If you are viewing this course as a recorded course after the live webinar, you can use the scroll bar at the bottom of the player window to pause and navigate the course.

This handout is for reference only. It may not include content identical to the powerpoint. Any links included in the handout are current at the time of the live webinar, but are subject to change and may not be current at a later date.

Making Speech More Distinct

Donald J. Schum, PhD
Vice President, Audiology & Professional Relations
Oticon, Inc.

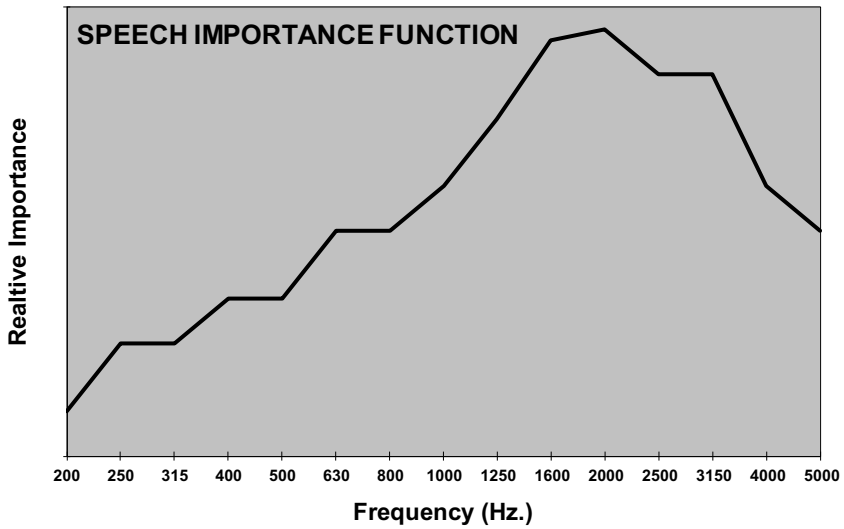
DJS@Oticonusa.com

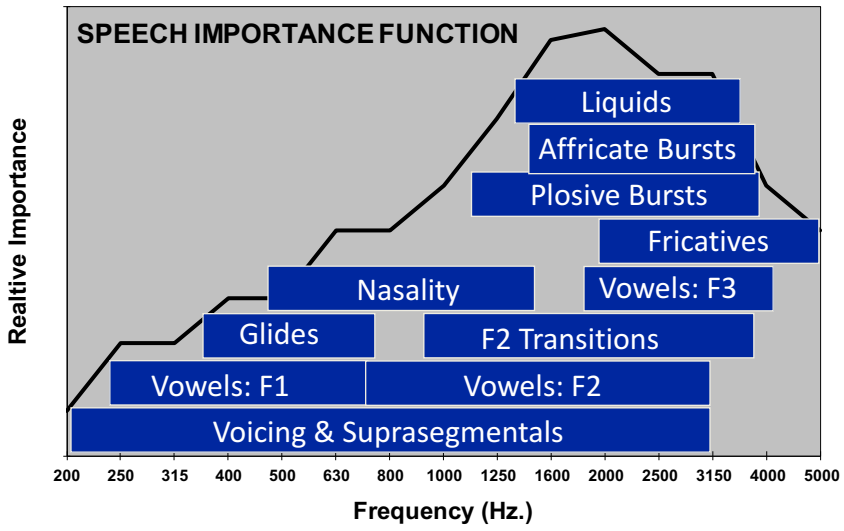
Agenda

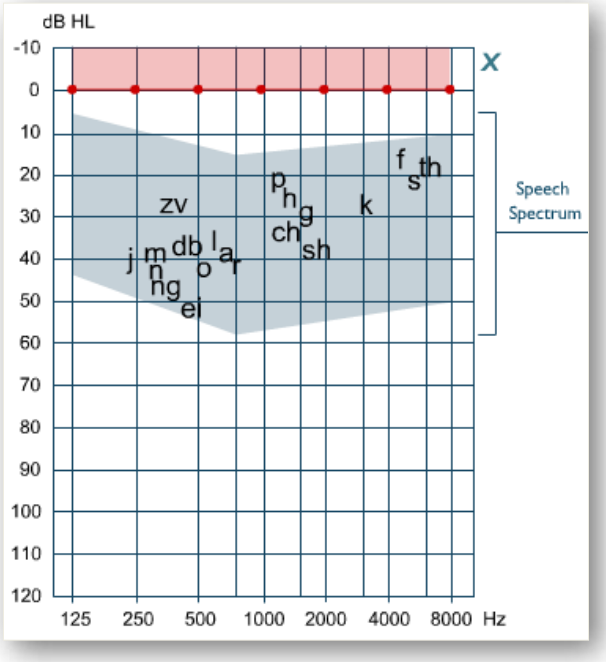
- The Nature of Speech Understanding
- The Acoustics of Speech in Conversation
- Naturally Produced Enhanced Speech
- Requirements for Computer Enhancement
- Some Past Examples
- Looking Forward

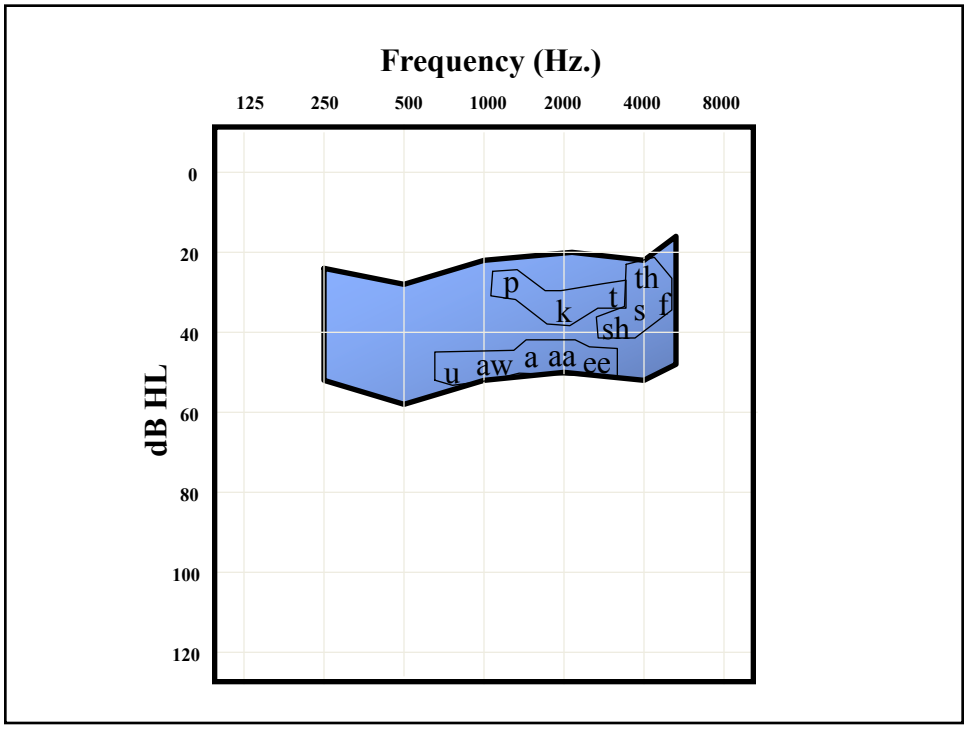
Agenda

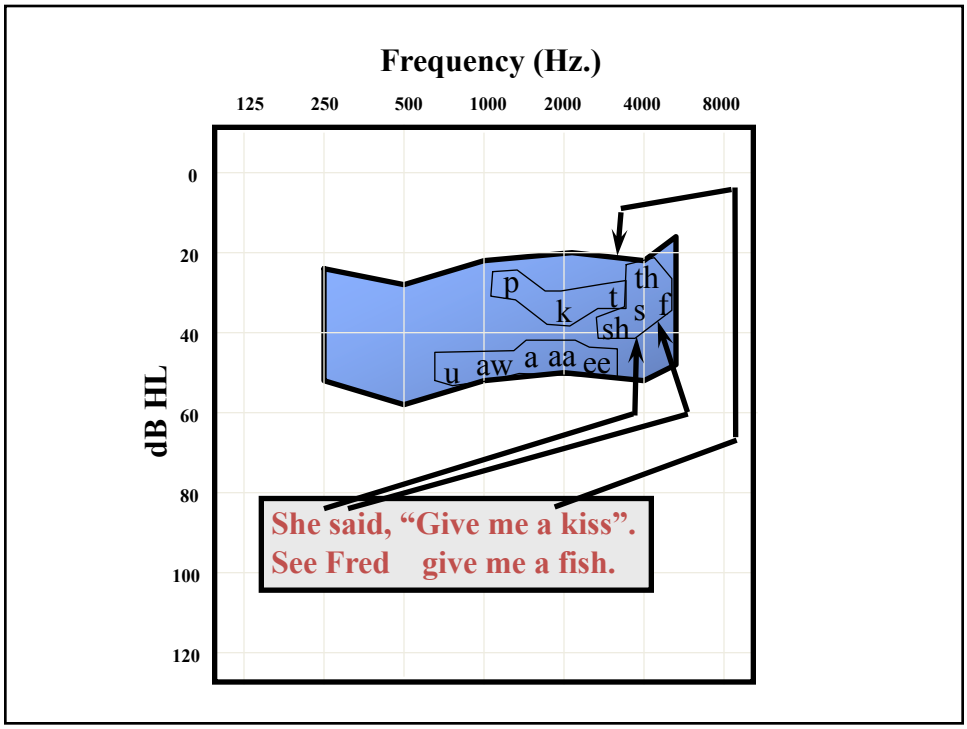
- The Nature of Speech Understanding
- The Acoustics of Speech in Conversation
- Naturally Produced Enhanced Speech
- Requirements for Computer Enhancement
- Some Past Examples
- Looking Forward



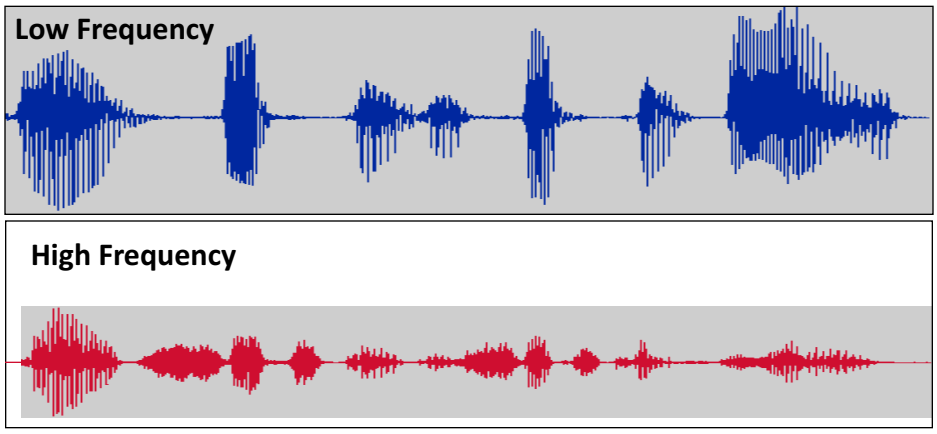

SPEECH IMPORTANCE FUNCTION
Realtive Importance
Liquids
Affricate Bursts
Plosive Bursts
Fricatives
Nasality
Vowels: F3
Glides
F2 Transitions
Vowels: F1
Vowels: F2
Voicing & Suprasegmentals
200 250 315 400 500 630 800 1000 1250 1600 2000 2500 3150 4000 5000
Frequency (Hz.)


dB HL
-10 0 10 20 30 40 50 60 70 80 90 100 110 120
X
Speech Spectrum
zv
j m n ng
db
l a r
o
ei
ph g
ch sh
k
f s th
125 250 500 1000 2000 4000 8000 Hz


Frequency (Hz.)
125
250
500
1000
2000
4000
8000
dB HL
0
20
40
60
80
100
120
p
t
k
th
f
s
sh
u
aw
a
aa
ee


Frequency (Hz.)
125
250
500
1000
2000
4000
8000
dB HL
0
20
40
60
80
100
120
p
t
k
th
f
s
sh
u
aw
a
aa
ee
She said, “Give me a kiss”.
See Fred give me a fish.

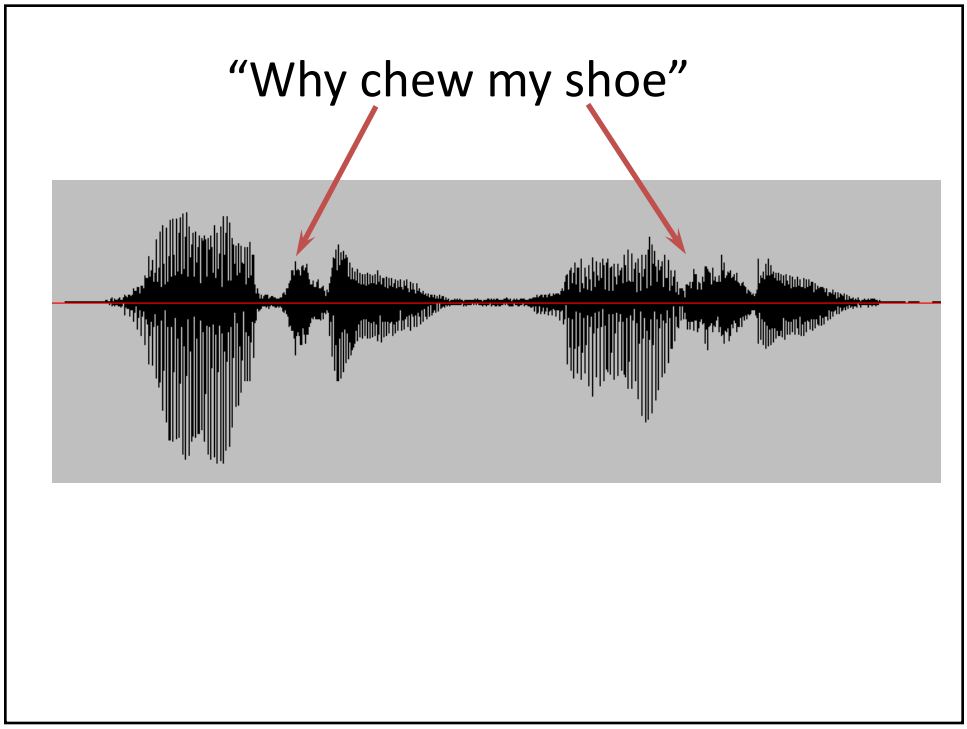
Low Frequency Inputs Versus High Frequency Inputs

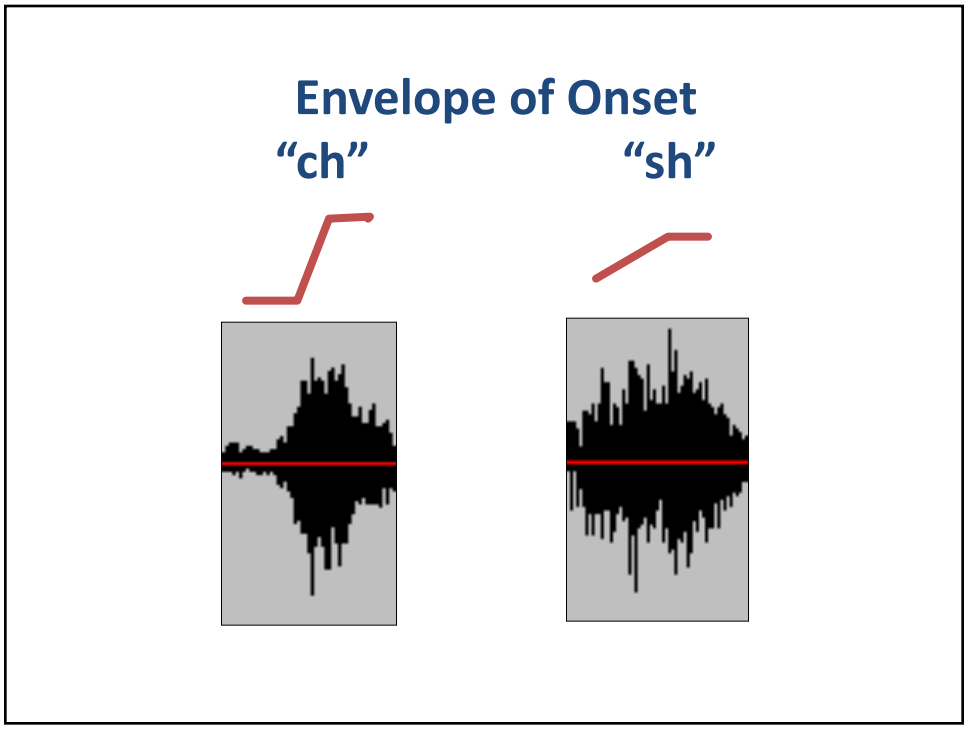


A ge s i x ty to s i x ty f i ve

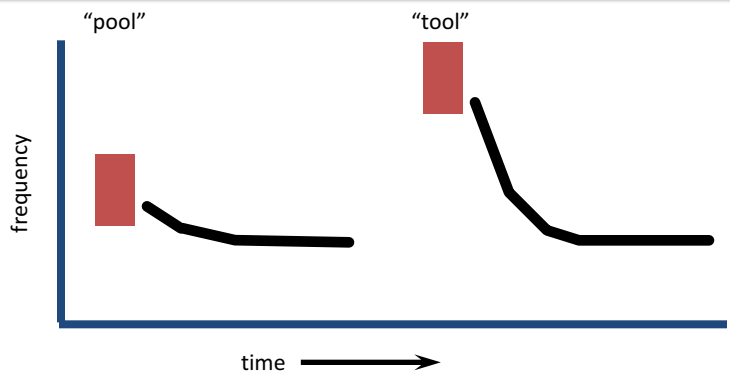
Examples of Dynamic Based Contrasts

- fricatives versus affricates: sh vs. ch
- plosives versus fricatives: d vs. z
- dipthongs versus vowels: oi vs. o
- CV & VC transitions


"Why chew my shoe"


Envelope of Onset
"ch"
"sh"

CV F2 Transition



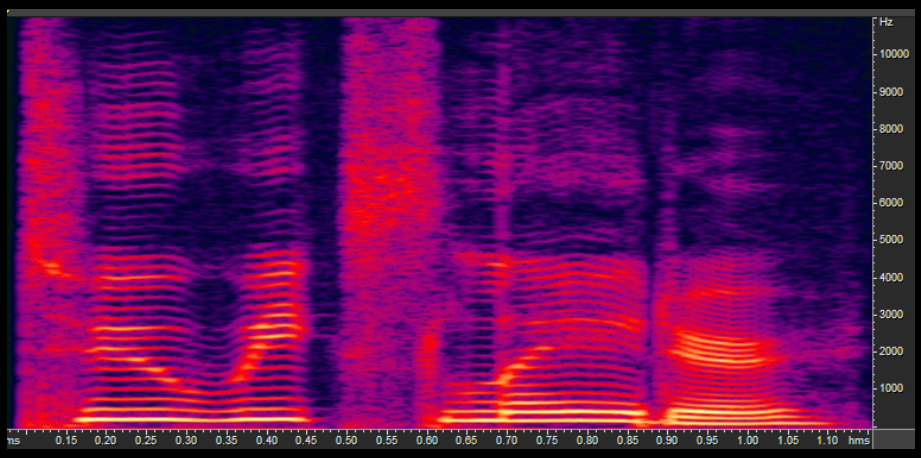
Agenda

- The Nature of Speech Understanding
- The Acoustics of Speech in Conversation
- Naturally Produced Enhanced Speech
- Requirements for Computer Enhancement
- Some Past Examples
- Looking Forward

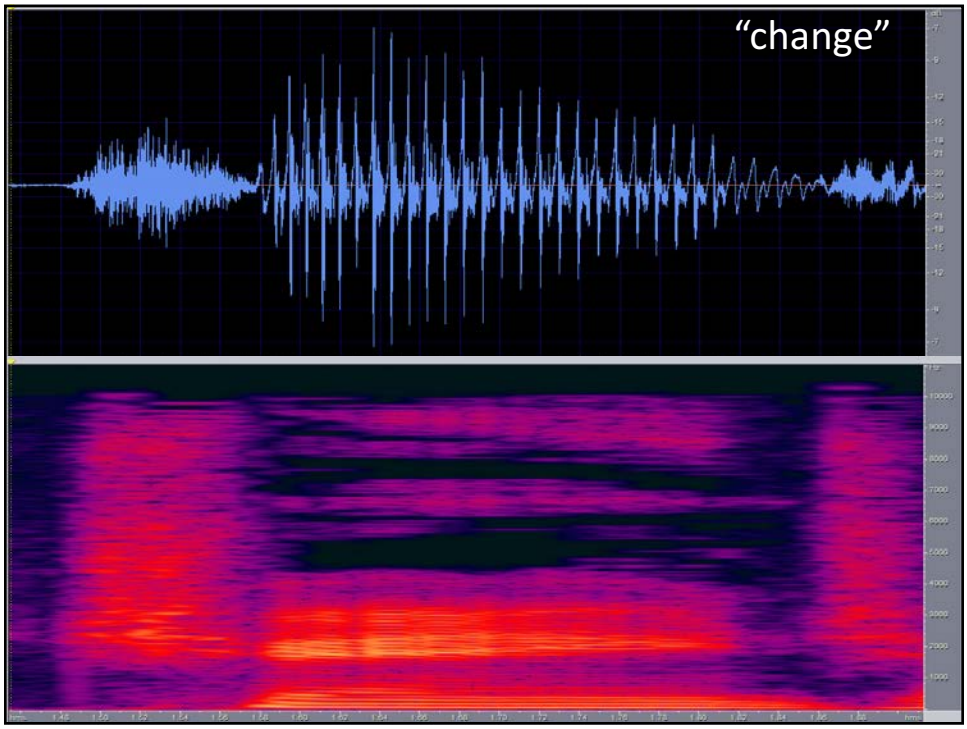
Speech is Movement

Slip knot

“two weeks later”



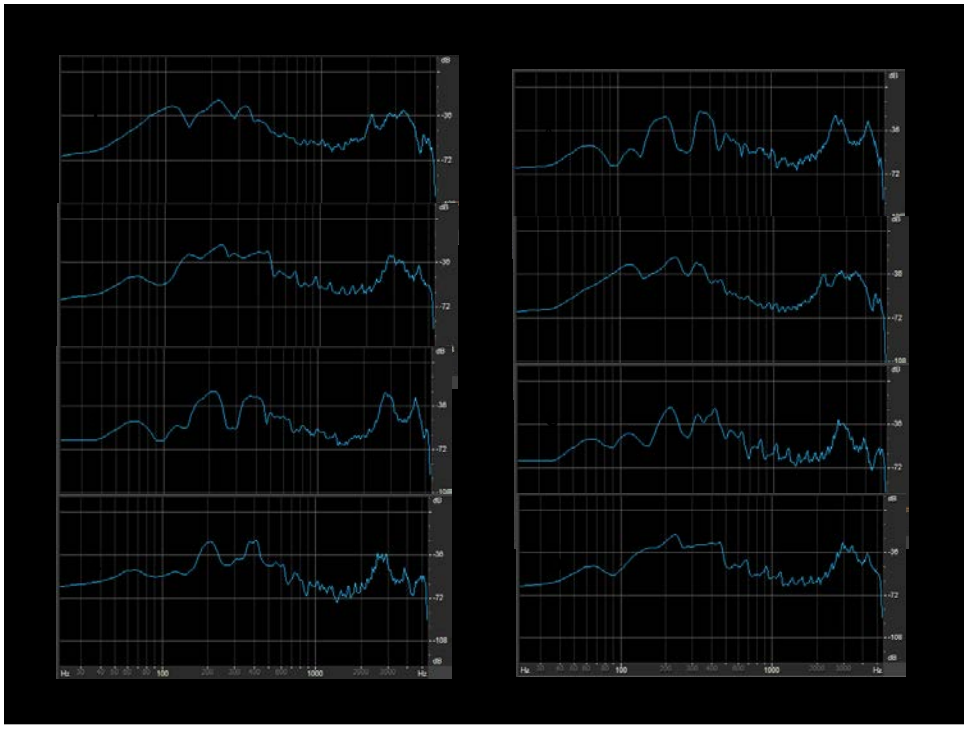
t o wee k s l a t er

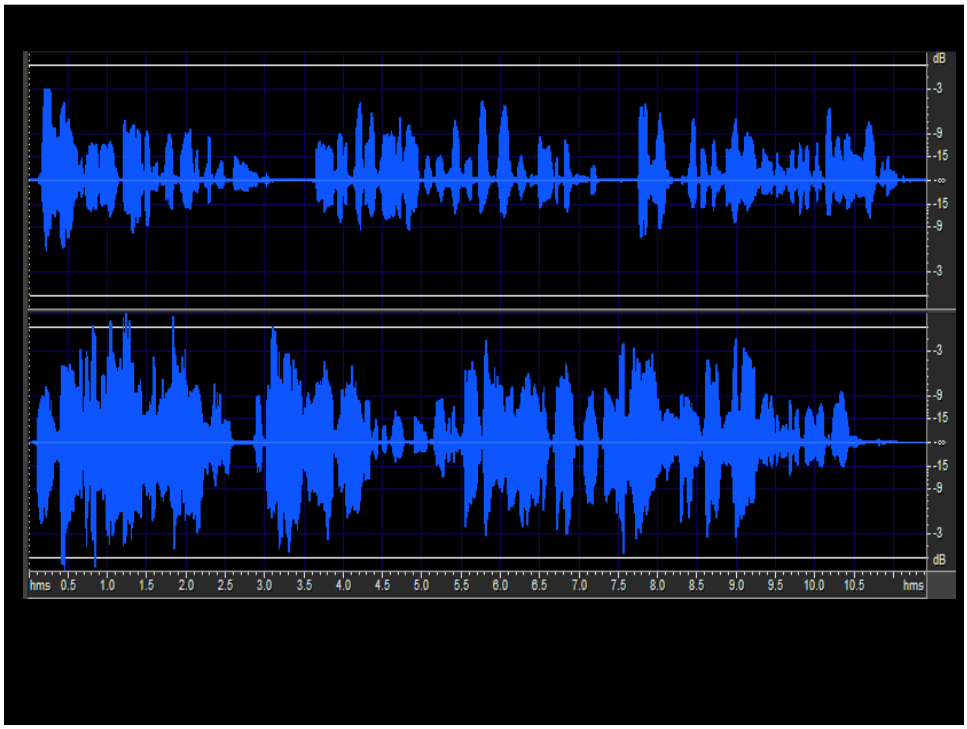


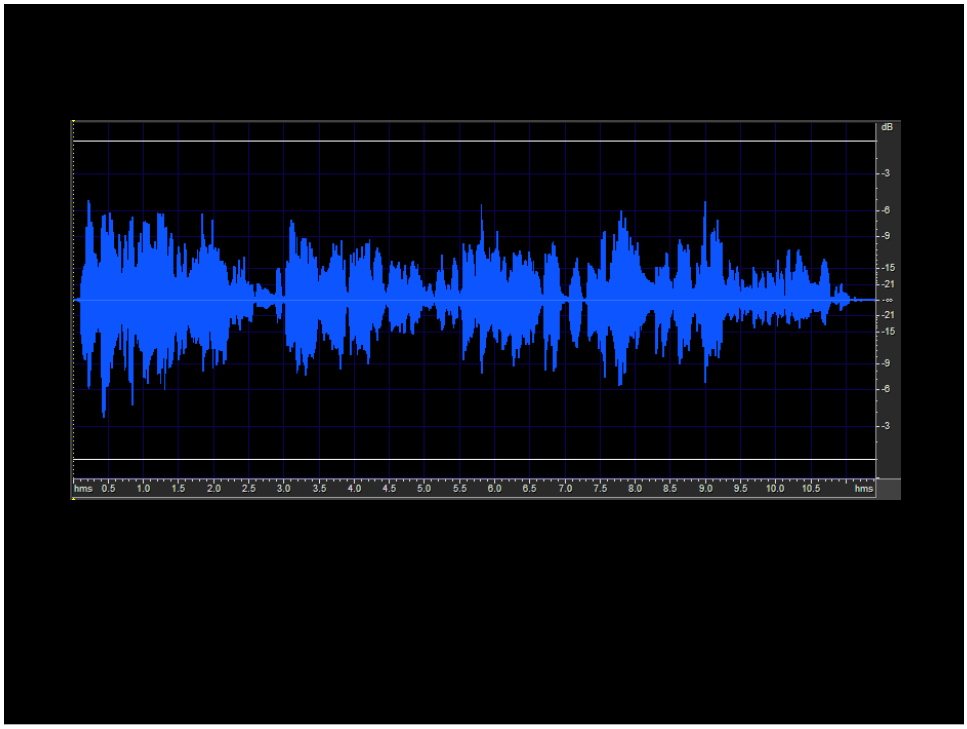
How Do We Talk?

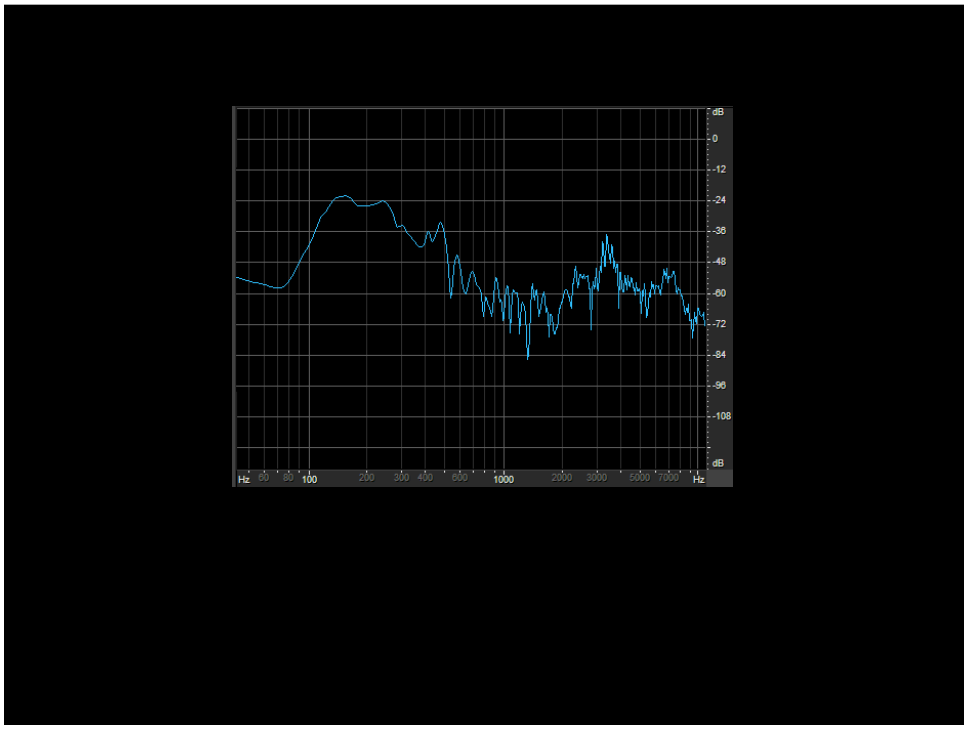
- Minimal Effort
  - Minimal Contrasts
  - Minimal Durations
  - Reductions
- Only as clear as we need to be
  - General human tendency: Conservation of energy
- We adjust to environment, but. . .
- Assume normally hearing listener

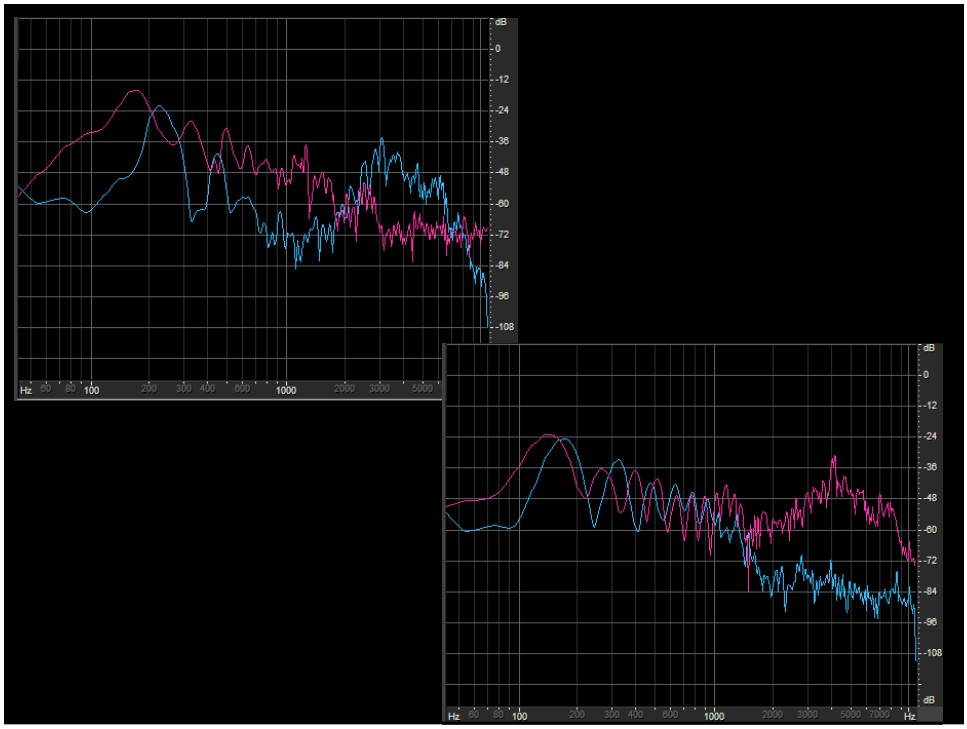
Speaker-to-Speaker Variability




dB
-3
-9
-15
-∞
-15
-9
-3
hms 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5 7.0 7.5 8.0 8.5 9.0 9.5 10.0 10.5 hms


dB
-3
-6
-9
-15
-21
-∞
-21
-15
-9
-6
-3
hms 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5 7.0 7.5 8.0 8.5 9.0 9.5 10.0 10.5 hms


dB
0
-12
-24
-36
-48
-60
-72
-84
-96
-108
dB
Hz 80 80 100 200 300 400 600 1000 2000 3000 5000 7000 Hz

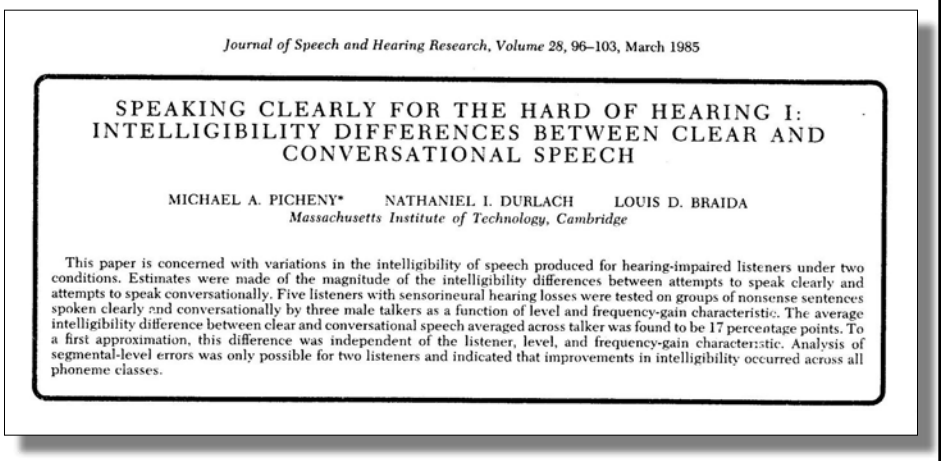


Agenda

- The Nature of Speech Understanding
- The Acoustics of Speech in Conversation
- Naturally Produced Enhanced Speech
- Requirements for Computer Enhancement
- Some Past Examples
- Looking Forward

Clear Speech

- Natural acoustic changes which occur when a talker attempts to produce speech which is precise and accurate.



Journal of Speech and Hearing Research, Volume 28, 96–103, March 1985

SPEAKING CLEARLY FOR THE HARD OF HEARING I: INTELLIGIBILITY DIFFERENCES BETWEEN CLEAR AND CONVERSATIONAL SPEECH

MICHAEL A. PICHENY* NATHANIEL I. DURLACH LOUIS D. BRAIDA
Massachusetts Institute of Technology, Cambridge

This paper is concerned with variations in the intelligibility of speech produced for hearing-impaired listeners under two conditions. Estimates were made of the magnitude of the intelligibility differences between attempts to speak clearly and attempts to speak conversationally. Five listeners with sensorineural hearing losses were tested on groups of nonsense sentences spoken clearly and conversationally by three male talkers as a function of level and frequency-gain characteristic. The average intelligibility difference between clear and conversational speech averaged across talker was found to be 17 percentage points. To a first approximation, this difference was independent of the listener, level, and frequency-gain characteristic. Analysis of segmental-level errors was only possible for two listeners and indicated that improvements in intelligibility occurred across all phoneme classes.

Clear Speech: Research History

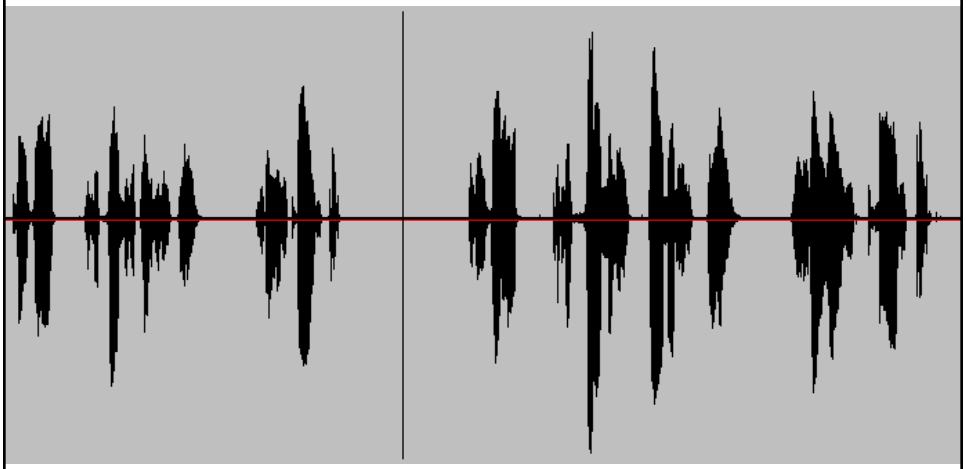
- Picheny, Braida, & Durlach: Mid 1980's
  - coined the term, contrasted to "conversational style speech"
  - documented acoustic changes
  - documented intelligibility improvement

Acoustic Changes

- Slower rate of speech
- More frequent and longer pauses
- Longer phoneme durations (consonants & vowels)
- More released word-final stops
- Greater differentiation of vowels
- Improved Consonant/Vowel ratio

"Tuborg is a famous Danish beer. So is Carlsberg."

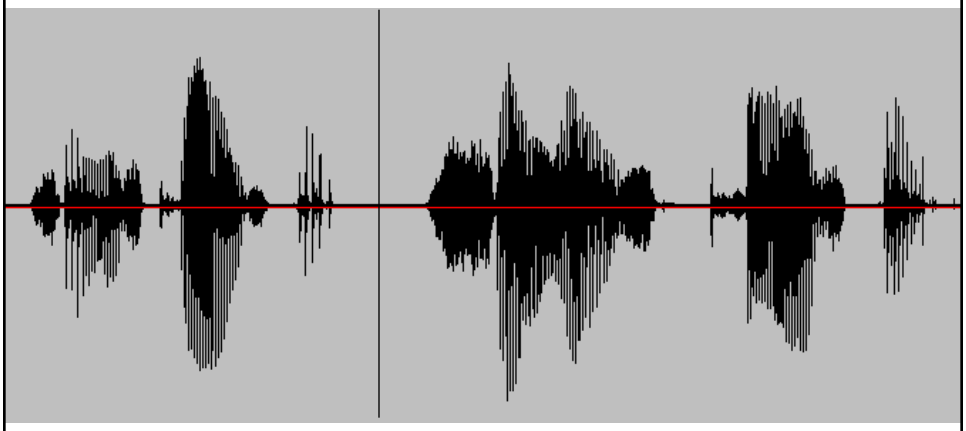
Conversational Clear



- Slower Rate
- Higher Level
- More & Longer Pauses

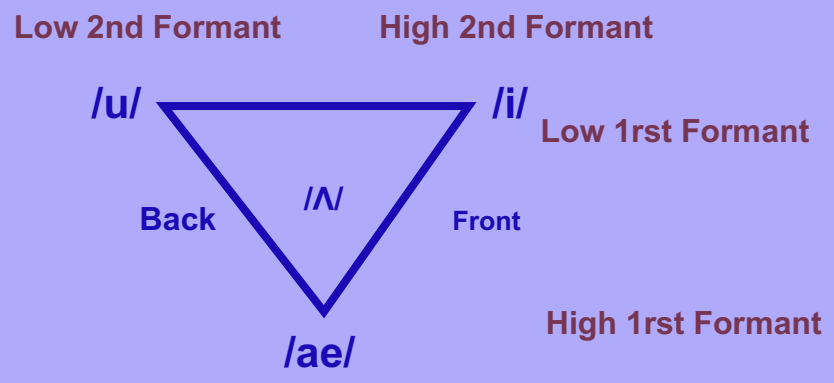
"So is Carlsberg."

Conversational Clear



- Longer, more intense consonants (improved C/V)

Vowel Triangle



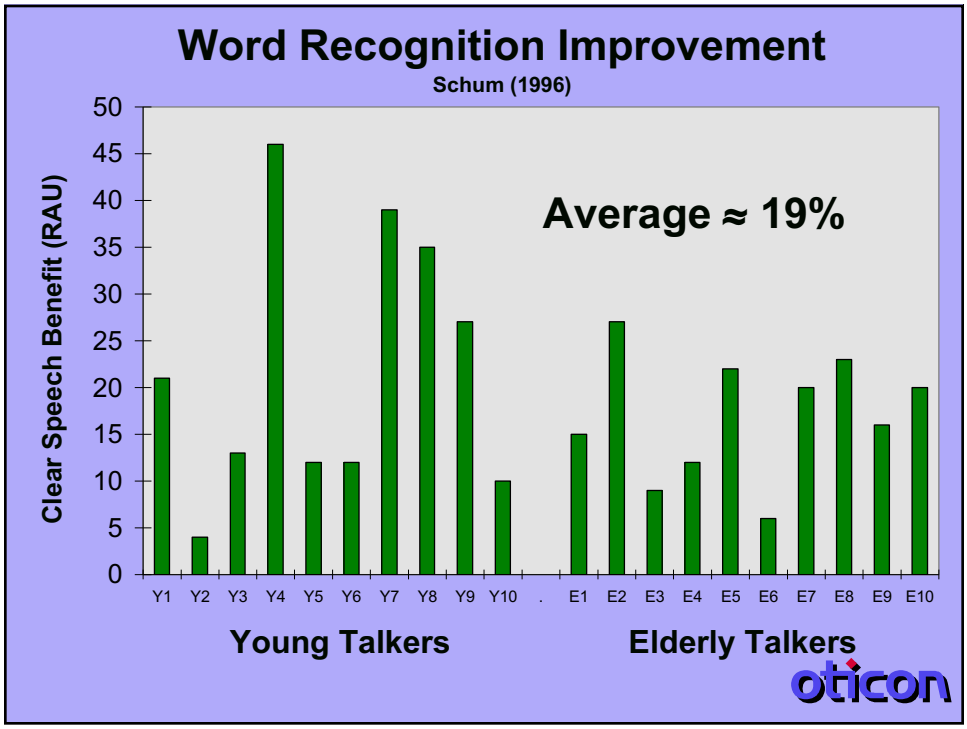
oticon

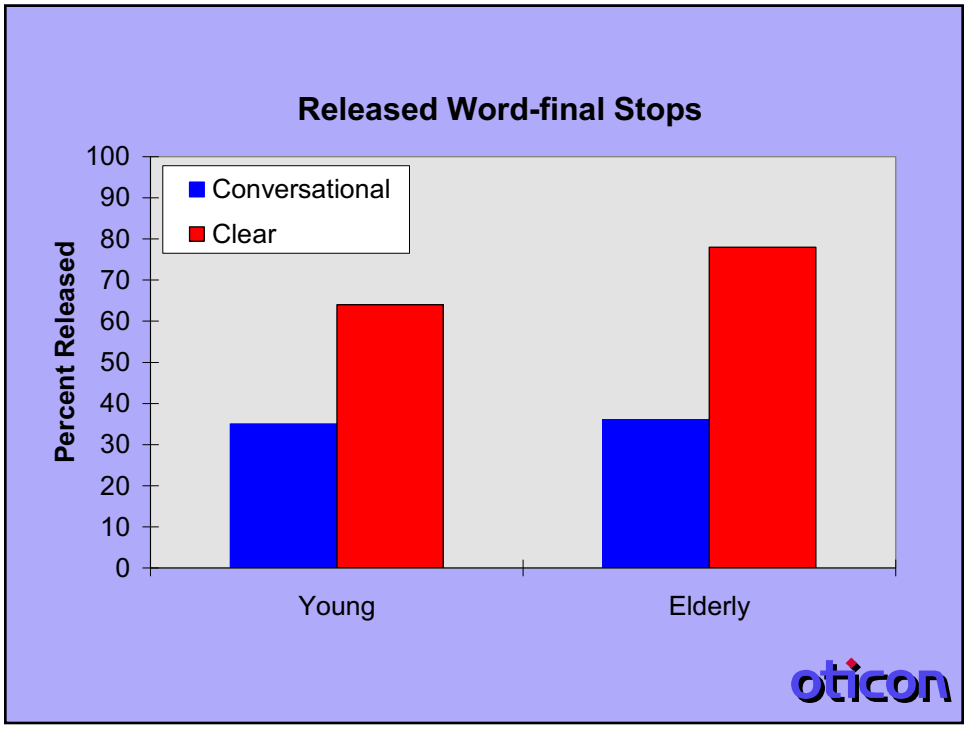
Effect on Intelligibility?

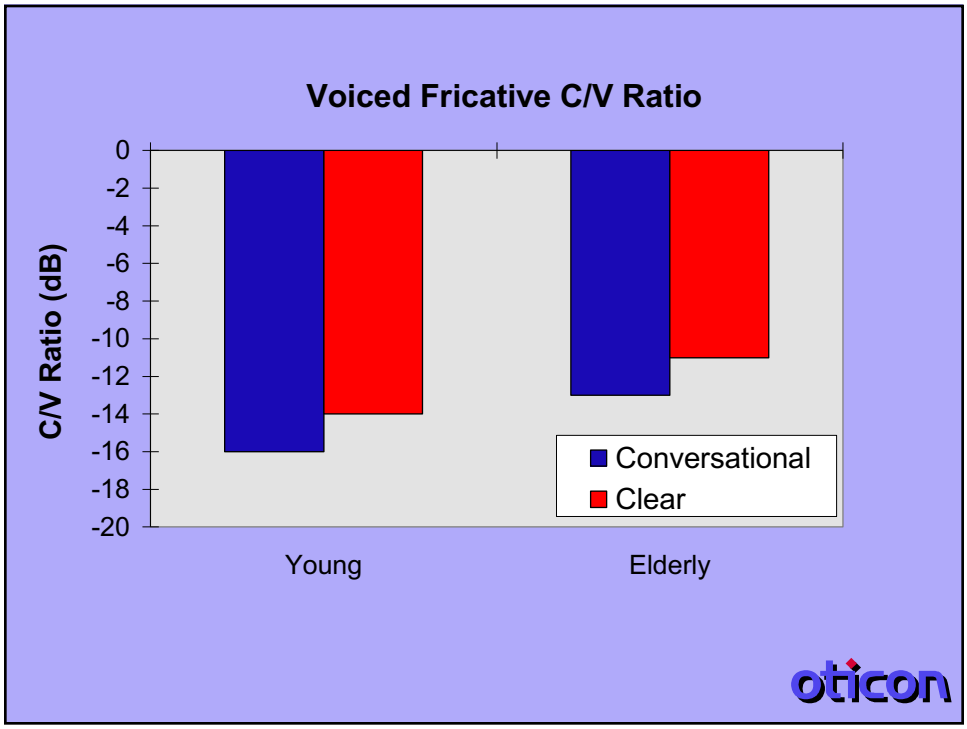
At least 15 to 20% in word recognition

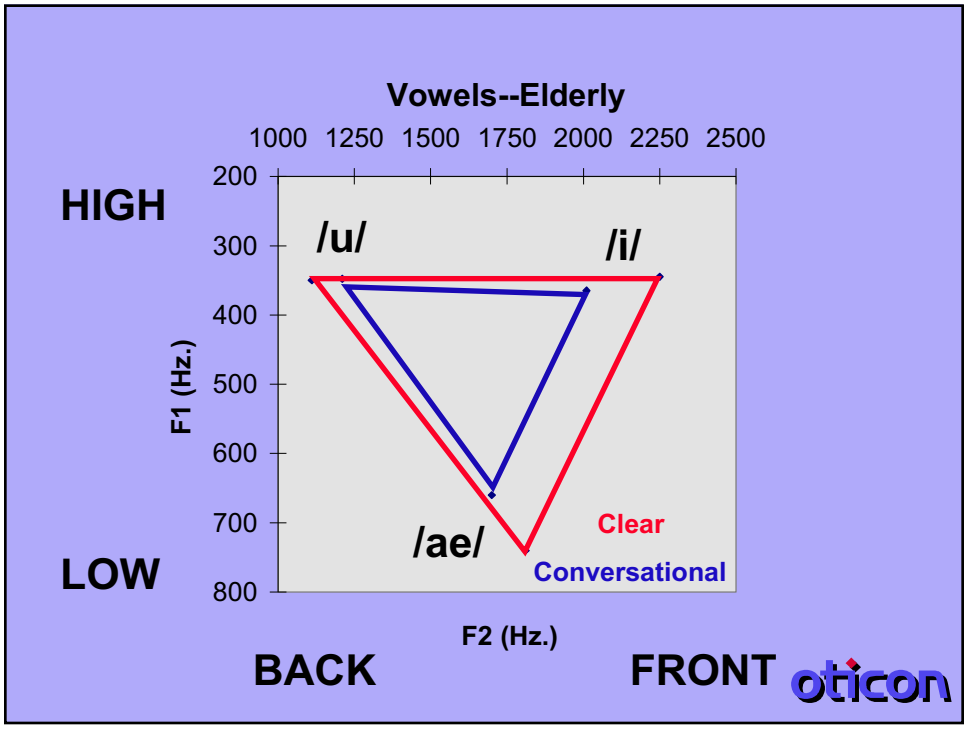
Schum (1996) Study

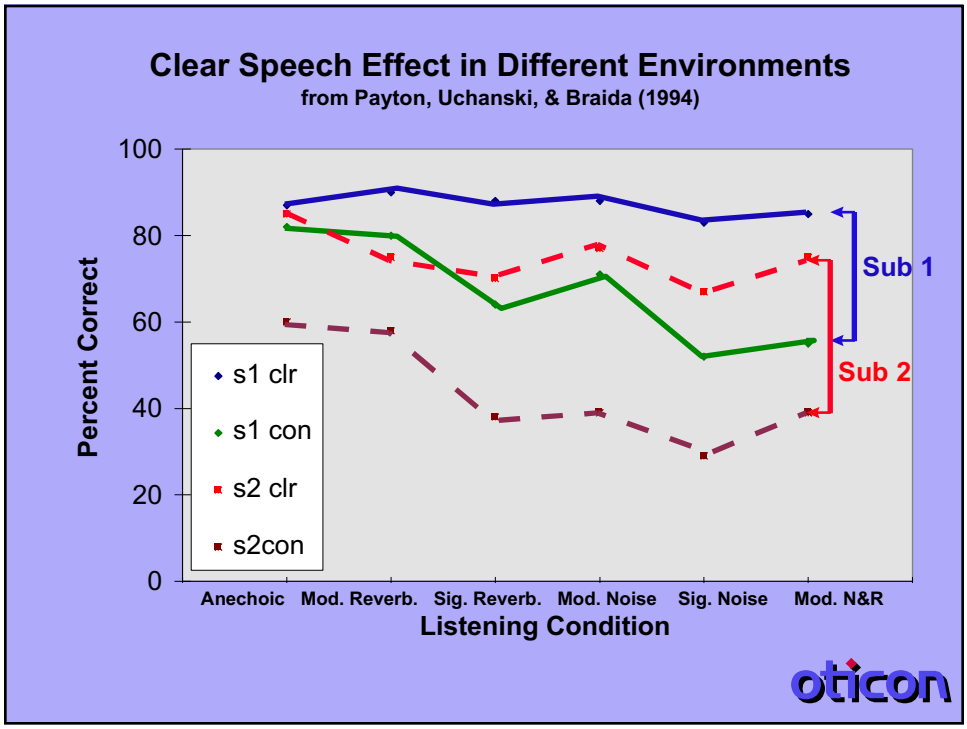
- Elderly & young, inexperienced talkers
- 5-10 minutes of training and practice
- Materials played back in noise to patients with SNHL (mild-mod sloping)




Released Word-final Stops
Percent Released
100
90
80
70
60
50
40
30
20
10
0
Conversational
Clear
Young
Elderly
oticon


Voiced Fricative C/V Ratio
C/V Ratio (dB)
0
-2
-4
-6
-8
-10
-12
-14
-16
-18
-20
Conversational
Clear
Young
Elderly
oticon


Vowels--Elderly
1000 1250 1500 1750 2000 2250 2500
200
300
400
500
600
700
800
F1 (Hz.)
HIGH
LOW
/u/
/i/
/ae/
Clear
Conversational
F2 (Hz.)
BACK
FRONT
oticon


Clear Speech Effect in Different Environments
from Payton, Uchanski, & Braida (1994)
100
80
60
40
20
0
Percent Correct
s1 clr
s1 con
s2 clr
s2con
Sub 1
Sub 2
Anechoic
Mod. Reverb.
Sig. Reverb.
Mod. Noise
Sig. Noise
Mod. N&R
Listening Condition
oticon

Agenda

- The Nature of Speech Understanding
- The Acoustics of Speech in Conversation
- Naturally Produced Enhanced Speech
- Requirements for Computer Enhancement
- Some Past Examples
- Looking Forward

Automatic Speech Recognition

Accurate
Automatic Speech Recognition

Robust
Automatic Speech Recognition

Automatic Speech Recognition:

How good is it?

How good is it?

0 dB S/N +5 dB S/N +10 dB S/N +15 dB S/N +20 dB S/N

+5 dB S/N Babble

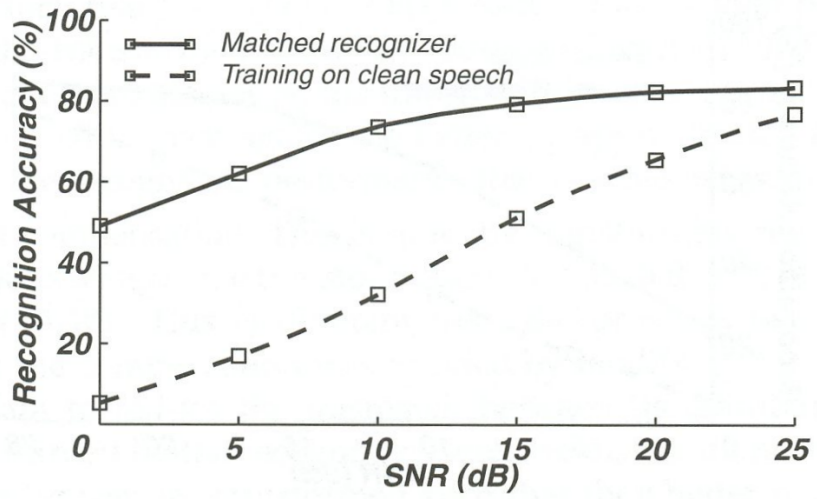
1. The boy fell from the window
2. The wife helped her husband.
3. Big dogs can be dangerous.
4. Her shoes were very dirty.
5. The player lost a shoe.
6. Somebody stole the money.
7. The fire was very hot.
8. She's drinking from her own cup.
9. The picture came from a book.
10. The car was going too fast.

The boy know from
the wife helped her husband
and gods in the game
or shoes.
Her player lost issue
somebody still for
my player was very
he's drinking from her own
secure gain from
the car was going to that

Sentences in Quiet

1. The boy fell from the window
2. The wife helped her husband.
3. Big dogs can be dangerous.
4. Her shoes were very dirty.
5. The player lost a shoe.
6. Somebody stole the money.
7. The fire was very hot.
8. She's drinking from her own cup.
9. The picture came from a book.
10. The car was going too fast.

To myself from the window
the life helped her husband
big dogs can be dangerous
her shoes were very dirty
the player lost a shoe
somebody stole the money
the fire was very hot
cheese drinking from her own cop
the picture came from a book
the car was going too fast



Davis, 2002

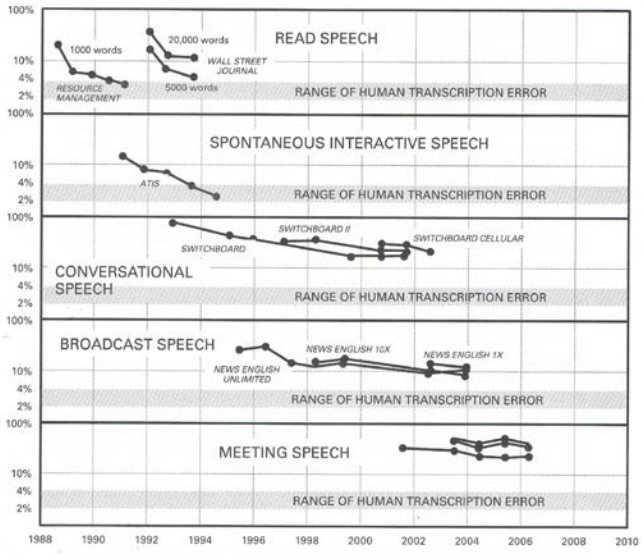


Figure 6.5
Evolution of the word error rate through the years in various projects evaluated by NIST.

PIERACCINI, 2012

Real Time
Automatic Speech Recognition

Multi-dimensional
Enhancement Scheme

Natural Clear Speech

- Individual correlations between 15 x 3 acoustic variables and Clear Speech Performance benefit for 20 talkers
- Best single correlation .45
- Most <.2
- …
- Stepwise analysis drove correlations to >.9
- . . .but needed 9 or 10 dimensions

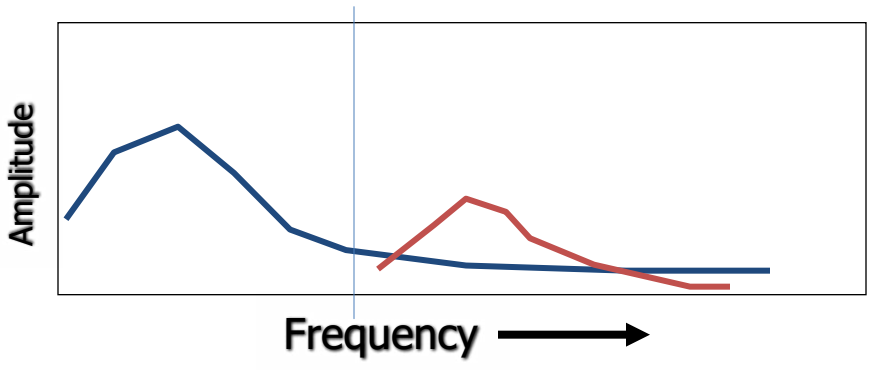
Naturalness

(“The Avatar Effect”)

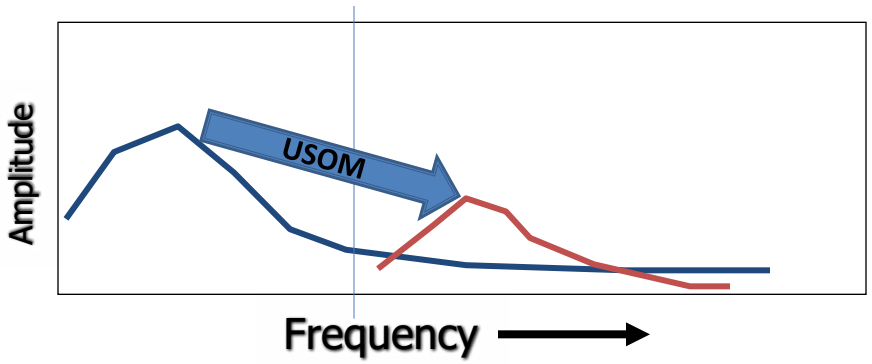
Agenda

- The Nature of Speech Understanding
- The Acoustics of Speech in Conversation
- Naturally Produced Enhanced Speech
- Requirements for Computer Enhancement
- Some Past Examples
- Looking Forward

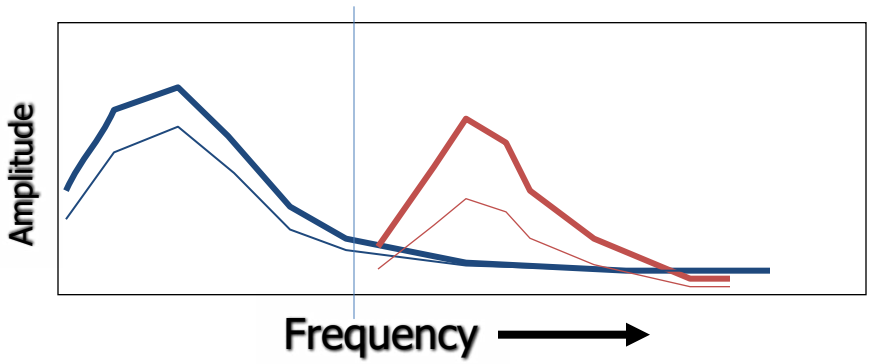
Multichannel Compression



Multichannel Compression



Multichannel Compression



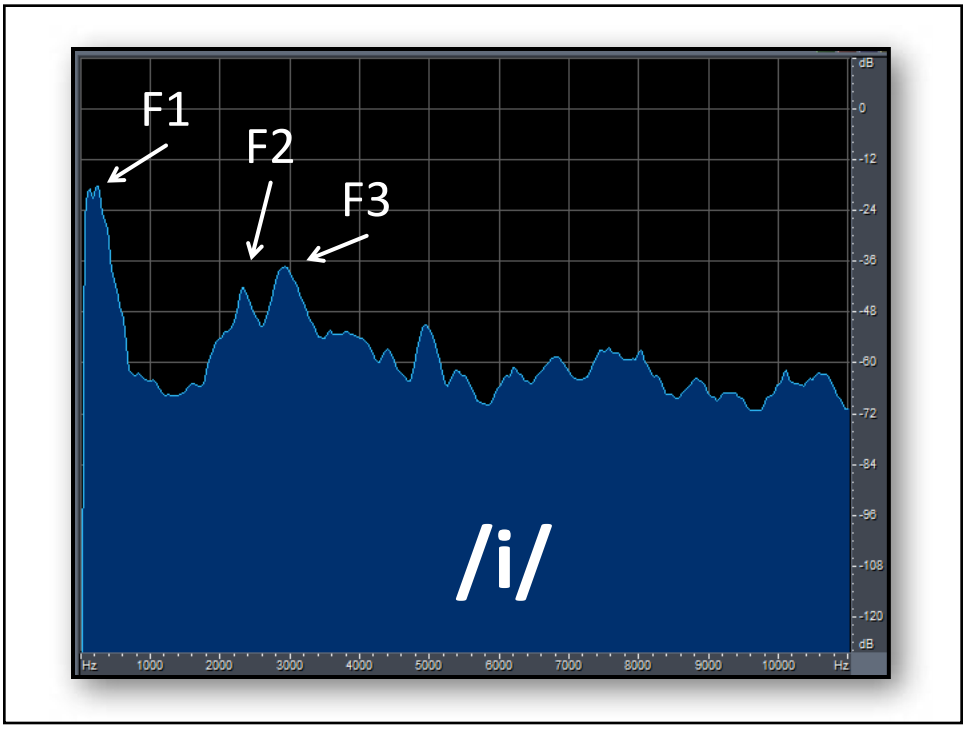
Clear Speech as a Signal Processing Technique

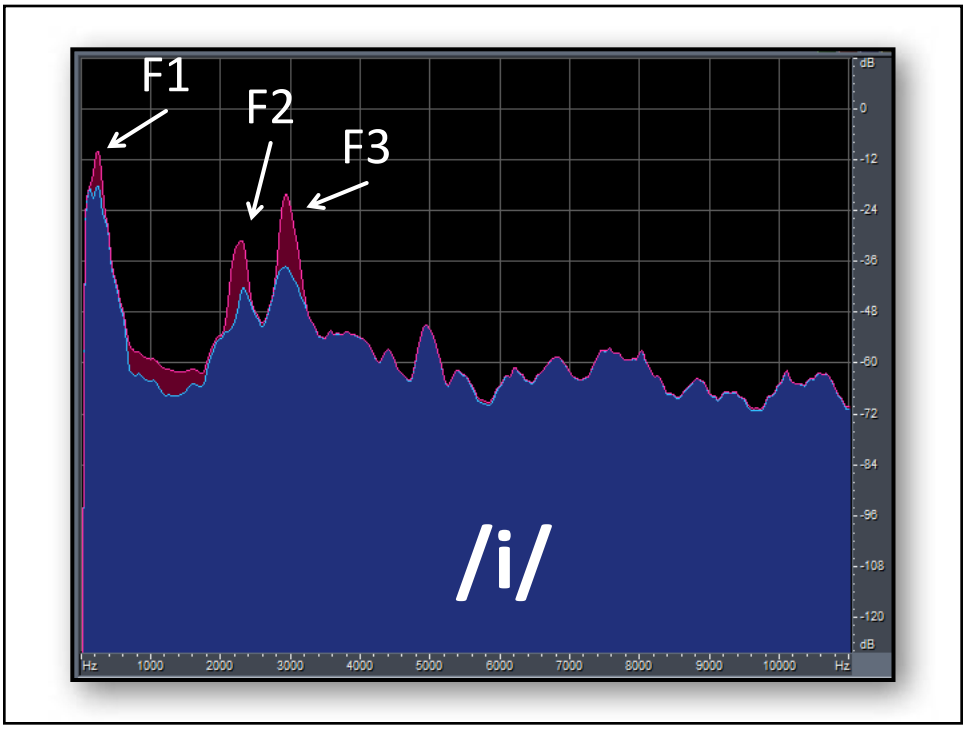
Can a computer based hearing instrument be taught to produce Clear Speech?

Can Clear Speech by Created by a DSP based Hearing Instrument?

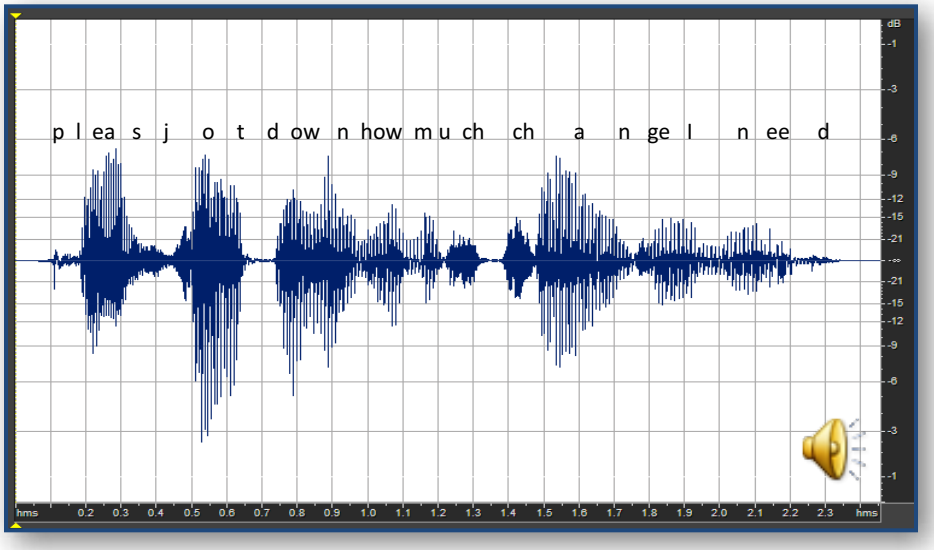
- Naturally produced Clear Speech is an extremely complex combination of acoustic changes
  - Not just. . .increased C/V ratio
  - . . .longer consonant durations
  - . . .more pauses
  - . . .changes in intonation contours (linguistically meaningful)
- DSP approach unlikely to mimic this effect
  - Finding/manipulating phonemes is difficult
  - Time marches on
  - Manipulating one or two dimensions is not enough

Phoneme Enhancement

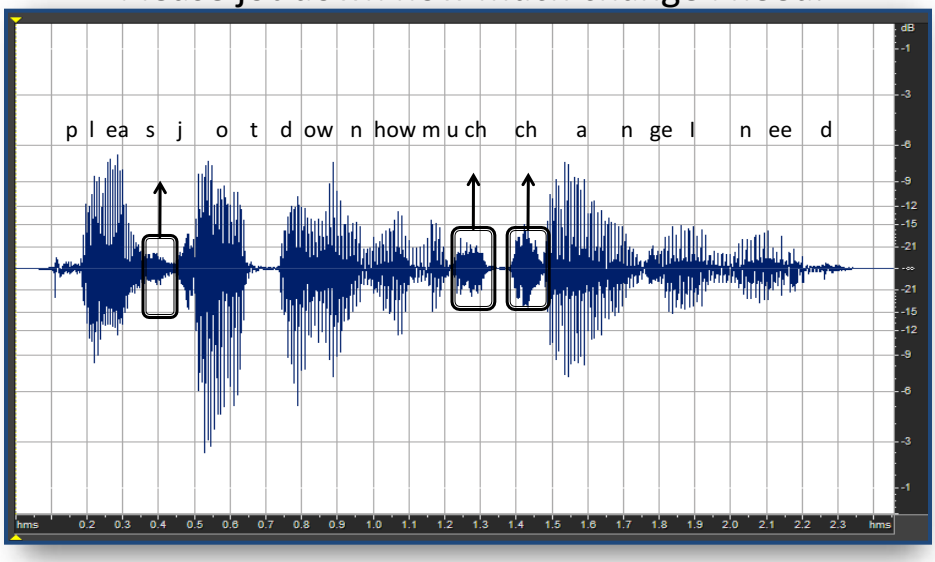

F1
F2
F3
/i/
dB
0
-12
-24
-36
-48
-60
-72
-84
-96
-108
-120
dB
Hz
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
Hz


F1
F2
F3
/i/
dB
0
-12
-24
-36
-48
-60
-72
-84
-96
-108
-120
dB
Hz
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
Hz

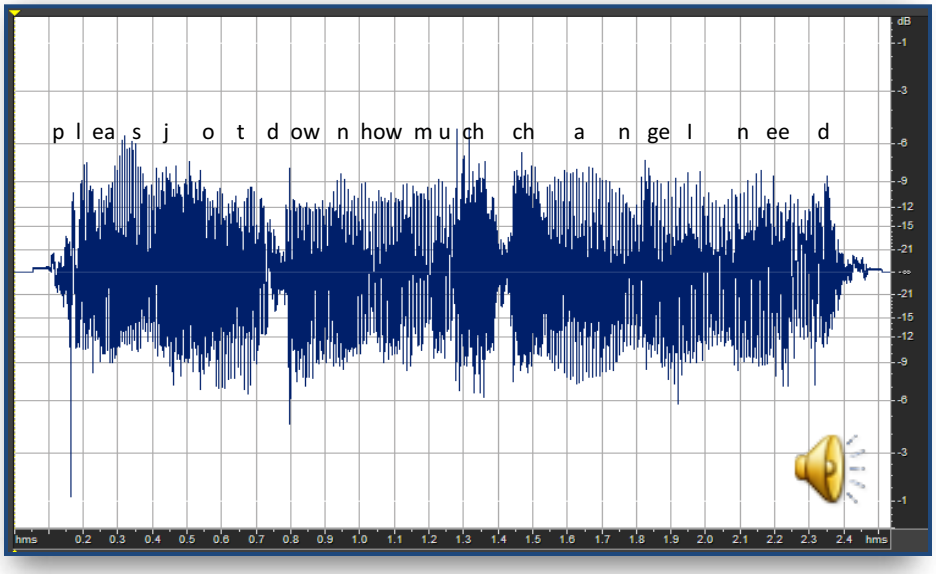
Please jot down how much change I need.



Please jot down how much change I need.

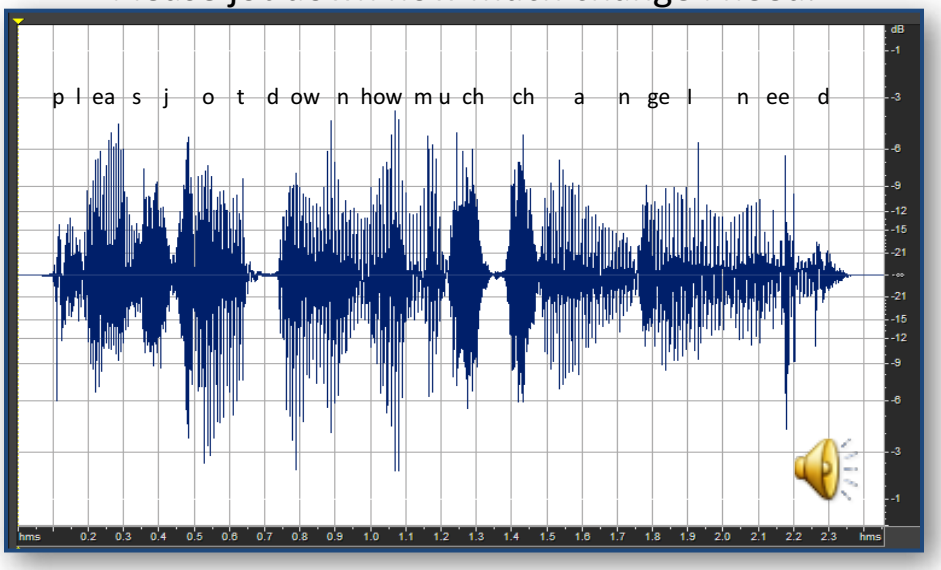


Please jot down how much change I need.



Actual effect of Fast Acting Compression

Please jot down how much change I need.



Agenda

- The Nature of Speech Understanding
- The Acoustics of Speech in Conversation
- Naturally Produced Enhanced Speech
- Requirements for Computer Enhancement
- Some Past Examples
- Looking Forward

Making Speech More Distinct

Donald J. Schum, PhD
Vice President, Audiology & Professional Relations
Oticon, Inc.

DJS@Oticonusa.com