  - Imprecise consonants
  - Changes in resonance

How do people with dysarthria try to communicate?

- Not effective to use speech alone

- They usually facilitate communication by:
  - Facial expressions
  - Body language
  - Situational cues

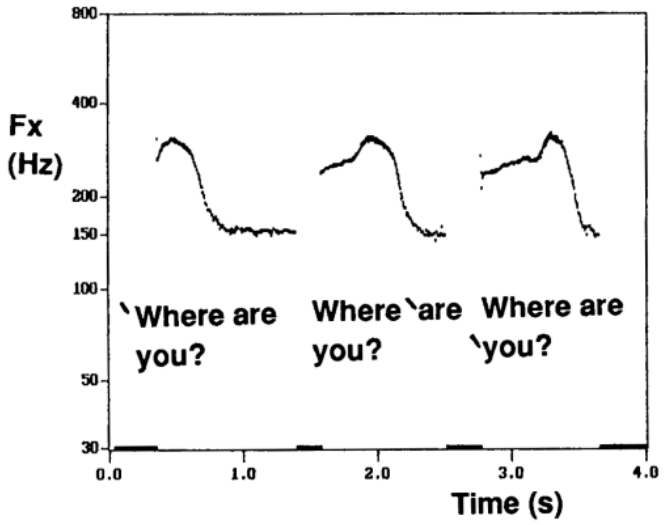
- Familiarity with the speaker and semantic and syntactic context can also help

Prosody as an Information Carrier

- Refers to following aspects in speech:
  - Stress
  - Rhythm
  - Intonation
  - Pause structure
- Associated acoustic parameters:
  - Fundamental frequency (F0)
  - Amplitude
  - Duration
  - Segment quality
- It supplements the linguistic structure of the spoken language

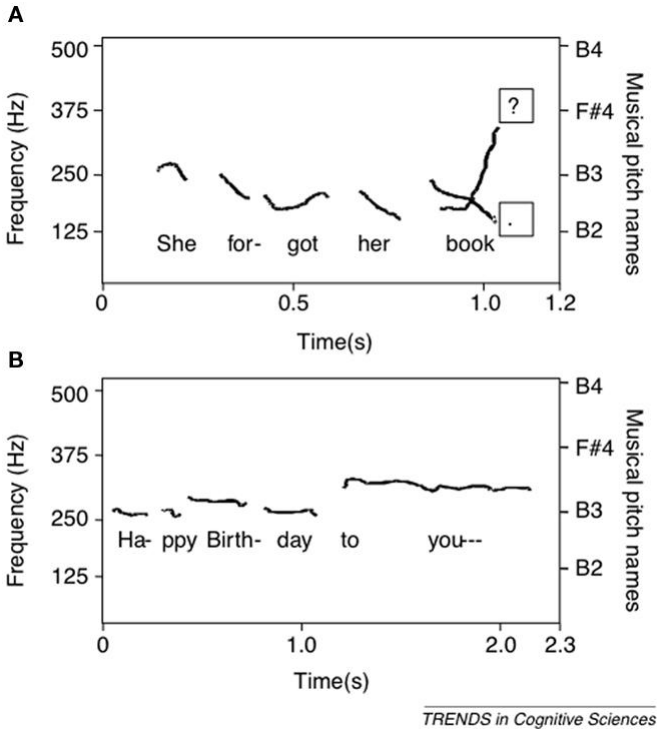
Pitch contour (F0 contour)

Effect of stress on the curve



UCL Division of Psychology and Language Sciences - PALS1004 Introduction to Speech Science

Question / Statement



Tillmann, Barbara, et al. "Congenital amusia (or tone-deafness) interferes with pitch processing in tone languages." *Frontiers in psychology* 2 (2011).

Previous Studies on Prosodic Control in Dysarthria

- Dorze et al. (1994) compared the ability to mark question-statement contrasts:
  - Smaller intonational differences
  - Overall slower rate of utterances

- Yorkston et al. (1984) studied stress patterning in mild dysarthria
  - Reduced variation of frequency, intensity and duration

Can people with dysarthria use prosody for communication?

- The possibility to communicate the difference between statements and questions has been not studied
- Prosodic parameters generally vary at slower time scales than segmental features
- The relatively slow and gradually varying prosodic features may still convey information

The current study

- Investigating the prosodic control abilities of people with severe dysarthria
  - The ability to mark question-statement contrasts are studied
  - Effects of F0 contours and durational cues on speech intelligibility is examined

Experiment method

- 8 speakers with severe dysarthria and 48 naive listeners participated in the study
- Speakers were asked to repeat three word phrases as questions and as statements
- Listeners had to classify each recorded phrase as question or statement
- Some of the recorded phrases were systematically changed to verify the effects of different prosody cues

Participant Speakers

- The 8 participants with dysarthria were recruited from speech and language clinics in Greater Toronto Area

- Age range was 27 to 44 years with mean age of 37 years

- Several evaluations were performed to pass the speakers for the experiment
  - Oral peripheral examination to diagnose dysarthria from verbal apraxia and/or aphasia
  - Assessment of Intelligibility of Dysarthric Speech (AIDS)
  - Pure tone audiometric evaluation with threshold at 25 dB HL in at least one ear
  - All participants had adequate receptive language and cognitive skills

Participant Listeners

- The 48 listener participant were normal hearing English monolingual speakers
- Age range was 22 to 50 years with mean age of 28 years
- They were all evaluated by pure tone audiometric with threshold at 25 dB HL in at least one ear
- The listeners were randomly divided into 8 groups of 6 people, each group assigned to one speaker

Phrases

- 10 three-word phrases were used in the experiment. The phrases had three syllables and the consonant clusters were minimized

Phrase	Abbreviation	Statement context	Question context
She was here (?)	SWH	Bob asks if you have seen Jane. You say…	John tells you that the Queen of England visited your house. You ask…
It was hot (?)	IWH	You just got back from holidays in Florida. Jane asks if the weather was nice. You say…	Jane says her vacation to Alaska was too hot. You ask…
She loves dogs (?)	SLD	Molly asks if Sue likes dogs. You say…	Jane doesn't like animals. Someone tells you she loves dogs. You ask…
Play it loud (?)	PIL	You can't hear the message on your answering machine. You tell Jane to …	It's midnight and someone asks you to turn up the volume on the stereo. You ask…
Use some soap (?)	USS	You ask Jenny to get a out a stain from your shirt. She asks how. You say…	You run out of shampoo while washing your hair. Your attendant tells you to just use soap. You ask…
She said no (?)	SSN	Bill proposed to Sue but she refused. Someone asks what she said. You say…	John offered Beth a million dollars. She declined. You ask…
It's Thursday (?)	ITD	Someone asks what day it is. You say…	Someone tells you it's Thursday on the weekend. You ask…
He lives there (?)	HLT	Doug asks you where Chris lives. You point to the blue house and say…	Chris is a millionaire. Paul says he lives in a small run down apartment downtown. You ask…
Pass some salt (?)	PSS	Your food is a little bland. You say…	While eating desert Jim asks you for the salt. You ask…
Give it back (?)	GIB	Someone grabs your pen. You say…	Someone gives you a gift and then asks for it back. You ask…

Recording procedure

- Speech recordings were made using a digital audio recorder in a sound treated audiometric booth
- Speakers were asked to produce each phrase 5 times as a statement and 5 times as a question (Total of 100 recordings)
- The order of the 10 phrase types and the order of question/statement tokens were randomized for each speaker

Systematic Manipulation of the recordings

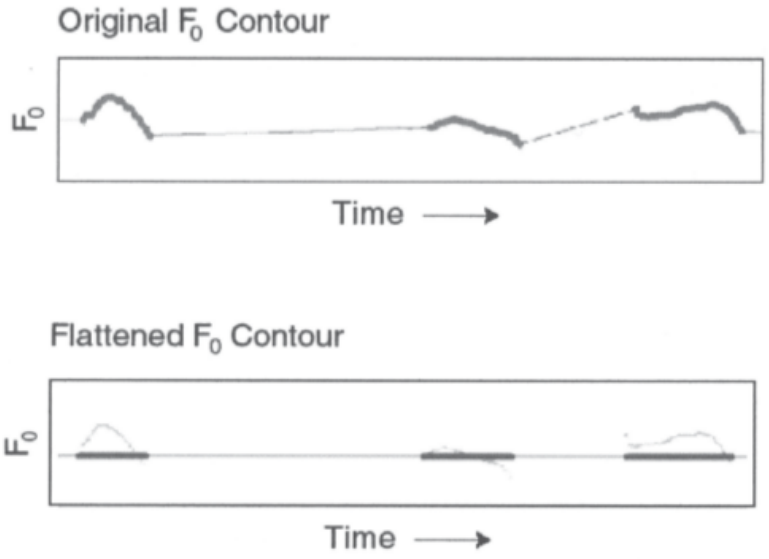
- For each speaker a random subset of 60 original recording was selected (3 questions and 3 statements for each sample)

- Praat speech analysis software was used to flatten the F0 and/or equalizing syllable durations

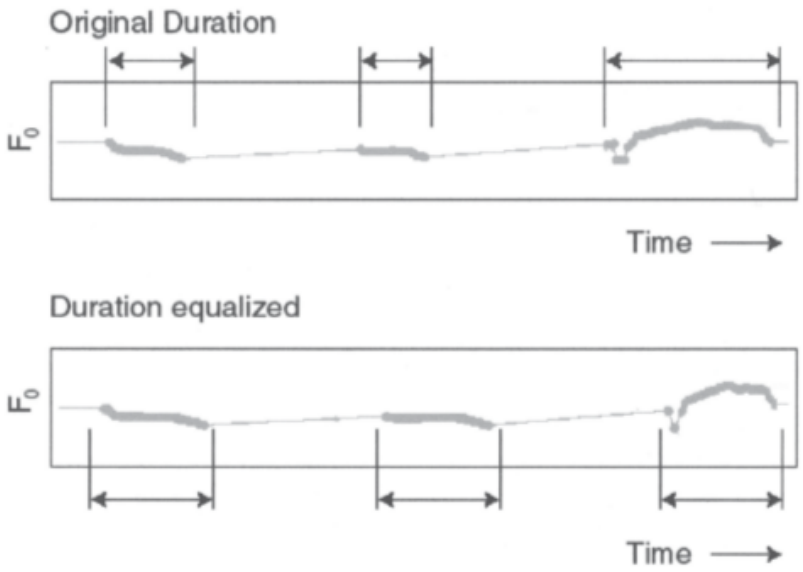
- 4 stimulus sets were created (total of 240 recordings):
    1. The original recording
    2. F0 contour flattened
    3. Syllable durations equalized
    4. Both Manipulations

Systematic Manipulation of the recordings (contd.)

Pitch Flattening



Syllable durations equalizing



Listening procedure

- The listening was conducted in a sound treated audiometric booth through headphones
- Listeners could listen to each vocalization as many times as they needed
- They had to categorize each vocalization as a question or a statement by using a selection button
- In addition to the 240 produced vocalizations, 40 random repeats were used to judge the reliability
  - two listeners had reliability ratings of less than 90% and were excluded for the final analysis

Results

Accuracy scores across all speakers for each stimulus set

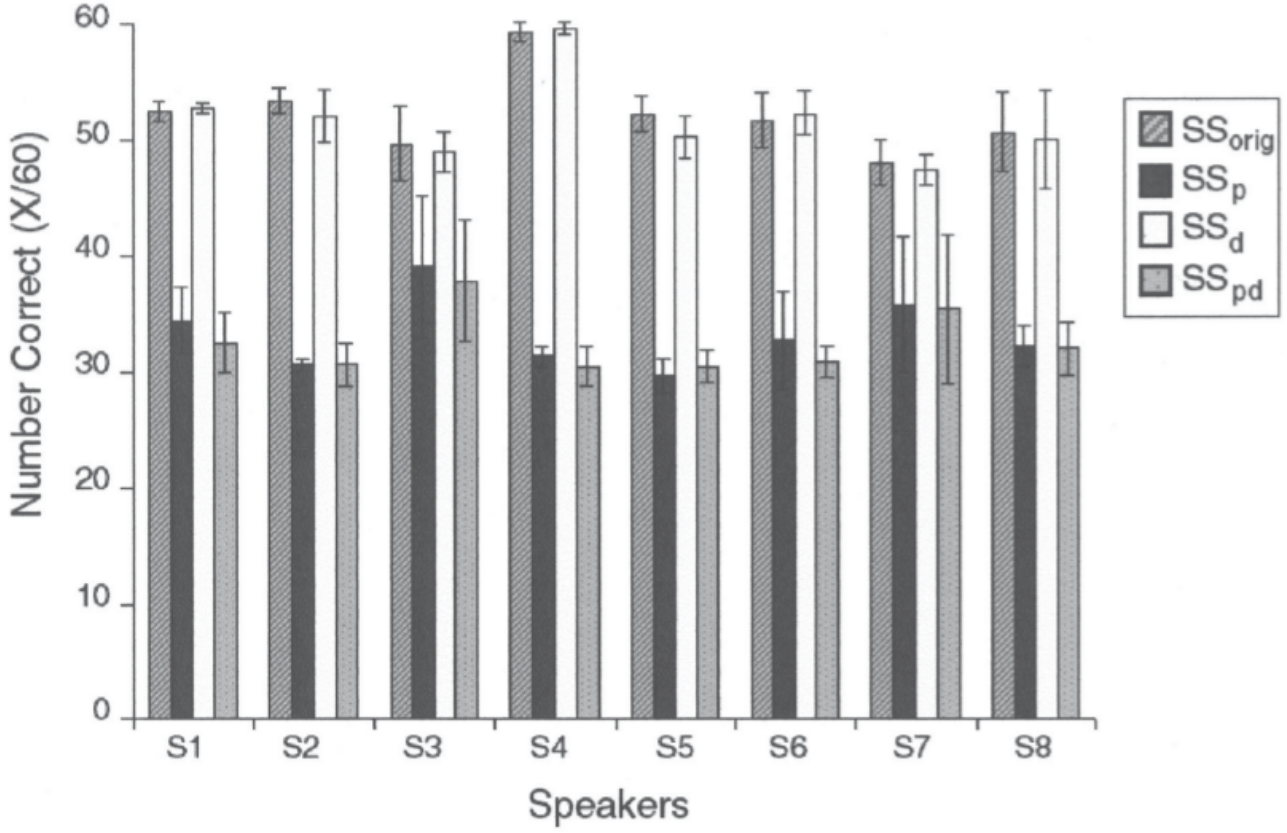
Manipulation	Sum of correct responses across all speaker groups	Proportion of correct responses
SS_{orig}	2506	87.0
SS_{p}	1596	55.4
SS_{d}	2482	86.2
SS_{pd}	1561	54.2

- Flattening the F0 countour reduced classification accuracy by 32%
- Removal of durational cues reduced the accuracy by only 0.8%

Accuracy scores across all speakers for questions and statements

Manipulation	Percentage error for statements	Percentage error for questions
SS_{orig}	2.6	23.2
SS_{p}	4.0	85.0
SS_{d}	2.8	24.7
SS_{pd}	3.0	88.2

Effect of manipulation on accuracy for each speaker



Analyzing effects of different factors

Effect	Degrees of freedom	Wald chi-square value	Probability > chi-square
Speaker	7	58.1	< .0001
Manipulation	3	818.4	< .0001
Phrase	9	20.6	0.0145
Speaker * Manipulation	21	119.7	< .0001
Speaker * Phrase	63	248.0	< .0001

- A logistic model was used for analysis

Results of pairwise contrasts between all manipulations

Contrast	Degrees of freedom	Chi-square value	Probability > chi-square
SS_{orig} vs SS_{p}	1	466.4	< .0001
SS_{orig} vs SS_{d}	1	154.8	< .0001
SS_{orig} vs SS_{pd}	1	519.8	< .0001
SS_{p} vs SS_{d}	1	324.0	< .0001
SS_{d} vs SS_{pd}	1	341.9	< .0001
SS_{p} vs SS_{pd}	1	0.9	0.3332

Conclusions

- Speakers with dysarthria are able to exert sufficient control on prosody to communicate intentions

- The information seems, at least partially, to be encoded in F0 contour and to a lesser extent in syllable duration

Questions?

CHANTI: Predictive Text Entry Using Non-verbal Vocal Input

Sporka et al.

Presented by Aryan Arbabi

A text entry application for users with physical disabilities

- Depending on the type and severity, number of different input signals can be very limited and thus be very slow

- Speech recognition does not always work as voice of many people with physical disabilities is subject to dysarthria

Non-verbal Voice Interaction (NVVI)

- Based on interpretation of non-verbal sounds such as humming or whistling

- Various acoustic parameters of the sound signal are measured, such as pitch or volume

Keyboard input methods

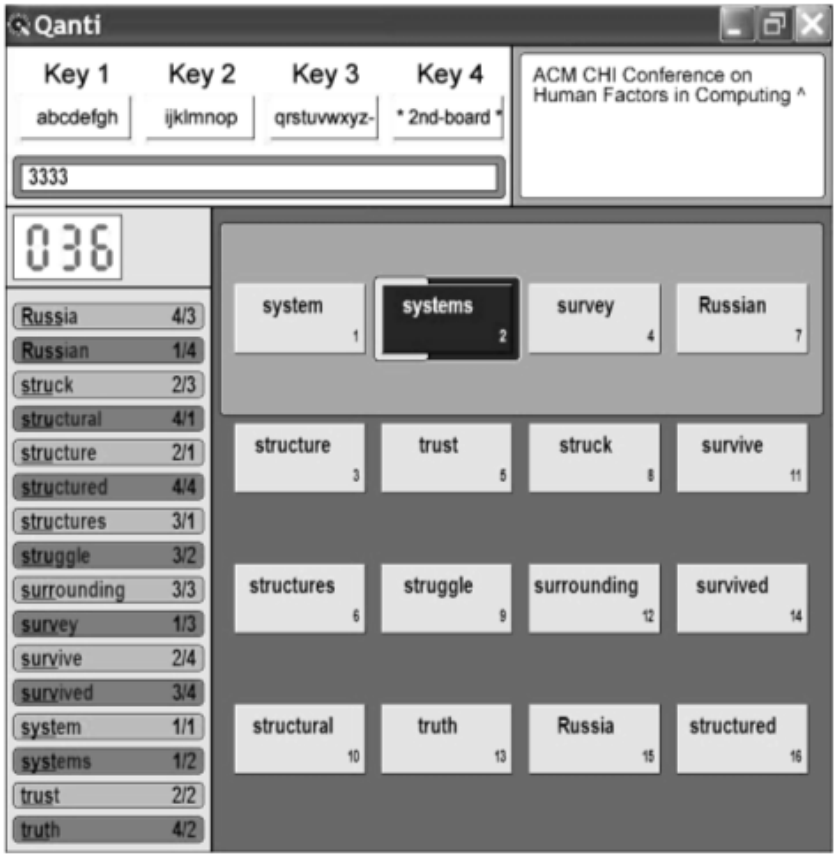
- Ambiguous keyboard
  - A single key is associated with more than one character
  - A dictionary is checked for matched candidates of a key sequence
  - If there were multiple matches or an autocomplete is possible the user selects from a list
  - A famous example is the T9 application for phone keypad

Keyboard input methods (Contd.)

- Scanning keyboard
  - The keys are cyclically highlighted
  - User can select the currently highlighted using a dedicated switch

- Scanning ambiguous keyboard (SAK)
  - Is a mixture of both keyboards

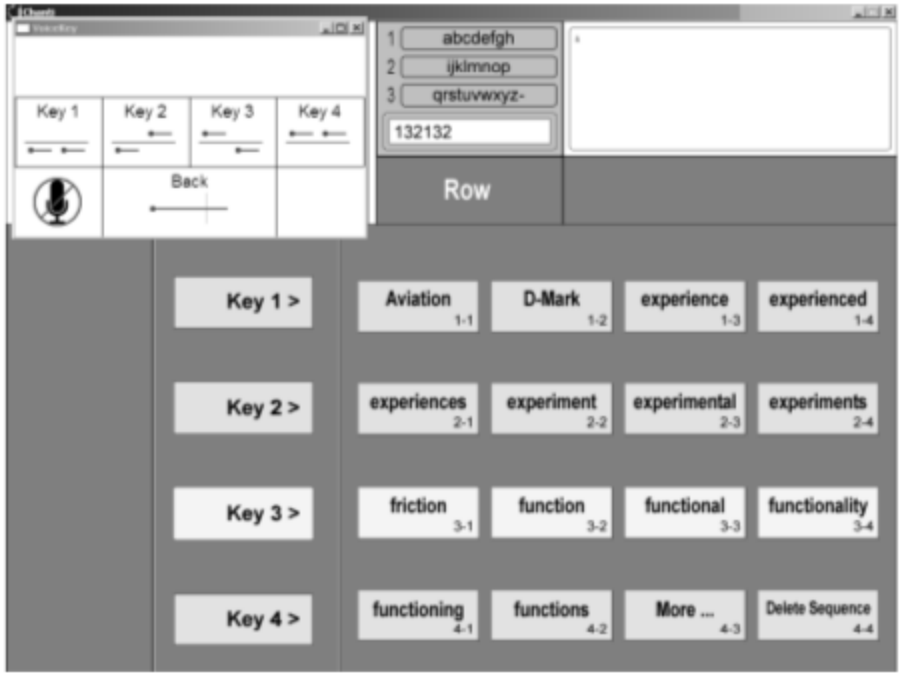
QANTI an implementation of a SAK



The proposed method: CHANTI

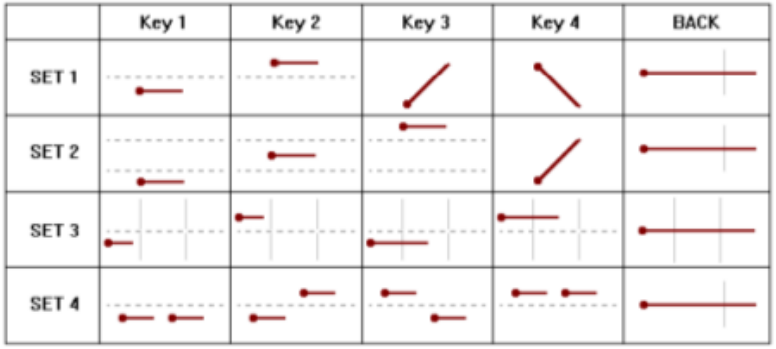
- Combines QANTI and NVVI
- The structure of the user interface is close to QANTI
- Directly selects items by accepting NVVI gestures

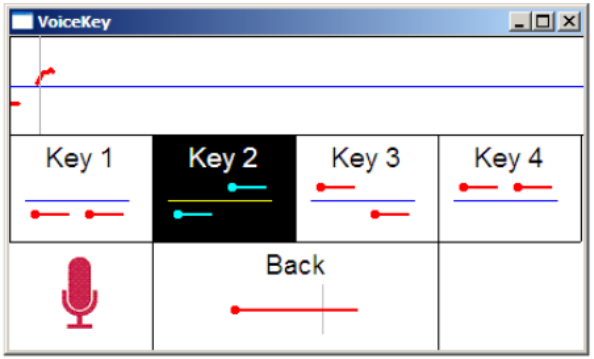
CHANTI User Interface



NVVI gestures in CHANTI

- There are 4 possible sets of keys, each including four gestures for the item selection keys





The study

- 8 participants from three countries (Germany, Czech Republic and USA)
- Range of no speech impairment to severe dysarthria
- The participants ability to produce NVVI gestures was tested in Day 1
  - Three participants were excluded
- They were asked to use CHANTI for minimum of 30 minutes each day for 7 days

Results

Participant (country and language)	Disability, condition	Dysarthria	Typical type rate (without CHANTI)	Mean type rate on first day*	Mean type rate on last day*	Gesture set	Would use CHANTI again?
Miloš (CZ)	Congenital malformation	No	100 CPM	14 CPM	21 CPM	#1	no
Petr (CZ)	Quadriplegia (accident)	No	30 CPM	6 CPM	12 CPM	#4	yes
Gabriele (DE)	Friedreich ataxia	Yes	very low	very low	5 CPM	#1	yes
Rolf (DE)	Friedreich ataxia	Yes	15—30 CPM**	4 CPM	11 CPM	#1	yes
Sarah (US)	Paraplegia (accident), thyroid problem	No	0—150 CPM	5 CPM	11 CPM***	#1***	yes

* Including the time spent on any corrections

** Depending on current condition

*** Sarah was using the *Full Keyboard* mode. Her peak performance was 16 CPM but during her last session the condition deteriorated and reached 8 CPM after an attack.

Conclusions

- This study shows that NVVI can be a viable interaction tool for text entry
- It can be used by people not capable of speech intelligible by automatic speech recognition
- By the end of the experiment the typical rates were between 10 to 15 CPM

Questions?