

FROM TEXT TO SPEECH THE MITALK SYSTEM

Presented by

나집

2013-6-14

MEASURES OF INTELLIGIBILITY AND COMPREHENSION

- Different tests were conducted to provide information about
 - Phoneme Recognition
 - Word recognition in sentences
 - Listening comprehension
- In carrying out these tests a total of 27,128 responses were collected from 160 naïve listeners

PHONEME RECOGNITION

- ◉ Modified Rhyme Test was used to measure the intelligibility of the speech produced by the system
- ◉ Six lists of 50 monosyllabic words were prepared on MIT TTS system
- ◉ Approximately half of the items in a given test list differed in the initial consonant while the remaining half differed in the final consonant
- ◉ Forced choice response forms were provided to the subjects to record their judgements

Sample test trials from the Modified Rhyme Test

1.	a) bad	b) back	c) ban	d) bass	e) bat	f) bath
2.	a) beam	b) bead	c) beach	d) beat	e) beak	f) bean
3.	a) bus	b) but	c) bug	d) buff	e) bun	f) buck
4.	a) case	b) cave	c) cape	d) cane	e) cake	f) came
5.	a) cuff	b) cut	c) cuss	d) cub	e) cup	f) cud
6.	a) dip	b) din	c) dill	d) dig	e) dim	f) did
7.	a) dub	b) dun	c) dung	d) dug	e) duck	f) dud
8.	a) fizz	b) fin	c) fill	d) fig	e) fib	f) fit
9.	a) hear	b) heath	c) heal	d) heave	e) heat	f) heap
10.	a) kid	b) kit	c) kill	d) kin	e) king	f) kick
11.	a) lace	b) lame	c) lane	d) lay	e) lake	f) late
12.	a) man	b) math	c) mad	d) mat	e) mass	f) map
13.	a) pace	b) pane	c) pave	d) page	e) pay	f) pale
14.	a) path	b) pat	c) pack	d) pad	e) pass	f) pan
15.	a) peas	b) peak	c) peal	d) peace	e) peach	f) peat
16.	a) pip	b) pick	c) pin	d) pill	e) pit	f) pig
17.	a) puff	b) pus	c) pub	d) pun	e) puck	f) pup
18.	a) rate	b) race	c) ray	d) raze	e) rave	f) rake
19.	a) safe	b) sake	c) same	d) sane	e) save	f) sale
20.	a) sat	b) sag	c) sack	d) sap	e) sass	f) sad
21.	a) seed	b) seek	c) seen	d) seep	e) seem	f) seethe
22.	a) sill	b) sick	c) sing	d) sit	e) sin	f) sip
23.	a) sup	b) sud	c) sun	d) sum	e) sub	f) sung
24.	a) tap	b) tang	c) tam	d) tan	e) tab	f) tack
25.	a) tease	b) tear	c) teak	d) teal	e) team	f) teach

PHONEME RECOGNITION

- Although the modified rhyme test employed real words, our interest was focused on the phoneme errors and the resulting perceptual confusions
- Overall performance on the test was very good with a total of only 6.9 percent

PHONEME RECOGNITION

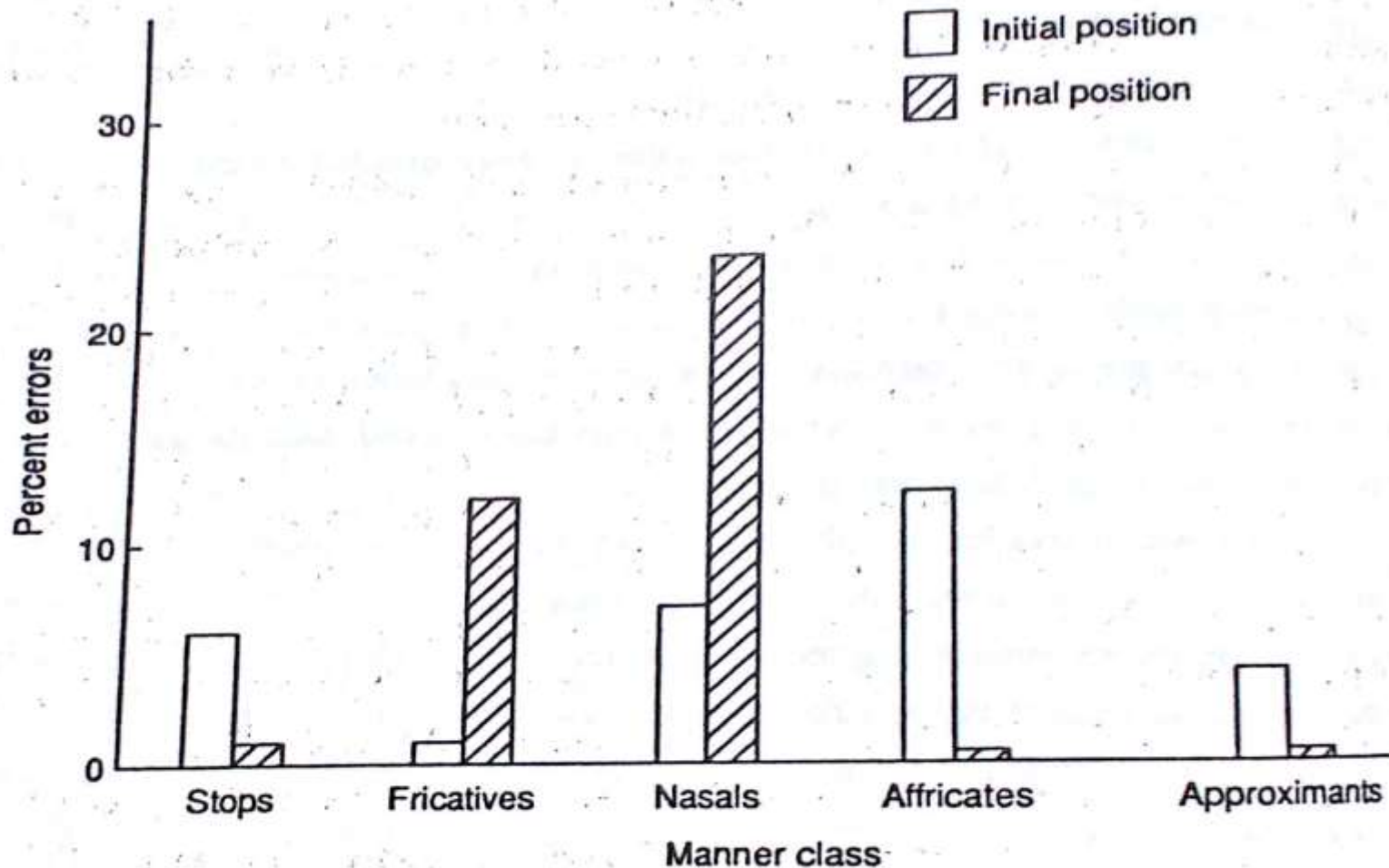


Figure 13-1: Average percent errors across various manner classes

Phoneme presented

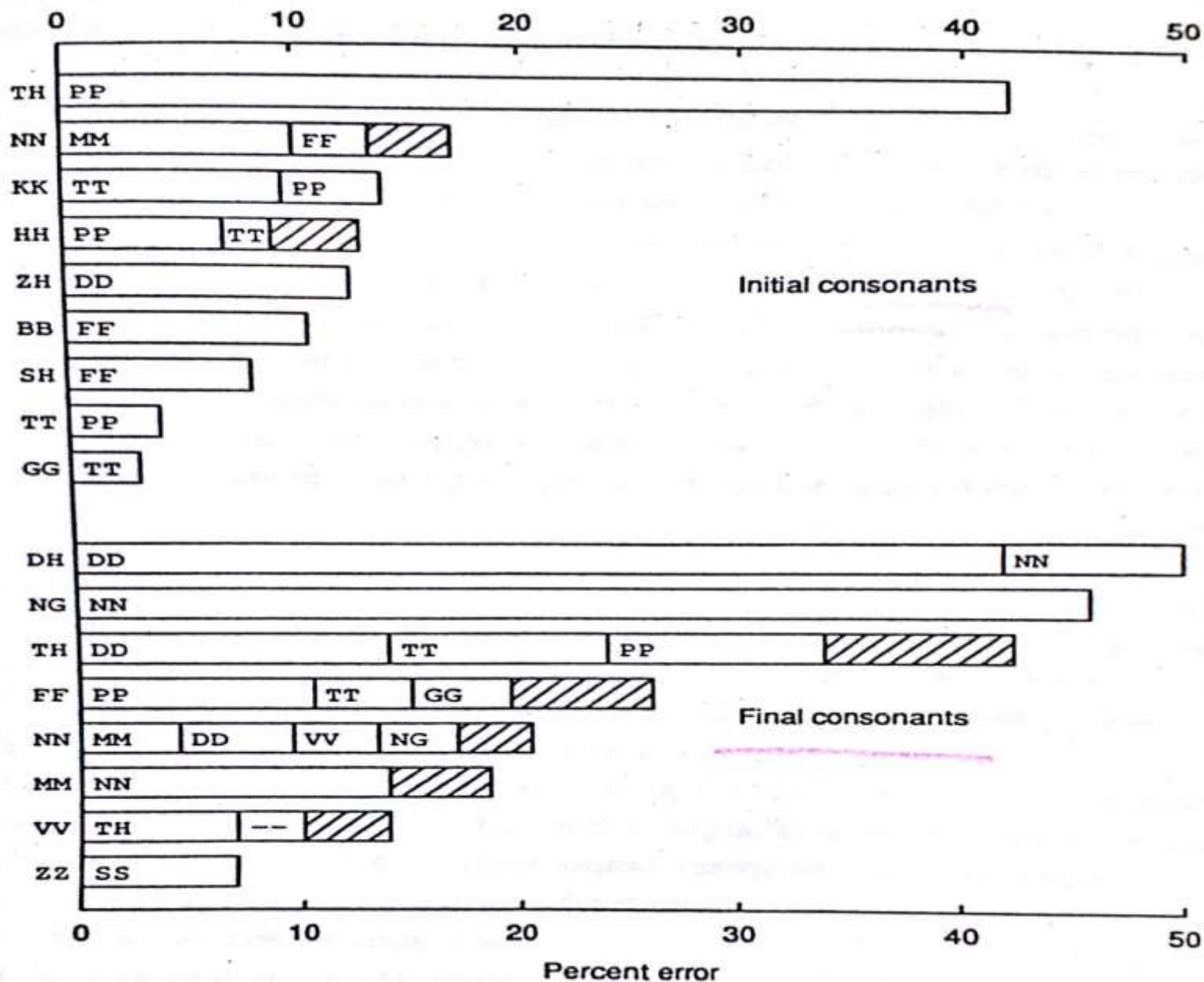


Figure 13-2: Distribution of errors and most frequent perceptual confusions

CONTD...

- ◉ The advantage of initial over final consonants observed in the present study is consistent with data obtained from natural speech by House et al
- ◉ With this high phoneme recognition performance, it is difficult to pick up subtle details of the error patterns that might be useful in improving the quality of the phonetic component of the system at present time

WORD RECOGNITION IN SENTENCES

WORD RECOGNITION IN SENTENCES

- ◉ To evaluate word recognition in sentence context, two quite different sets of data was obtained

WORD RECOGNITION IN SENTENCES

- ◉ To evaluate word recognition in sentence context, two quite different sets of data was obtained
 - Harvard Psychoacoustic Sentences: these sentences are all meaning full and contain a wide range of syntactic constructions

WORD RECOGNITION IN SENTENCES

- ◉ To evaluate word recognition in sentence context, two quite different sets of data was obtained
 - Harvard Psychoacoustic Sentences: these sentences are all meaning full and contain a wide range of syntactic constructions
 - Haskins Anomalous Sentences: these are meaningless and all have the same syntactic structure

Sample test materials from the Harvard Psychoacoustic Sentences

1. The birch canoe slid on the smooth planks
2. Glue the sheet to the dark blue background
3. It's easy to tell the depth of a well
4. These days a chicken leg is a rare dish
5. Rice is often served in round bowls
6. The juice of lemons makes fine punch
7. The box was thrown beside the parked truck
8. The hogs were fed chopped corn and garbage
9. Four hours of steady work faced us
10. A large size in stockings is hard to sell
11. The boy was there when the sun rose
12. A rod is used to catch pink salmon
13. The source of the huge river is the clear spring
14. Kick the ball straight and follow through
15. Help the woman get back to her feet
16. A pot of tea helps to pass the evening
17. Smoky fires lack flame and heat
18. The soft cushion broke the man's fall
19. The salt breeze came across from the sea
20. The girl at the booth sold fifty bonds
21. The small pup gnawed a hole in the sock
22. The fish twisted and turned on the bent hook
23. Press the pants and sew a button on the vest
24. The swan dive was far short of perfect
25. The beauty of the view stunned the young boy

Sample test materials from the Haskins Anomalous Sentences

1. The wrong shot led the farm
2. The black top ran the spring
3. The great car met the milk
4. The old corn cost the blood
5. The short arm sent the cow
6. The low walk read the hat
7. The rich paint said the land
8. The big bank felt the bag
9. The sick seat grew the chain
10. The salt dog caused the show
11. The last fire tried the nose
12. The young voice saw the rose
13. The gold rain led the wing
14. The chance sun laid the year
15. The white bow had the bed
16. The near stone thought the ear
17. The end home held the press
18. The deep head cut the cent
19. The next wind sold the room
20. The full leg shut the shore
21. The safe meat caught the shade
22. The fine lip tired the earth
23. The plain can lost the men
24. The dead hand armed the bird
25. The fast point laid the word

CONTD...

- ◉ We would anticipate lower levels of word recognition performance on Haskins Anomalous sentences than on Harvard Sentences
- ◉ Because in Harvard sentences the syntactic and semantic context is readily available and can be used freely by the listener
- ◉ Subjects were told to write down each sentence as they heard it in the appropriate location on their response sheets

CONTD...

- ⦿ Performance on the Harvard sentences was quite good with an overall mean of 93.2 percent correct word recognition
- ⦿ Word recognition performance on Haskins sentences was 78.7 percent correct
- ⦿ Comparisons of word recognition performance in the first and second half of the tests indicated the presence of a reliable learning effect with an improved performance of 2%

COMPREHENSION

- Comprehension is a complex cognitive process, initially involving the input and subsequent encoding of sensory information, the retrieval of previously stored knowledge from long term memory and the subsequent interpretation, integration or assimilation of various knowledge sources that may be available to a listener at the time
- One of the factors that obviously plays an important role in listening comprehension is the quality of the input signal expressed in terms of overall intelligibility

COMPREHENSION

- Fifteen narrative passages and an appropriate set of test question were selected from standardized adult reading comprehension tests
- A parallel set of this data was also used for reading comprehension

COMPREHENSION

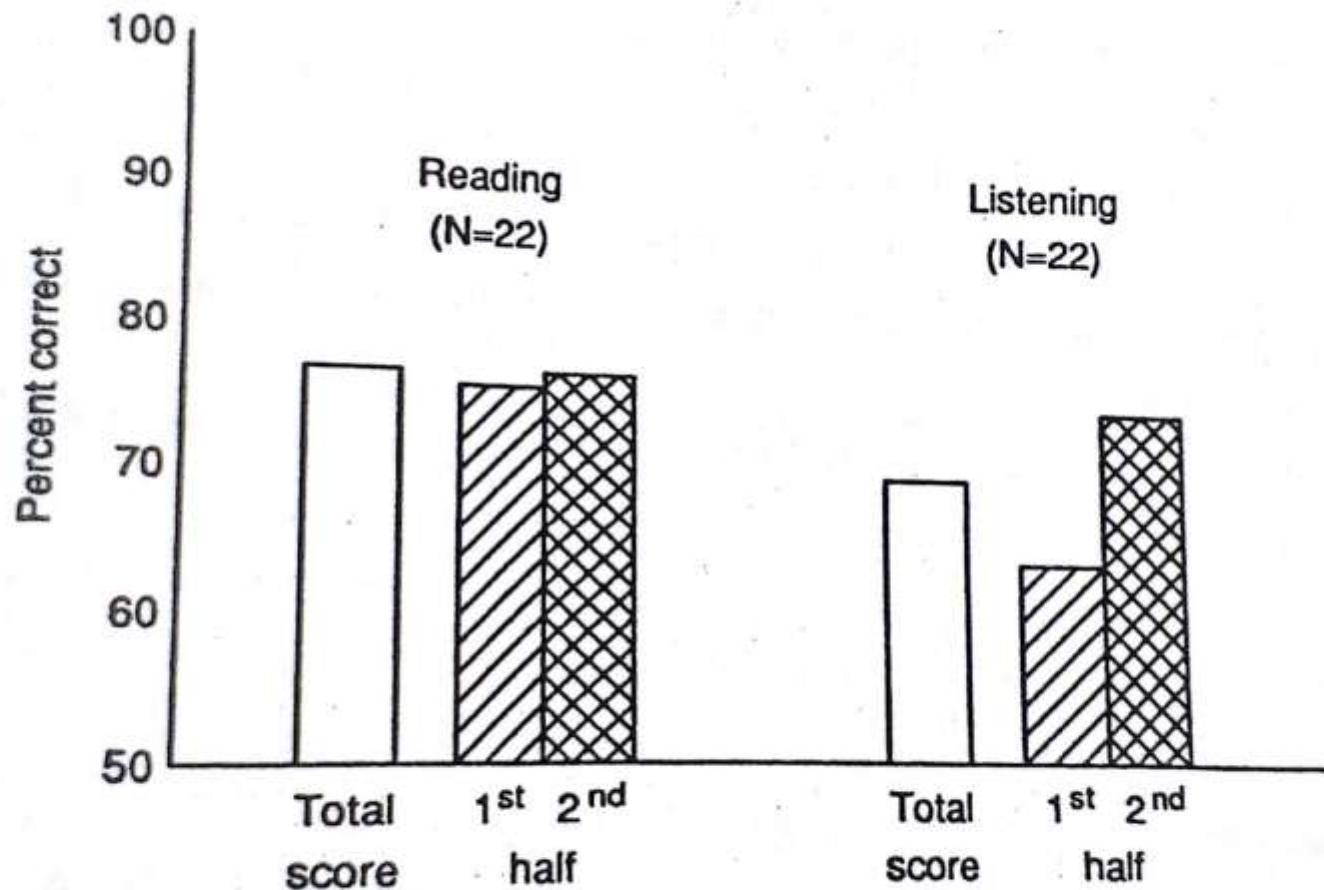


Figure 13-3: Percent correct comprehension scores for reading and listening groups

GENERAL DISCUSSION

- ◉ The results of the three tests designed to evaluate intelligibility, word recognition and listening comprehension indicated very high levels of performance
- ◉ The authors were able to identify a number of errors in the operation of the system which ordinarily might not have been detected
- ◉ Errors consisted of incorrect parsings, pronunciations or stress assignments.
- ◉ Error located at one module often affected analyses carried out by other modules

GENERAL DISCUSSION

- ◉ Two errors in the first module FORMAT were detected
 - U.S. (space between U. and S.)
 - 19th (“one nine-T-H”)
- ◉ A number of errors were also detected in the module DECOMP which decomposes the words into morphs. Several words were identified that should have been in the lexicon since their pronunciation cannot be handled correctly by spelling to sound rules
- ◉ For example “tain” in “britain” was pronounced as that of “maintain”

GENERAL DISCUSSION

- ◉ POS errors were also detected for example “close” can be a verb or adjective each with a different pronunciation
- ◉ The errors described above are considered to be minor and can be corrected easily
- ◉ There were no errors detected which escaped explanation

GENERAL DISCUSSION

- ◉ There are several limitations to our present evaluation
- ◉ We did not carry out any of the control conditions for the three types of tests using natural speech
- ◉ The modified rhyme test was too easier for listeners, additional open response tests should be employed
- ◉ The comprehension test was limited in ability to distinguish between new knowledge acquired from listening to text and the previous knowledge of the listener

GENERAL DISCUSSION

- ◎ The present results although preliminary support the general conclusion that very high quality synthetic speech can be produced automatically from unrestricted English text and that such a system could be implemented in an applied setting.

IMPEMENTATION

- ◉ Conceptual design

- ◉ Implementation History

- BPCL + FORTRAN → UNIX + PASCAL+ C

IMPLEMENTATION

- ◉ Conceptual design

Synthesis

Timing

Fundamental
Frequency

Phonetic
Targets

Continuation
Smoothing

Parameter
Conversion

Waveform
Generation

- ◉ Implementation History

- BPCL + FORTRAN → UNIX + PASCAL+ C

IMPLEMENTATION

◉ Conceptual design

Analysis

Symbols to
Standard
Form

Phonetic
Transcription

Parsing

Lexical
Stress

Phonological
Recoding

Semantic
Analysis

Synthesis

Timing

Fundamental
Frequency

Phonetic
Targets

Continuation
Smoothing

Parameter
Conversion

Waveform
Generation

◉ Implementation History

- BPCL + FORTRAN → UNIX + PASCAL + C

