Rhythm and Dynamics of Speech

Robert Port
Linguistics, Cognitive Science
Indiana University

May, 2012
COST School, IIASPP, Vietri sul Mare, Italy
In modern linguistics, time is not treated as a property of language itself

- Linguistics ignores the timing of speech

- It is usually assumed that a language is a formal system consisting of timeless symbols in serial order. (eg, Chomsky)

- **Timing lies outside linguistics proper**
  (on the standard view)
Why do I say this?

- The minimal unit of language is assumed to be the "phoneme" or "phonological segment." (Assumed now for a century.)
- These are represented graphically using the letters of the phonetic alphabet.
- The phoneme is thought to be the greatest "discovery" of modern linguistics.
- But ....
Phonemes and phones are:

1. discretely different from each other (a small fixed set of types)
2. occur discretely in space (no overlap, etc) (modeling the perceptual sequence of phones)
3. serially ordered in space (modeling time)
4. static, invariant over time (they remain on the page indefinitely)
5. invariant across speakers, across contexts, speaking rate, etc.

Example: \[ [ t^h \theta \ l m e j \ r o ] \] 6 segments

SO phones and phonemes are really LETTERS reinterpreted as SOUNDS!
But how can sounds be timeless??
Where did the `phoneme’ come from?

1. David Abercrombie pointed out (1949) that from classical times until the 20\textsuperscript{th} C, ``letters” (L. \textit{litera}) were viewed equally as sound or graphics. Eg, Wm Holder (1669) ``The elements of language are \textit{letters} – simple discriminations of breath or voice”.

2. Only in the late 19\textsuperscript{th} C was a clear distinction made between graphics and sound. Saussure (in 1890s) said ``The domain of linguistics is speech, not written texts”

3. But 19\textsuperscript{th} C linguists noted that description of languages using the Roman alphabet was difficult since:
   – Orthographies were internally inconsistent
   – Letters had different interpretations in each language
4. So the *International Phonetic Association* was formed in 1886 to establish a consistent alphabet for scientific and pedagogical use. The ``IPA”
the international phonetic alphabet (2005)

**consonants (pulmonic)**

<table>
<thead>
<tr>
<th>Nasal</th>
<th>m</th>
<th>m</th>
<th>n</th>
<th>n</th>
<th>n</th>
<th>n</th>
<th>n</th>
<th>n</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plosive</td>
<td>p</td>
<td>b</td>
<td>t</td>
<td>d</td>
<td>t</td>
<td>d</td>
<td>c</td>
<td>j</td>
<td>k</td>
</tr>
<tr>
<td>Fricative</td>
<td>φ</td>
<td>β</td>
<td>f</td>
<td>v</td>
<td>θ</td>
<td>θ</td>
<td>s</td>
<td>z</td>
<td>s</td>
</tr>
<tr>
<td>Approximant</td>
<td>u</td>
<td>ɾ</td>
<td>ι</td>
<td>j</td>
<td>j</td>
<td>ι</td>
<td>j</td>
<td>w</td>
<td></td>
</tr>
<tr>
<td>Trill</td>
<td>B</td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tap, flap</td>
<td>t</td>
<td>ɾ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral fricative</td>
<td>t</td>
<td>ɾ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral approximant</td>
<td>l</td>
<td>l</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral flap</td>
<td>l</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Where symbols appear in pairs, the one to the right represents a modally voiced consonant, except for murmured h.

Shaded areas denote articulations judged to be impossible. Light grey letters are unofficial extensions of the IPA.

**consonants (non-pulmonic)**

<table>
<thead>
<tr>
<th>clicks</th>
<th>implosives</th>
<th>ejectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bilabial fricated (&quot;dental&quot;)</td>
<td>Bilabial</td>
<td></td>
</tr>
<tr>
<td>Laminal alveolar fricated (&quot;dental&quot;)</td>
<td>Dental or alveolar</td>
<td>p'</td>
</tr>
<tr>
<td>Apical (post)alveolar abrupt (&quot;retroflex&quot;)</td>
<td>Retroflex</td>
<td>t'</td>
</tr>
<tr>
<td>Subapical retroflex</td>
<td>Palatal</td>
<td>k'</td>
</tr>
<tr>
<td>Laminal postalveolar abrupt (&quot;palatal&quot;)</td>
<td>Velar</td>
<td>t'</td>
</tr>
<tr>
<td>Lateral alveolar fricated (&quot;lateral&quot;)</td>
<td>Lateral affricate</td>
<td>Alveolar fricated</td>
</tr>
</tbody>
</table>

Examples:
- Bilabial fricated:
- Dental or alveolar:
- Retroflex:
- Palatal:
- Velar:

**consonants (co-articulated)**

- m: Voiceless labialized velar approximant
- w: Voiced labialized velar approximant
- u: Voiced labialized palatal approximant
- ə: Voiceless palatalized postalveolar (alveolo-palatal) fricative
- z: Voiced palatalized postalveolar (alveolo-palatal) fricative
- h: Simultaneous x and f (existence disputed)
- ʃ: Affricates and double articulations may be joined by a tie bar

**vowels**

- Front | Near front | Central | Near back | Back |
- Close | i | y | i | u |
- Near close | I | Y | ɾ | ɾ |
- Close mid | e | œ | ə | ə |
- Mid | æ | ø | ø | ø |
- Open mid | ɛ | œ | ɔ | ɔ |
- Near open | æ | œ | œ | œ |
- Open | a | œ | œ | œ |

**suprasegmental**

- Primary stress | Extra stress

- Secondary stress
- Long [fɔnə'tʃən] | Short ⧫
- Extra-Short ⧫

- Syllable break ⧫ Linking (no break)

- intonation
- Minor (foot) break ⧫
- Major (intonation) break ⧫

**(tone)**

- Level tones
- Contour tones (e.g.)

- Top ⧫
- High | Falling
- Mid ⧫
- Low | Low rising
- Bottom ⧫
- High falling

- tone terracing ⧫
- Low falling

- Upstep ⧫
- Peaking ⧫
5. Then Polish linguist Beaudoin de Courtenay (1899) proposed the **phoneme as the psychological counterpart** of the phonetic symbols (ie, *letters*).

So the ``phoneme” was an invention that **collapsed the graphic and acoustic once again**!
4. Presumed properties of phonemes were:
   I. Discrete
   II. Serially ordered
   III. Timeless and non-overlapping
   IV. Invariant across contexts
   V. Fixed inventory

5. Yet they were called "sounds" and assumed to have "acoustic correlates" – even though timeless, etc!
4. Presumed properties of phonemes were:
   I. Discrete
   II. Serially ordered
   III. Timeless and non-overlapping
   IV. Invariant across contexts
   V. Fixed inventory

5. Yet they were called "\textit{sounds}" and assumed to have "\textit{acoustic correlates}" – even though \textit{timeless}, etc!

\textbf{All are properties of letters!}
• Amazingly these **contradictions have been overlooked** (by most*) all these years!

• After early debates (eg, Twaddell, 1935), the view of the **phoneme as a cognitive spelling code** became standard for nearly all modern linguists.

• As technology to study acoustic waveforms and spectra improved through the 20\textsuperscript{th} C, **nothing with the expected properties were found in sound**.

Yet we **could not give up** on **phonemes with acoustic invariants**.

[Turns out the phoneme is a ``**cognitive blend**’’ of **letters** with **sound**. (See Fauconnier and Turner, 2002)]
Why are phonemes such a powerful idea?

1. They are very simple and economical for writing.
2. They are intuitive.
3. Capture the abstractness of language.
4. We have spent our lives sharpening skill at translating sound into letters and back!
Why are phonemes such a powerful idea?

1. They are very simple and economical for writing.
2. They are intuitive.
3. Capture the abstractness of language.
4. We have spent our lives sharpening skill at translating sound into letters and back!

Because
NO ALPHABET CAN DESCRIBE WHAT SPEAKERS KNOW ABOUT WORDS!
Some evidence:

1. Speech sounds vary continuously – not discretely.
2. Words exhibit limitless variation – both in production and perception.
   In one speech corpus, 20% of words lack at least 1 phoneme. Of 4-6 syllable words, 22% lack at least 2 syllables! (K. Johnson, 2004)
3. Data show that speech memory retains indexical (speaker-specific) features.
   Recognition memory for a word list is better when the voice is the same (Palmeri, et al, 1993; Goldinger, 1996)
4. (Almost) no acoustic invariants can be found for phonemes – across all contexts, speakers, styles, etc.
5. Languages exhibit many temporal patterns -- not simply serial order.
   V/C ratio for English rabid/rapid, fuzzy/fussy
Let’s look at some common temporal patterns in speech, especially periodic ones.
Speech Rhythm

People everywhere produce speech that is *sometimes periodically* structured.

(A bold hunch not based on systematic data)
• There may be neural oscillators underlying periodic speech (and perhaps limb motions as well).
• Oscillations with frequencies in the range (0.5 Hz – 5 Hz) are found in speech.
• It appears that speech gestures that produce salient (loud) auditory patterns are attracted to certain phase angles.
Repetitive events can be modeled using a circle usually with a unique point defined as phase zero. A circle, in turn, can be represented as a sinusoid.

Phase zero – a unique, distinctive point in the cycle
Rhythmic speech is more common than most of us notice.

Here are some examples of rhythmic speech in various languages

1. **Chants and songs**
   - Chant of a Buddhist sutra, Thailand
     Obviously periodic, normatively periodic
I went to the garden of love
And saw what I never had seen
A chapel was built in the midst
Where I used to play on the green.

Black arrows mark vowel onsets.
Red arrows are halfway between nearest black arrows

Speaker unknown
Example 4: Actress - teasing

```
Oohooo, you great big Latin lover you!
```

Lucille Ball
Desi Arnaz
Example 5: Story teller

Drive around in his Chevy Impala talking about

*religion* and *theater* and *school* and our *families*

Our security will not be gained by timid measures.
Our security requires constant vigilance and decisive action.
I believe America has only one option.
We must fight this war until our work is done.
Example 7: News broadcaster: signoff

Daniel Schorr

```
...that would help to make sure that the other districts would likely go Republican.
It’s going to take some time, I think, to see how this all works out.”
```
8. Laboratory Speech-Cycling task

It is easy to get people to speak rhythmically in the lab.

Participants repeat a phrase (chantlike)

   Eg, "Take a pack of cards"

Repetition of any phrase encourages entrainment to a periodic oscillator (an internal one).
`Speech Cycling Task’  

Cummins-Port (1998) *J. Phonetics*

- Subjects repeat a short text over and over. Eg, `*Take a pack of cards, Take a pack of cards, …’`
- *Take* and *cards* are stressed.
- If *take* defines the beginning of a cycle (phase zero), are there any constraints on the timing of *cards*?
  - Yes, powerful ones.
• Ss heard a **double metronome**:  
  – ‘Beep’ for first word (phase zero)  
  – ‘Boop’ for final word (target)
• **Boop** is located at a *continuous range* of phase angles from  
  $\varphi = 0.25$ to $0.70$
• Ss asked to put ‘*Take on Beep and cards on Boop.*’
• Onsets are found by **lo-pass filtering** (about 900 Hz), **smoothing** and locating steep rises in energy.
• Many thousand trials.
Results

Subjects had strong *preference* to locate *onsets* at `harmonic fractions’ of the cycle: 1/3, 1/2, 2/3.

Why these 3 phase angles?
Preferred phases suggest *two oscillators* coupled to each other.

A pair of oscillators is proposed with integer-ratio frequencies. The model is based on the *Haken-Kelso-Bunz model* for finger wagging.
Meters in music, song and speech cycling seem to depend on **coupled pulse oscillators**.

These provide **target times** for salient speech events (vowel onsets) that are equally spaced.
Hypothesis

• The pulses (phase zeros) of the internal oscillators attract salient acoustic events (ie, syllable onsets).

• The resonant patterns bias the timing of speakers by creating temporal attractors.

• To demonstrate periodicity in speech, try to demonstrate the presence of the attractors, not absolute isochrony!

The preferred timing patterns can be viewed as natural resonances of English speech timing. These resonances probably differ from language to language.
9. Japanese mora timing

Japanese words consist of an integral number of moras.

- Ta-ka-gi 3 moras
- To-o-kyo-o 4 moras
- Ho-n-da 3 moras
- cho-t-to 3 moras
- Ta-ka-gi-sa-n 5 moras

It is said moras each take the same amount of time. (Traditional Japanese pedagogy – but possibly inspired by kana orthography.)
• Our proposal is that speakers entrain perceptual pulses at vowel onsets (just like English) to an internal oscillator.
How we find `speech pulses’

In Japanese phrase ‘Takagi Toko-san’

- Waveform
- Low-pass filtered (800 Hz) and rectified
- Smoothed
- Positive first derivative
- Threshold to find onsets (some manual cleanup)

Data from Brady, Port & Nagao (2006) ASA Annual Meeting, Honolulu
Generate a `mora sinusoid.'

Find optimal period.

Onsets map to phase circle

look for phase clustering

Work done with Michael C Brady, ICPhS 2007
`Geminate K’ lengthens stop closure but adds a full mora of duration

Ta     ka      gi        To     k k o     sa n     ...

\[\text{Anticipatory IOI} \quad \text{Target IOI}\]
Phase clustering

discordant onset
\( \bar{R} \) as a measure of clustering

If vectors tend to point in the same direction, \( \bar{R} \) length will be large. If vectors tend to point in divergent directions, \( \bar{R} \) length will be small.

\[
\bar{R} = \frac{R}{n}
\]

\( 0 \leq \bar{R} \leq 1 \)
Example of Japanese for teaching foreigners

Ano otoko no shito no na-ma-e wa Shimada Yukio de

Ano onna no shito no na-ma-e wa Takahashi Noriko des(u)

``That woman’s name is Takahashi Noriko”
`Emphatic Periodicity’ in English

- Periodicity *comes and goes* in many speakers

- Especially *professional speakers*: radio announcers, preachers, politicians, teachers, ...

- Periodic speech is common in `*punchlines,*’ or emphasized phrases.
I’m here to tell you,
When we are at our lowest,
When we are suffering and hurting,
If you’ll stand real still, you’ll feel somebody standing right behind you.
(congregational response)
Hallelujah. If you listen real close you’ll hear Him speaking your name.

God will never leave us or forsake us.

Rev. Pope, South Carolina, 2002
the Sunday after a hurricane
First excerpt: Nonperiodic speech

Im here t’ tell ya ... when we are at our lowest .. when we are ... suffrin’ and hurtin’

$R = 0.22$
Punchline excerpt: Very periodic

2 frequencies give good clusters
Clustering at range of periods

Syllables: 240 ms

Stresses: 480 ms

\[ \bar{R} \]

Sinusoid period in ms

R shows 2 good peaks in 2:1 ratio
Example 11  Actor: conclusion

Concluding speech by Orson Wells in ‘War of the Worlds’ radio show 1939

Ah, strange it now seems to sit in my peaceful study .... Princeton

Longer excerpt

$\bar{R} = 0.67$
Conclusions

1. Linguists thought language is constructed from a small inventory of tokens. But no. There is vast richness that disappears when we write words down with letters! Phonemes distort language. Our intuitions are biased by learning to read and write.

2. Speech is a complex motor act. One that resembles other animal behaviors like walking, arm movements, etc.

3. Speech is not always periodic, but slips in and out of periodicity easily.

4. This periodicity is often applied to draw listeners‘ attention.

More research is required to study these overlooked phenomena