

# REVIEW OF TEXT- TO-SPEECH CONVERSION FOR ENGLISH

Presented by

나집 칸

2013-3-29

# INTRODUCTION

# INTRODUCTION

Trace the history of progress toward the development of systems for converting text to speech.

# INTRODUCTION

Trace the history of progress toward the development of systems for converting text to speech.

Last Presentation

# INTRODUCTION

Trace the history of progress toward the development of systems for converting text to speech.

## Last Presentation

- Linear prediction: There are limitations on its applicability to general text synthesis.

# INTRODUCTION

Trace the history of progress toward the development of systems for converting text to speech.

## Last Presentation

- Linear prediction: There are limitations on its applicability to general text synthesis.
- Articulatory Model: Computational costs and lack of data upon which to base rules prevent immediate application of this approach.

# INTRODUCTION

Trace the history of progress toward the development of systems for converting text to speech.

## Last Presentation

- Linear prediction: There are limitations on its applicability to general text synthesis.
- Articulatory Model: Computational costs and lack of data upon which to base rules prevent immediate application of this approach.
- Acoustic properties of phonetic Segments

# PROSODY AND SENTENCE PHONETIC RECODING



# PROSODY AND SENTENCE PHONETIC RECODING

- ⦿ A sentence cannot be synthesized by simply stringing together a sequence of phonemes or words.

# PROSODY AND SENTENCE PHONETIC RECODING

- ⦿ A sentence cannot be synthesized by simply stringing together a sequence of phonemes or words.
- ⦿ For intelligible and moderately natural

# PROSODY AND SENTENCE PHONETIC RECODING

- A sentence cannot be synthesized by simply stringing together a sequence of phonemes or words.
- For intelligible and moderately natural
  - Intensity

# PROSODY AND SENTENCE PHONETIC RECODING

- ⦿ A sentence cannot be synthesized by simply stringing together a sequence of phonemes or words.
- ⦿ For intelligible and moderately natural
  - Intensity
  - Duration

# PROSODY AND SENTENCE PHONETIC RECODING

- A sentence cannot be synthesized by simply stringing together a sequence of phonemes or words.
- For intelligible and moderately natural
  - Intensity
  - Duration
  - Pitch

# PROSODY AND SENTENCE PHONETIC RECODING

- ⊙ A sentence cannot be synthesized by simply stringing together a sequence of phonemes or words.
- ⊙ For intelligible and moderately natural
  - Intensity
  - Duration
  - Pitch
  - Allophonic Detail

# INTENSITY RULES

# INTENSITY RULES

- ⦿ Stress syllables may be more intense by a few dB.



# INTENSITY RULES

- ◉ Stress syllables may be more intense by a few dB.
- ◉ Intensity is not a very effective perceptual cue to stress due to variations associated with vowel height,  $f_0$  and laryngeal state.

# INTENSITY RULES

- ⊙ Stress syllables may be more intense by a few dB.
- ⊙ Intensity is not a very effective perceptual cue to stress due to variations associated with vowel height,  $f_0$  and laryngeal state.
- ⊙ One can achieve stress related increase by rules that only manipulate  $f_0$ .

# INTENSITY RULES

- ◉ Stress syllables may be more intense by a few dB.
- ◉ Intensity is not a very effective perceptual cue to stress due to variations associated with vowel height,  $f_0$  and laryngeal state.
- ◉ One can achieve stress related increase by rules that only manipulate  $f_0$ .
- ◉ At a phrase level, it appears that syllables at the end of an utterance can become weaker in intensity.

# DURATION RULES...

# DURATION RULES...

- ◉ The syntactic structure of the sentence to be produced determines the locations of prosodic boundaries at which segments are longer in duration.

# DURATION RULES...

- ◉ The syntactic structure of the sentence to be produced determines the locations of prosodic boundaries at which segments are longer in duration.
- ◉ The lexicon or stress rules determine which consonants and vowels are stressed and hence longer in duration.

# DURATION RULES...

- ◉ The syntactic structure of the sentence to be produced determines the locations of prosodic boundaries at which segments are longer in duration.
- ◉ The lexicon or stress rules determine which consonants and vowels are stressed and hence longer in duration.
- ◉ The phonological component select the allophones and select their durations according to phonetic context.

# KLATT MODEL



# KLATT MODEL

- The Klatt model assumes that

# KLATT MODEL

- The Klatt model assumes that
  - Each phonetic segment type has an inherent duration.

# KLATT MODEL

- The Klatt model assumes that
  - Each phonetic segment type has an inherent duration.
  - Each rule tries to effect a percentage increase or decrease in the duration of the segment

# KLATT MODEL

- The Klatt model assumes that
  - Each phonetic segment type has an inherent duration.
  - Each rule tries to effect a percentage increase or decrease in the duration of the segment
  - The segments cannot be compressed shorter than a certain minimum duration

# KLATT MODEL

- The Klatt model assumes that
  - Each phonetic segment type has an inherent duration.
  - Each rule tries to effect a percentage increase or decrease in the duration of the segment
  - The segments cannot be compressed shorter than a certain minimum duration

$$DUR = \frac{(INH DUR - MIN DUR) \times PRCNT}{100}$$

TABLE II. Duration rules proposed by Klatt (1979a).

1. **PAUSE INSERTION RULE:** Insert a brief pause before each sentence-internal main clause and at other boundaries delimited by an orthographic comma (Goldman-Eisler, 1968; Cooper *et al.*, 1978).
2. **CLAUSE-FINAL LENGTHENING:** The vowel or syllabic consonant in the syllable just before a pause is lengthened (Gaitenby, 1965). Any consonants in the rhyme (between this vowel and the pause) are also lengthened (Oller, 1973; Klatt, 1975a).
3. **PHRASE-FINAL LENGTHENING:** Syllabic segments (vowels and syllabic consonants) are lengthened if in a phrase-final syllable (Klatt, 1975a). Durational increases at the noun-phrase/verb-phrase boundary are more likely in complex noun phrase or when subject-verb-object order is violated; durational changes are much less likely for pronouns (Harris *et al.*, 1981). The lengthening is perceptually important (Lehiste *et al.*, 1976; Umeda and Quinn, 1981).
4. **NON-WORLD-FINAL SHORTENING:** Syllabic segments are shortened slightly if not in a word-final syllable (Oller, 1973). [This rule is disputed by Umeda (1975).]
5. **POLYSYLLABIC SHORTENING:** Syllabic segments in a polysyllabic word are shortened slightly (Lehiste, 1975a). [This rule is also disputed by Umeda (1975).]
6. **NON-INITIAL-CONSONANT SHORTENING:** Consonants in non-word-initial position are shortened (Klatt, 1974; Umeda, 1977).
7. **UNSTRESSED SHORTENING:** Unstressed segments are shorter and more compressible than stressed segments (Fry, 1958; Umeda, 1975, 1977; Lehiste, 1975a).
8. **LENGTHENING FOR EMPHASIS:** An emphasized vowel is significantly lengthened (Bolinger, 1972; Umeda, 1975).
9. **POSTVOCALIC CONTEXT OF VOWELS:** The influence of a postvocalic consonant (in the same word) on the duration of a vowel is such as to shorten the vowel if the consonant is voiceless (House and Fairbanks, 1953; Peterson and Lehiste, 1960). The effects are greatest at phrase and clause boundaries (Klatt, 1975a).
10. **SHORTENING IN CLUSTERS:** Segments are shortened in consonant-consonant sequences (disregarding word boundaries, but not across phrase boundaries) (Klatt, 1973a; Haggard, 1973).
11. **LENGTHENING DUE TO PLOSIVE ASPIRATION:** A stressed vowel or sonorant preceded by a voiceless plosive is lengthened (Peterson and Lehiste, 1960).

# OTHER DURATIONAL RULE SYSTEMS

# OTHER DURATIONAL RULE SYSTEMS

- ◉ Gaitenby:



# OTHER DURATIONAL RULE SYSTEMS

- ◉ Gaitenby:

- Multiple stress levels conditioned by syntactic category (Verbs have less stress than nouns)

# OTHER DURATIONAL RULE SYSTEMS

## ◉ Gaitenby:

- Multiple stress levels conditioned by syntactic category (Verbs have less stress than nouns)
- Conditioned by word frequency (common words and words that are repeated in a discourse are reduced in stress)

# OTHER DURATIONAL RULE SYSTEMS

- ◉ Gaitenby:

- Multiple stress levels conditioned by syntactic category (Verbs have less stress than nouns)
- Conditioned by word frequency (common words and words that are repeated in a discourse are reduced in stress)

- ◉ Klatt rules capture durational differences between nouns and verbs by phrase final lengthening and de-stressing of common verbs

# FUNDAMENTAL FREQUENCY RULES

# FUNDAMENTAL FREQUENCY RULES

- ◉ Intonation

# FUNDAMENTAL FREQUENCY RULES

## ◉ Intonation

- The pitch pattern over time that distinguishes statements from questions or imperative.

# FUNDAMENTAL FREQUENCY RULES

- ◉ Intonation

- The pitch pattern over time that distinguishes statements from questions or imperative.

- ◉  $F_0$  changes affect stress judgments significantly

# FUNDAMENTAL FREQUENCY RULES

- ◉ Intonation

- The pitch pattern over time that distinguishes statements from questions or imperative.

- ◉  $F_0$  changes affect stress judgments significantly

- ◉ A rise in  $f_0$  or a fall in  $f_0$  can indicate a stressed syllable.



# FUNDAMENTAL FREQUENCY RULES

- ◉ Intonation

- The pitch pattern over time that distinguishes statements from questions or imperative.
- ◉  $F_0$  changes affect stress judgments significantly
- ◉ A rise in  $f_0$  or a fall in  $f_0$  can indicate a stressed syllable.
- ◉ The  $f_0$  pattern plays a complex role in encoding information for the listener because it not only conveys information about syntactic structure and stress patterns, but it also helps indicate speaker gender, head size, psychological state, and attitude toward what is being spoken.

CONTD...

## CONTD...

- ◎ Pike 1945: English is like a tone language in that four different degrees of stress corresponded to different pitch levels.

## CONTD...

- ◉ Pike 1945: English is like a tone language in that four different degrees of stress corresponded to different pitch levels.
- ◉ Liberman 1967: A given stress level is manifested as a higher pitch at the beginning of a sentence than near the end.

## CONTD...

- ◉ Pike 1945: English is like a tone language in that four different degrees of stress corresponded to different pitch levels.
- ◉ Liberman 1967: A given stress level is manifested as a higher pitch at the beginning of a sentence than near the end.
- ◉ Female speakers use  $f_0$  values about 1.7 times the male value.

CONTD...

# CONTD...

- ◉ Shaughnessy: Modal auxiliaries, negatives, quantifiers and sentential adverbs tend to be emphasized (local  $f_0$  increase) when present in read sentences.

## CONTD...

- ◉ Shaughnessy: Modal auxiliaries, negatives, quantifiers and sentential adverbs tend to be emphasized (local  $f_0$  increase) when present in read sentences.
- ◉ Coker: the strength of  $f_0$  gesture depends on semantic factors that extend over more than one sentence.



## CONTD...

- ◉ Shaughnessy: Modal auxiliaries, negatives, quantifiers and sentential adverbs tend to be emphasized (local  $f_0$  increase) when present in read sentences.
- ◉ Coker: the strength of  $f_0$  gesture depends on semantic factors that extend over more than one sentence.
- ◉ A repeated word is reduced in  $f_0$  gesture, and the reduction is due to semantic recurrence rather than to reappearance of exactly the same item.

CONTD...

# CONTD...

- ◉ In addition to the rule-governed changes to fundamental frequency over a sentence, there are also local perturbations due to aspects of segmental articulation.

# CONTD...

- ◉ In addition to the rule-governed changes to fundamental frequency over a sentence, there are also local perturbations due to aspects of segmental articulation.
- ◉ The  $f_0$  contour is higher near a voiceless consonant than near a voiced consonant, and is higher on a high vowel.

CONTD...

## CONTD...

- ⦿ Rules are needed to predict when  $f_0$  will rise or fall, and what levels it will reach on individual stressed syllables of a sentence as a function of:

# CONTD...

- Rules are needed to predict when  $f_0$  will rise or fall, and what levels it will reach on individual stressed syllables of a sentence as a function of:
  - Syntactic structure

# CONTD...

- ⊙ Rules are needed to predict when  $f_0$  will rise or fall, and what levels it will reach on individual stressed syllables of a sentence as a function of:
  - Syntactic structure
  - Stress pattern



# CONTD...

- Rules are needed to predict when  $f_0$  will rise or fall, and what levels it will reach on individual stressed syllables of a sentence as a function of:
  - Syntactic structure
  - Stress pattern
  - Semantic variables (if known) such as the locations of the most important word in a sentence

## CONTD...

- Rules are needed to predict when  $f_0$  will rise or fall, and what levels it will reach on individual stressed syllables of a sentence as a function of:
  - Syntactic structure
  - Stress pattern
  - Semantic variables (if known) such as the locations of the most important word in a sentence
- Such theories are still in their infancy.

CONTD...

# CONTD...

- One theory is based on the view that  $f_0$  moves slowly from target to target tone.

## CONTD...

- ⊙ One theory is based on the view that  $f_0$  moves slowly from target to target tone.
- ⊙ Another class of theories include commands to raise and lower  $f_0$  at certain times , emphasizing the motions over the actual target achieved.

CONTD...

## CONTD...

- ◉ Mattingly 1966: First algorithm for determination of  $f_0$  contour.

## CONTD...

- ◉ Mattingly 1966: First algorithm for determination of  $f_0$  contour.
- ◉ Mattingly rules recognized three intonational tunes that could be placed on the last prominent syllable of a clause.



## CONTD...

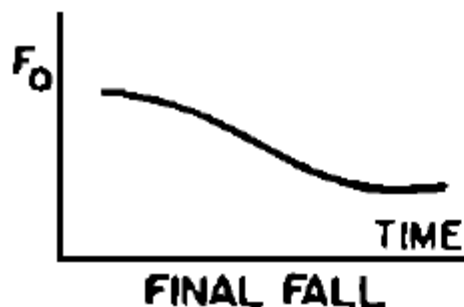
- ◉ Mattingly 1966: First algorithm for determination of  $f_0$  contour.
- ◉ Mattingly rules recognized three intonational tunes that could be placed on the last prominent syllable of a clause.
- ◉ Other prominent syllables could be marked by the user.

## CONTD...

- ◉ Mattingly 1966: First algorithm for determination of  $f_0$  contour.
- ◉ Mattingly rules recognized three intonational tunes that could be placed on the last prominent syllable of a clause.
- ◉ Other prominent syllables could be marked by the user.
- ◉ Unstressed symbols were generally lower in pitch because they were not assigned a target.

## CONTD...

- ◉ Mattingly 1966: First algorithm for determination of  $f_0$  contour.
- ◉ Mattingly rules recognized three intonational tunes that could be placed on the last prominent syllable of a clause.



CONTD...

## CONTD...

- ◉ Hart and Cohen (1973): Defined the intonational hat pattern as the tendency for intonation to

# CONTD...

- ◉ Hart and Cohen (1973): Defined the intonational hat pattern as the tendency for intonation to
  - Rise on the first stressed syllable of phrase

# CONTD...

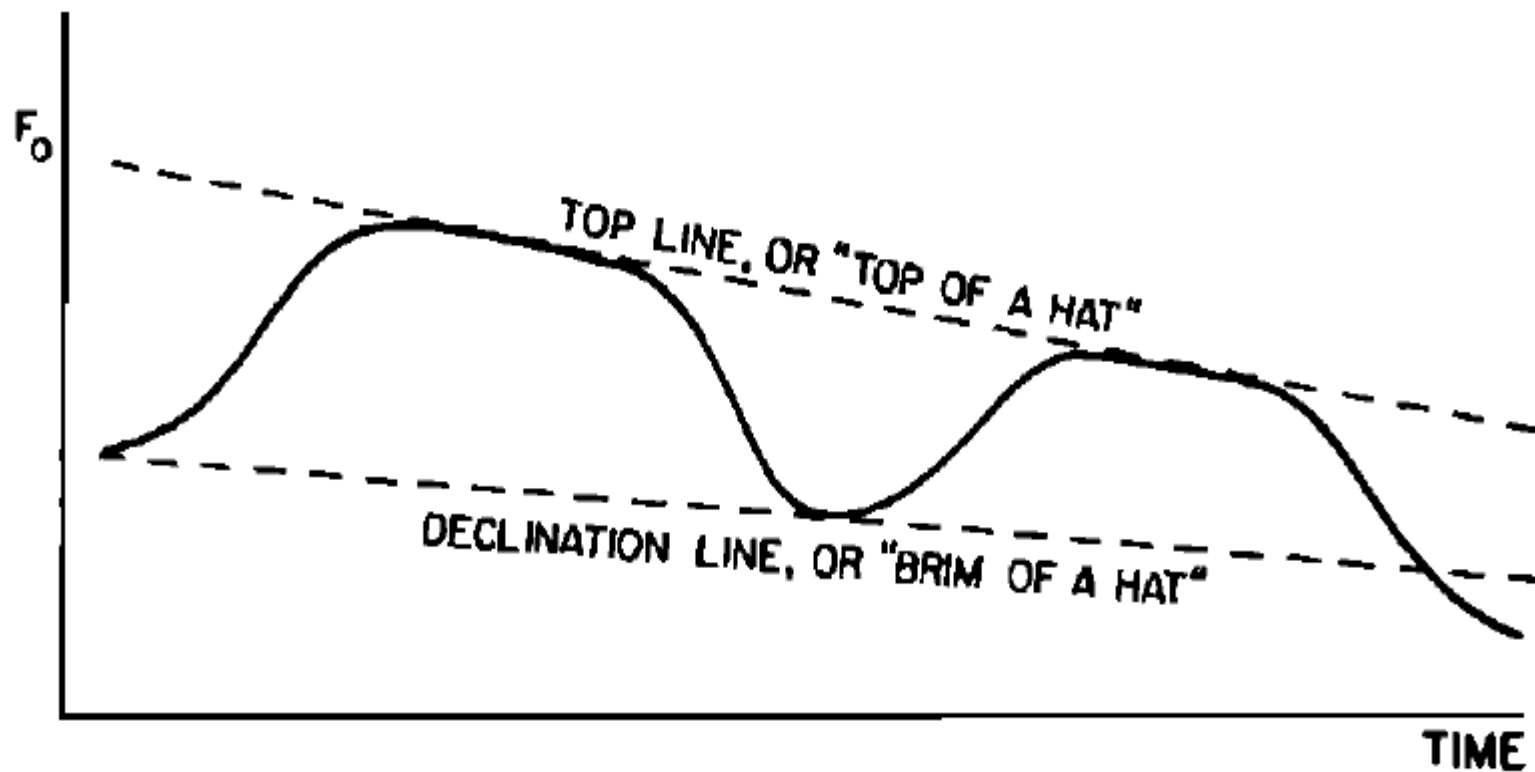
- ◉ Hart and Cohen (1973): Defined the intonational hat pattern as the tendency for intonation to
  - Rise on the first stressed syllable of phrase
  - Remain high until the final stressed syllable

## CONTD...

- ◉ Hart and Cohen (1973): Defined the intonational hat pattern as the tendency for intonation to
  - Rise on the first stressed syllable of phrase
  - Remain high until the final stressed syllable
  - Followed by either a dramatic fall or fall-rise depending on whether more material is to be spoken.



# CONTD...

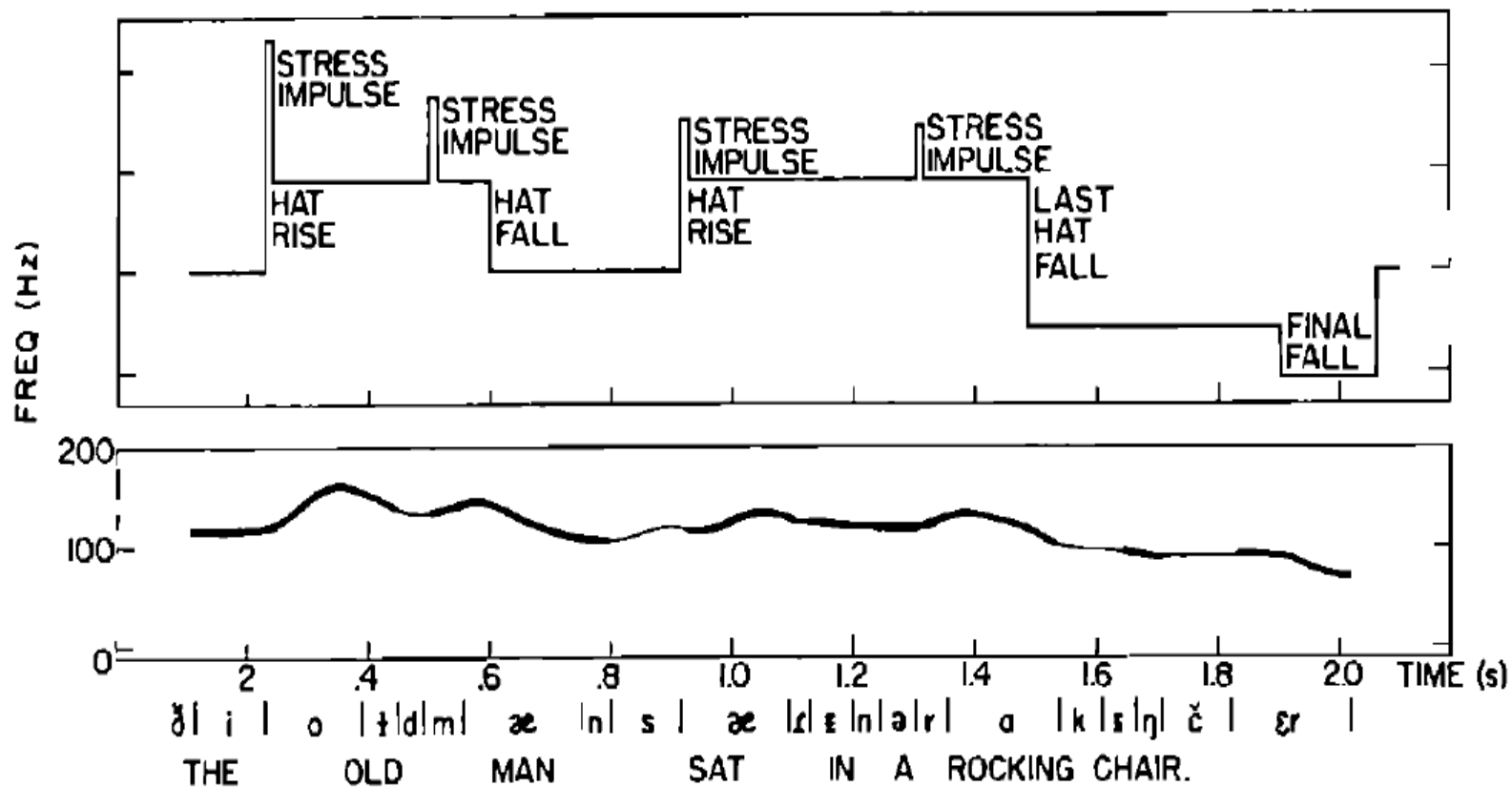


# KLATTALK $F_0$ RULES

# KLATTALK $F_0$ RULES

- ◉ Intonation contours can be modeled in terms of impulses and step commands fed to a linear smoothing filter.

# KLATTALK $F_0$ RULES



# ALLOPHONE SELECTION

# ALLOPHONE SELECTION

- ◉ Words are represented by phonemes and stress symbols.

# ALLOPHONE SELECTION

- ◉ Words are represented by phonemes and stress symbols.
- ◉ For cases where a phoneme can take on different allophones depending on the context of the word, a set of allophone selection rules is required.

# ALLOPHONE SELECTION

- ◉ Words are represented by phonemes and stress symbols.
- ◉ For cases where a phoneme can take on different allophones depending on the context of the word, a set of allophone selection rules is required.
- ◉ Phonetic form of function words



# ALLOPHONE SELECTION

- ◉ Words are represented by phonemes and stress symbols.
- ◉ For cases where a phoneme can take on different allophones depending on the context of the word, a set of allophone selection rules is required.
- ◉ Phonetic form of function words
  - ‘for’, ‘to’, ‘him’ reduced to [fɜ̃], [tə], [ɪm]

# ALLOPHONE SELECTION

- ◉ Words are represented by phonemes and stress symbols.
- ◉ For cases where a phoneme can take on different allophones depending on the context of the word, a set of allophone selection rules is required.
- ◉ Phonetic form of function words
  - ‘for’, ‘to’, ‘him’ reduced to [fɜ̃], [tə], [ɪm]
  - Klattalk: for is not reduced if next segment is vowel or silence.

# ALLOPHONES CONTD...

# ALLOPHONES CONTD...

- ◉ Selection of allophones involves many conflicting criteria and tradeoffs.

# ALLOPHONES CONTD...

- ◉ Selection of allophones involves many conflicting criteria and tradeoffs.
- ◉ The clearest cases are those where an phoneme is replaced by an allophone with distinctly different articulatory properties.

# ALLOPHONES CONTD...

# ALLOPHONES CONTD...

- ◉ Less clear are those where a small change is the result of a low level articulatory interaction.

# ALLOPHONES CONTD...

- ◉ Less clear are those where a small change is the result of a low level articulatory interaction.
- ◉ Or when many small changes can be made along an articulatory/acoustic dimension such as voicing onset time



# ALLOPHONES CONTD...

- ◉ Less clear are those where a small change is the result of a low level articulatory interaction.
- ◉ Or when many small changes can be made along an articulatory/acoustic dimension such as voicing onset time
- ◉ For example /t/ has VOT of +50ms but about 10ms longer in a word initial position e.g. “tone” than it is in pre-stressed word-medial positions e.g. “atone”.

# ALLOPHONES CONTD...

# ALLOPHONES CONTD...

- ◉ Should one create a separate symbol for each gradation along the VOT continuum, or handle these effects as low level adjustments to the time functions that control the synthesizer?

# ALLOPHONES CONTD...

- ◉ Should one create a separate symbol for each gradation along the VOT continuum, or handle these effects as low level adjustments to the time functions that control the synthesizer?
- ◉ The important thing is to be able to produce the appropriate acoustic changes in the synthetic speech.

# ALLOPHONES CONTD...

# ALLOPHONES CONTD...

- ◉ Allophones marks boundaries e.g. the distinction between “night rate” and “nitrate”

# ALLOPHONES CONTD...

- ◉ Allophones marks boundaries e.g. the distinction between “night rate” and “nitrate”
- ◉ A listener can easily tell which sequence was intended by the speaker because of the stronger frication/aspiration in the latter case.

# ALLOPHONES CONTD...

- ◉ Allophones marks boundaries e.g. the distinction between “night rate” and “nitrate”
- ◉ A listener can easily tell which sequence was intended by the speaker because of the stronger frication/aspiration in the latter case.
- ◉ Phonotactics or the specification of permitted phonetic sequences at the beginnings, middles and ends of words also can provide word boundary hypothesis for the listener.



# ALLOPHONES CONTD...

# ALLOPHONES CONTD...

- ◉ Phonological details differs for different dialects of English as well as for different speaking styles and speaking rates.

# ALLOPHONES CONTD...

- ◉ Phonological details differs for different dialects of English as well as for different speaking styles and speaking rates.
- ◉ TTS Systems need to select only rules appropriate to a single dialect of English and make modifications concerning rule applicability as a function of speaking rate.

# ACOUSTIC PHONETIC STUDIES

# ACOUSTIC PHONETIC STUDIES

- Lihiste noted the following kinds of word boundary indicators

# ACOUSTIC PHONETIC STUDIES

- Lihiste noted the following kinds of word boundary indicators
  - The presence of a laryngealized vowel onset usually signals the beginning of a word that starts with a vowel.

# ACOUSTIC PHONETIC STUDIES

- Lihiste noted the following kinds of word boundary indicators
  - The presence of a laryngealized vowel onset usually signals the beginning of a word that starts with a vowel.
  - A normally aspirated release of [p,t,k] becomes un-aspirated if a preceding [s] is part of the same word ("the spot" versus "this pot")

# ACOUSTIC PHONETIC STUDIES

- Lihiste noted the following kinds of word boundary indicators
  - The presence of a laryngealized vowel onset usually signals the beginning of a word that starts with a vowel.
  - A normally aspirated release of [p,t,k] becomes un-aspirated if a preceding [s] is part of the same word ("the spot" versus "this pot")
  - Selection between an initial or final allophone /r/ or /l/ intervocalically depends on the location of a word boundary on either side of the consonant



# ALLOPHONES CONTD...

# ALLOPHONES CONTD...

- A vowel is longer in duration in an open syllable (no word-final consonant), and shorter if followed by a voiceless word-final consonant.

# ALLOPHONES CONTD...

- A vowel is longer in duration in an open syllable (no word-final consonant), and shorter if followed by a voiceless word-final consonant.
- A word-final [t,d] is flapped or glottalized before a word beginning with a stressed vowel.

# ALLOPHONES CONTD...

# ALLOPHONES CONTD...

- ◉ Nakatani: phonetic cues and stress changes are perceptually more powerful cues to word boundary locations than are durational and pitch changes associated with syntactic boundary movements.

# ALLOPHONES CONTD...

- ◉ Nakatani: phonetic cues and stress changes are perceptually more powerful cues to word boundary locations than are durational and pitch changes associated with syntactic boundary movements.
- ◉ Additional phenomena noticed include

# ALLOPHONES CONTD...

- ◉ Nakatani: phonetic cues and stress changes are perceptually more powerful cues to word boundary locations than are durational and pitch changes associated with syntactic boundary movements.
- ◉ Additional phenomena noticed include
  - Geminate consonants are lengthened with respect to singletons (e.g. the /k/ in "drunk converse" versus "drunken verse").

# ALLOPHONES CONTD...

- ◉ Nakatani: phonetic cues and stress changes are perceptually more powerful cues to word boundary locations than are durational and pitch changes associated with syntactic boundary movements.
- ◉ Additional phenomena noticed include
  - Geminate consonants are lengthened with respect to singletons (e.g. the /k/ in "drunk converse" versus "drunken verse").
  - Vowels can be deleted and words resyllabified.



# ALLOPHONES CONTD...

- ◉ Nakatani: phonetic cues and stress changes are perceptually more powerful cues to word boundary locations than are durational and pitch changes associated with syntactic boundary movements.
- ◉ Additional phenomena noticed include
  - Geminate consonants are lengthened with respect to singletons (e.g. the /k/ in "drunk converse" versus "drunken verse").
  - Vowels can be deleted and words resyllabified.
  - There are restrictions on vowel reduction

# ALLOPHONES CONTD...

# ALLOPHONES CONTD...

- ◉ Zue and Laferriere(1979): Studied /t/ and /d/ and distinguished

# ALLOPHONES CONTD...

- ◉ Zue and Laferriere(1979): Studied /t/ and /d/ and distinguished
  - Within-word prestressed variants as in “return” and “reduce”

# ALLOPHONES CONTD...

- ◉ Zue and Laferriere(1979): Studied /t/ and /d/ and distinguished
  - Within-word prestressed variants as in “return” and “reduce”
  - Unstressed (shorter, less aspirated) versions as in “minty“, “mindy“, “moulted” and “molded”.

# ALLOPHONES CONTD...

- ◉ Zue and Laferriere(1979): Studied /t/ and /d/ and distinguished
  - Within-word prestressed variants as in “return” and “reduce”
  - Unstressed (shorter, less aspirated) versions as in “minty“, “mindy“, “moulted” and “molded”.
  - Voiced flaps as in “rarer” and “raider”

# ALLOPHONES CONTD...

- ◉ Zue and Laferriere(1979): Studied /t/ and /d/ and distinguished
  - Within-word prestressed variants as in “return” and “reduce”
  - Unstressed (shorter, less aspirated) versions as in “minty“, “mindy“, “moulted” and “molded”.
  - Voiced flaps as in “rarer” and “raider”
  - Glottalized or nasal released stops, as in “sweeten” and “Sweden”

# ALLOPHONES CONTD...

- ◉ Zue and Laferriere(1979): Studied /t/ and /d/ and distinguished
  - Within-word prestressed variants as in “return” and “reduce”
  - Unstressed (shorter, less aspirated) versions as in “minty“, “mindy“, “moulted” and “molded”.
  - Voiced flaps as in “rarer” and “raider”
  - Glottalized or nasal released stops, as in “sweeten” and “Sweden”
  - Deleted allophones, as sometimes occur in “pentagon”.



# ALLOPHONES CONTD...

# ALLOPHONES CONTD...

- ◉ Klatt measured burst durations and VOT for plosives in consonant clusters as a function of stress/phonetic/structural environments.

# ALLOPHONES CONTD...

- ◉ Klatt measured burst durations and VOT for plosives in consonant clusters as a function of stress/phonetic/structural environments.
- ◉ Proposed a set of quantitative rules to account for the data.

# ALLOPHONES CONTD...

- ⊙ Klatt measured burst durations and VOT for plosives in consonant clusters as a function of stress/phonetic/structural environments.
- ⊙ Proposed a set of quantitative rules to account for the data.
  - VOT for /p,t,k/ is longer in clusters with a following sonorant consonant

# ALLOPHONES CONTD...

- ⊙ Klatt measured burst durations and VOT for plosives in consonant clusters as a function of stress/phonetic/structural environments.
- ⊙ Proposed a set of quantitative rules to account for the data.
  - VOT for /p,t,k/ is longer in clusters with a following sonorant consonant
  - Shorter in a cluster with a preceding /s/

# ALLOPHONES CONTD...

- ◉ Klatt measured burst durations and VOT for plosives in consonant clusters as a function of stress/phonetic/structural environments.
- ◉ Proposed a set of quantitative rules to account for the data.
  - VOT for /p,t,k/ is longer in clusters with a following sonorant consonant
  - Shorter in a cluster with a preceding /s/
  - Shorter if the syllable is unstressed

# ALLOPHONES CONTD...

- ◉ Klatt measured burst durations and VOT for plosives in consonant clusters as a function of stress/phonetic/structural environments.
- ◉ Proposed a set of quantitative rules to account for the data.
  - VOT for /p,t,k/ is longer in clusters with a following sonorant consonant
  - Shorter in a cluster with a preceding /s/
  - Shorter if the syllable is unstressed
  - Longer in word initial position

# ALLOPHONES CONTD...

- ◉ Klatt measured burst durations and VOT for plosives in consonant clusters as a function of stress/phonetic/structural environments.
- ◉ Proposed a set of quantitative rules to account for the data.
  - VOT for /p,t,k/ is longer in clusters with a following sonorant consonant
  - Shorter in a cluster with a preceding /s/
  - Shorter if the syllable is unstressed
  - Longer in word initial position
  - Shorter if preceded by a voiced segment of preceding word



# ALLOPHONES CONTD...

# ALLOPHONES CONTD...

- ◉ Morpheme structure can be important in determining the acoustic realization of the consonants.

# ALLOPHONES CONTD...

- ⊙ Morpheme structure can be important in determining the acoustic realization of the consonants.
- ⊙ For example, /p,t,k/ are not strongly aspirated in /sp,st,sk/ clusters, except for the case where there is an obvious morpheme boundary after the /s/ as in “discourteous” and “miscalculate”

# ALLOPHONES CONTD...

- ◉ Morpheme structure can be important in determining the acoustic realization of the consonants.
- ◉ For example, /p,t,k/ are not strongly aspirated in /sp,st,sk/ clusters, except for the case where there is an obvious morpheme boundary after the /s/ as in “discourteous” and “miscalculate”
- ◉ The morpheme boundary symbol must be present in the abstract linguistic description for such words if the aspiration feature is to be computed correctly.

# ALLOPHONES CONTD...

# ALLOPHONES CONTD...

- Coker and Umeda

# ALLOPHONES CONTD...

- ◉ Coker and Umeda
- ◉ Prevocalic and postvocalic allophones may differ in acoustic aspects related to temporal buildup/decay of the sound source.

# ALLOPHONES CONTD...

- ◉ Coker and Umeda
- ◉ Prevocalic and postvocalic allophones may differ in acoustic aspects related to temporal buildup/decay of the sound source.
- ◉ The pre-voicing for [b, d, g] is weaker and less rich in higher harmonics in utterance initial positions.



# ALLOPHONES CONTD...

- ◉ Coker and Umeda
- ◉ Prevocalic and postvocalic allophones may differ in acoustic aspects related to temporal buildup/decay of the sound source.
- ◉ The pre-voicing for [b, d, g] is weaker and less rich in higher harmonics in utterance initial positions.
- ◉ [m, n, l] were a few dB weaker in intensity in word-initial position than in medial and final positions.

# ALLOPHONES CONTD...

- ◉ Coker and Umeda
- ◉ Prevocalic and postvocalic allophones may differ in acoustic aspects related to temporal buildup/decay of the sound source.
- ◉ The pre-voicing for [b, d, g] is weaker and less rich in higher harmonics in utterance initial positions.
- ◉ [m, n, l] were a few dB weaker in intensity in word-initial position than in medial and final positions.
- ◉ Noise intensity for [s] was about 3dB more intense word-initially than medially & finally.

# ALLOPHONES CONTD...

# ALLOPHONES CONTD...

- Oshika:

# ALLOPHONES CONTD...

- ◉ Oshika:
- ◉ Palatalization of word final alveolar consonants if the next word-final alveolar consonants if the next word begins with a palatal consonants.

# ALLOPHONES CONTD...

- ◉ Oshika:
- ◉ Palatalization of word final alveolar consonants if the next word-final alveolar consonants if the next word begins with a palatal consonants.
  - “Did you” [dɪju] and “this shoe” [ðʃʃiu]

# ALLOPHONES CONTD...

- ◉ Oshika:
- ◉ Palatalization of word final alveolar consonants if the next word-final alveolar consonants if the next word begins with a palatal consonants.
  - “Did you” [dɪju] and “this shoe” [ðʃʃiu]
- ◉ Zue found the effect to be asymmetrical applying to the [s] in “this shoe” but not to the [s] or [ʃ] of “wish some”

# ALLOPHONES CONTD...



# ALLOPHONES CONTD...

- Fertig measured formant values at ten equally spaced locations throughout each syllable for C-ə-C and performed averaging over time to obtain formant values associated with vowels.

# ALLOPHONES CONTD...

- ◉ Fertig measured formant values at ten equally spaced locations throughout each syllable for C-ə-C and performed averaging over time to obtain formant values associated with vowels.
- ◉ At least for this vowel the data suggested that formant motions can be additively decomposed into

# ALLOPHONES CONTD...

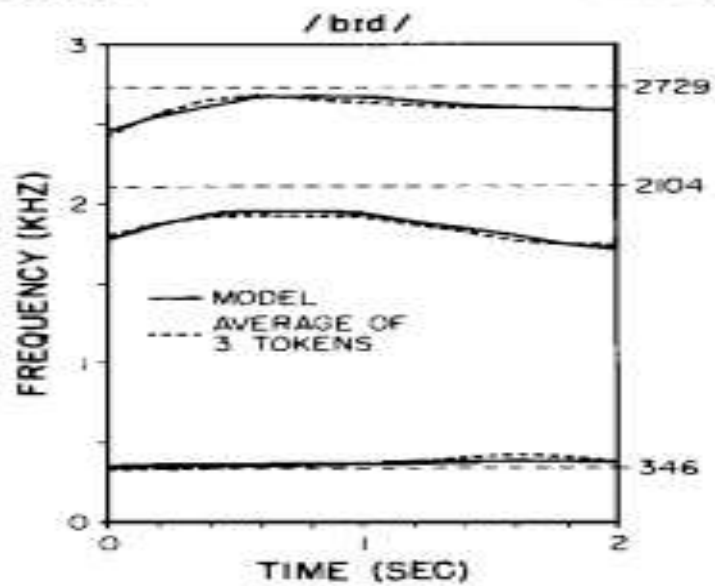
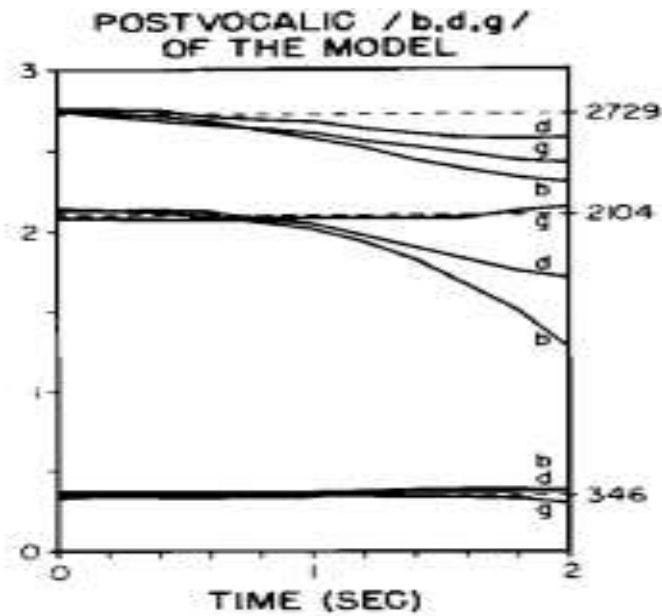
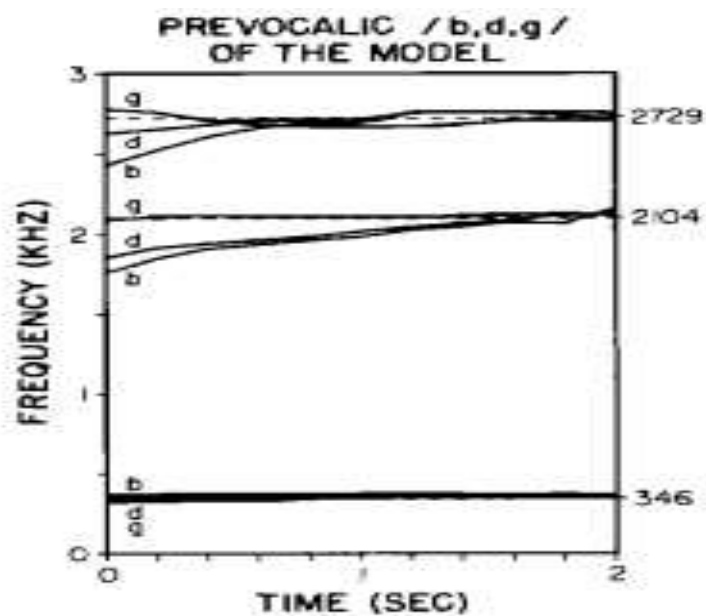
- ◉ Fertig measured formant values at ten equally spaced locations throughout each syllable for C-ə-C and performed averaging over time to obtain formant values associated with vowels.
- ◉ At least for this vowel the data suggested that formant motions can be additively decomposed into
  - An underlying vowel target

# ALLOPHONES CONTD...

- Fertig measured formant values at ten equally spaced locations throughout each syllable for C-ə-C and performed averaging over time to obtain formant values associated with vowels.
- At least for this vowel the data suggested that formant motions can be additively decomposed into
  - An underlying vowel target
  - A pre-vocalic consonants transition

# ALLOPHONES CONTD...

- ◉ Fertig measured formant values at ten equally spaced locations throughout each syllable for C-ə-C and performed averaging over time to obtain formant values associated with vowels.
- ◉ At least for this vowel the data suggested that formant motions can be additively decomposed into
  - An underlying vowel target
  - A pre-vocalic consonants transition
  - A post-vocalic consonants transition



# ALLOPHONES CONTD...

# ALLOPHONES CONTD...

- ◉ Of particular interest are rules that modify vowels in sentence contexts depending on consonantal context, stress and duration.



# ALLOPHONES CONTD...

- ◉ Of particular interest are rules that modify vowels in sentence contexts depending on consonantal context, stress and duration.
- ◉ The vowel shrinks when one goes from isolation to sentences, but it is not clear whether vowels tend to neutralize toward schwa, or simply accede to articulatory demands of adjacent consonants.

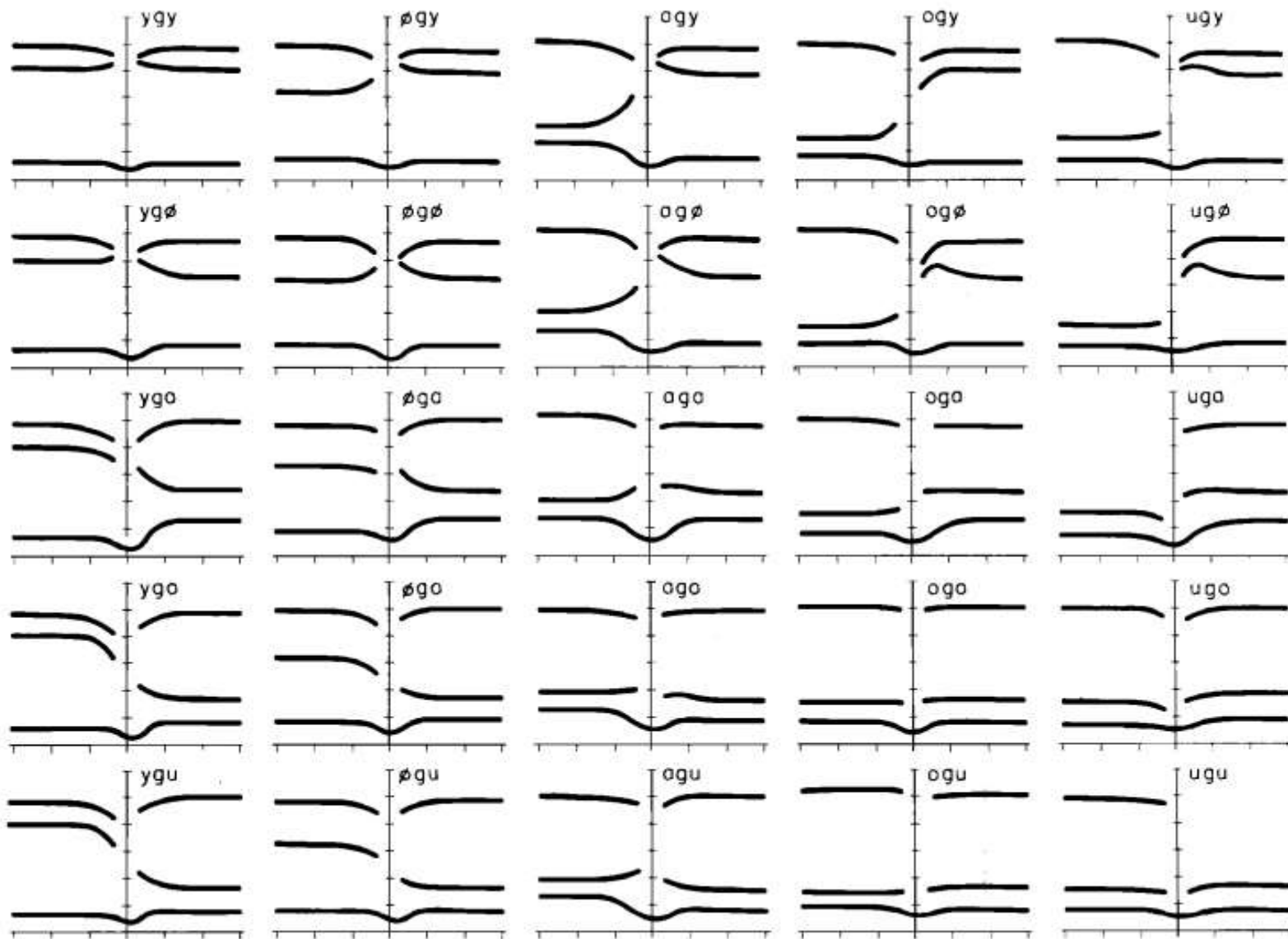
# ALLOPHONES CONTD...

- ◉ Of particular interest are rules that modify vowels in sentence contexts depending on consonantal context, stress and duration.
- ◉ The vowel shrinks when one goes from isolation to sentences, but it is not clear whether vowels tend to neutralize toward schwa, or simply accede to articulatory demands of adjacent consonants.
- ◉ Subjective impression of unnaturalness of synthesis by rule system can attributed to details of this sort.

# ALLOPHONES CONTD...

# ALLOPHONES CONTD...

- ◉ The formant transitions for CV syllable depend to some extent on the nature of the phonetic segment that precedes the consonant.



# CONCLUSION

# CONCLUSION

- ◉ TTS Systems have only begun to simulate the details of the allophone selection discussed.

# CONCLUSION

- ⦿ TTS Systems have only begun to simulate the details of the allophone selection discussed.
- ⦿ The area of allophonic detail and prosodic specification is one of the weaker aspects of rule systems.



