Speech Production 1

Paper 9:
Foundations of Speech Communication
Lent Term: Week 3

Katharine Barden
Outline

• **Complexity** of speech production
  – Timing and coordination
  – Redundancy
  – Compensatory articulation
  – Contextual variation
  – ‘Frames of reference’

• **Methods and techniques** for studying speech production

• **Coarticulation**
  – What can studies of coarticulation tell us?
  – Models of coarticulation
What is “speech production”? 

• Transformation of ideas into movements/sounds of speech

• Speaking involves many tasks, e.g.
  - Conceptualising an utterance
  - Building its syntactic structure
  - Retrieving the right words in the right order
  - Controlling the articulators to say the words
**Tight coordination** of more than one articulator is necessary e.g. ‘spa’:

<table>
<thead>
<tr>
<th>Target</th>
<th>Critical</th>
<th>Supporting</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>tongue tip constriction area $&gt;0 \text{ mm}^2$, $&lt; x \text{ mm}^2$</td>
<td>jaw high; tongue body fronted</td>
</tr>
<tr>
<td>p</td>
<td>lip constriction area $0 \text{ mm}^2$, lip pressure</td>
<td>jaw high enough</td>
</tr>
<tr>
<td>a</td>
<td>pharynx constriction area narrow; (tongue back)</td>
<td>lip constriction area (open); jaw low</td>
</tr>
</tbody>
</table>

+ coordination of respiratory, glottal and velar activity…
Redundancy: the ‘degrees of freedom’ problem

• More than 1 way to do a skilled motor task because skilled tasks require coordination between different parts of the body

• e.g. to touch forehead with finger can use shoulder, elbow, wrist, knuckle in many different combinations → ‘motor equivalence’
Redundancy: the ‘degrees of freedom’ problem

• There are a huge number of muscles involved in speech:
  – respiration
  – tongue
  – jaw
  – lips (upper and lower)
  – tension in the vocal folds
  – velum

→ many ways to achieve a goal
→ difficult to explain how a trajectory is planned
Compensatory adjustments to achieve a goal

Unperturbed (‘normal’) /i/ articulation

Lowered jaw (due to bite block); compensatory adjusting of the tongue to produce /i/

In the face of perturbation, an acoustically equivalent alternative articulation may rapidly be achieved.

Figure from Lindblom (1983: Fig. 10.9). Original study: Lindblom et al. (1979).
Complexity of speech production

- Articulatory movements are subject to several types of **contextual variation**
  e.g. due to stress, rate, speaking style, prosody, neighbouring sounds
Reference frames: the ‘translation’ problem

• Many different frames of reference can be used to describe both:
  the goals of speech production
    – linguistic, acoustic, auditory, articulatory, proprioceptive
  & the movements involved
    e.g. articulator-oriented, motor commands, coordinative structures
Coordinative structures

• Groups of muscles organized for the achievement of a particular skilled activity

• Same muscles play a role in a large number of different coordinative structures

• Detailed operation of the high level command is executed automatically by groups of muscles that form coordinative structures
Reference frames: the ‘translation’ problem

• ‘Translation’ between different frames is difficult to explain
  – linguistic units traditionally thought of as abstract, discrete and time-free
  – articulation is concrete, continuous and happens in time
How do we transform an idea into the dynamic, context-sensitive neuromuscular activity that produces the acoustic signal of speech?
How can we study the organisation of speech production?

• No direct access to mental activity – instead infer processes from:
  
  - acoustic data
How can we study the organisation of speech production?

– movement data

• X-rays
• X-ray microbeam

Pellet positions on midsagittal trace

Westbury, Hashi and Lindstrom (1995: Fig.1)
X-ray µ-beam showing persistence of inaudible gestures

- **List Production**
  - "perfect, memory"

- **Phrasal Production**
  - "perfec(t) memory"

- Phrasal: /m/ closure overlaps /t/ release, making it inaudible; /t/ gesture is present nevertheless

U. Tokyo X-ray µ-beam
(Fujimura et al. 1973)
How can we study the organisation of speech production?

- movement data
  - EMA (electromagnetic articulography)

http://www.qmuc.ac.uk/ssrc/default.htm

Tabain and Perrier (2005: Fig. 2)
How can we study the organisation of speech production?

– movement data

• MRI (magnetic resonance imaging)

/\textit{t}/ \quad /\textit{t}^h/ \quad /\textit{t}'/

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{images}
\caption{Kim, Honda and Maeda (2005: from Fig. 5)}
\end{figure}
How can we study the organisation of speech production?

– movement data
  • ultrasound

http://www.qmuc.ac.uk/ssrc/default.htm
How can we study the organisation of speech production?

– movement data

• EPG (electropalatography)

http://www.qmuc.ac.uk/ssrc/default.htm
Ellis and Hardcastle (2002: from Fig. 4)
How can we study the organisation of speech production?

- airflow data
  - measures airflow from mouth and nose
  - records air pressure in mouth and pharynx

http://www.linguistics.ucla.edu/faciliti/facilities/physiology/aero/aero.htm
How can we study the organisation of speech production?

- Electromyography
- Experimentally-induced compensatory articulation (e.g. bite-blocks)
- Psycholinguistic data (e.g. reaction times to more/less frequent words)
- Neuroimaging
- Speech errors
- Speech pathologies
Coarticulation

- “Coarticulation … refers to the fact that a phonological segment is not realized identically in all environments, but apparently varies to become more like an adjacent or nearby segment”
  (Kühnert and Nolan 1999: 7)

  e.g. /k/ in ‘key’ → [ [+ ] ] (advanced)

  in ‘caw’ → [ הו ] (retracted & with lip rounding)

- Sometimes detectable by impressionistic observation, sometimes only apparent from instrumental investigation
Terms - Coarticulation

- right-to-left = anticipatory
  e.g. rounding on /s/ and /p/ of /spu:n/

- left-to-right = perseverative = carryover
  e.g. /i:t/ vs /bi:t/
  in context following /b/, /i:/ has lower formants
What can coarticulation studies tell us?

*Units of speech production…*

- Most theories of coarticulation assume that phoneme-sized segments play a part in speech production
- Coarticulation is viewed as a failure to achieve the target phonemes
- Do we actually store an articulatory plan for saying each word? (in which case coarticulation is not the same ‘problem’)
What can coarticulation studies tell us?

**Units of speech production…**

- Kuhnert and Nolan (1999) argue that similar coarticulatory patterns for the same phoneme in different words supports the psychological reality of the phoneme for production
  - it is more efficient to have phonemes plus coarticulation than to have a separate representation for each word/syllable

- A simple ‘phonemes plus coarticulation’ approach is not complete on its own e.g. it does not account for structural constraints on production
What can coarticulation studies tell us?

Organisational structures...

- Temporal or structural limits on coarticulation may shed light on the psychological reality of particular linguistic structures...
Temporal/structural limits on coarticulation?

- Coarticulatory effects can extend beyond adjacent segments
  - e.g. vowel-to-vowel coarticulation
“BARB a barb” vs “BEEB a beeb”

/babəbab/  /bibəbib/
Temporal/structural limits on coarticulation?

• Extensive coarticulatory effects of lip-rounding; more anticipatory than carryover

• In French, lip-rounding can occur as many as 6 segments before a rounded vowel (Benguerel and Cowan 1974)

• In English, rounding precedes /u/ onset by a fixed time interval (approx. 200 ms) - regardless of no. of intervening consonants or presence of a word boundary (Bell-Berti and Harris 1979)
Temporal/structural limits on coarticulation?

Nasalisation

• In a CVVn sequence, velar lowering can be initiated as early as the first consonant in English and French

• Nasal coarticulation not hindered by syllabic or word boundaries in English, French or Hindi

See Chafcouloff and Marchal, p. 74 for refs
Temporal/structural limits on coarticulation?

Tongue & jaw coarticulation

→ Generally less extensive spread than for lips or velum

→ Spread fairly equal in anticipatory and carryover directions
Temporal/structural limits on coarticulation?

Some structural influences…

- In VpV sequences in a carrier sentence, there is more coarticulation when unstressed than stressed (Engstrand, 1988)

- Consonants are generally more sensitive to coarticulatory effects in syllable-final position than in syllable-initial position (see Recasens, p. 102)

- $C_1$ and $C_2$ in /kl/ cluster can overlap less across a sentence or clause boundary than across a word or syllable boundary (Hardcastle, 1985)
Connected speech processes differ in content words and function words

<table>
<thead>
<tr>
<th>content</th>
<th>function</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>ban</em> <em>thatch</em></td>
<td><em>ban</em> <em>that</em></td>
</tr>
<tr>
<td>[banθətʃ]</td>
<td>[baŋətʃ]</td>
</tr>
<tr>
<td>[bānθətʃ]</td>
<td></td>
</tr>
<tr>
<td><em>[banːatʃ]</em></td>
<td></td>
</tr>
<tr>
<td><em>ban</em> <em>zips</em></td>
<td><em>ban</em> <em>this</em></td>
</tr>
<tr>
<td>[banzɪps]</td>
<td>[baŋəts]</td>
</tr>
<tr>
<td>[bānzɪps]</td>
<td></td>
</tr>
<tr>
<td>[bāzɪps]</td>
<td></td>
</tr>
<tr>
<td><em>[bānːɪps]</em></td>
<td></td>
</tr>
</tbody>
</table>

(Local, 2003)
Temporal/structural limits on coarticulation?

• The evidence doesn’t support idea of one particular linguistic unit governing the spreading of properties

• But there do seem to be some structural influences on the extent of coarticulation
What can coarticulation studies tell us?

Cognitive/physical factors

- Does the extent of coarticulation vary between languages?
  - If coarticulatory properties are found to be language-specific, then coarticulation cannot be purely due to physical factors (e.g. inertia, momentum of the articulators)
Is coarticulation language-specific?

Manuel (1999) presents evidence that the extent of coarticulation is sensitive to the system of phonological contrasts in a language.

- Vowel-to-vowel coarticulation is greater in Shona (which has a five-vowel system) than in English (with approx 15 vowels) (Manuel and Krakow, 1984)

- In Marshallese (which has no front-back dimension in vowels), F2 trajectories in vowels are dependent on target F2 values for surrounding consonants (Choi, 1992; 1995)
Models of coarticulation...

– attempt to elucidate the ‘phonology → articulation’ stage of production

– make hypotheses about how speech production targets are represented
Context-sensitive allophones (Wickelgren 1969, 1972)

• Input to speech production = $y x_z$
  (allophones appropriate for left and right phonemic contexts)

• Problems:
  – large number of allophones to store
  – ignores coarticulation across >1 segment
  – ignores structural constraints
Feature-spreading (Henke 1966)

• Utterance plan = matrix of essential phonological features, no redundant features marked

<table>
<thead>
<tr>
<th></th>
<th>s</th>
<th>p</th>
<th>u:</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>rounding</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

• Features from later phonemes copied onto earlier segments by look-ahead process

• But:
  – no good for gradient effects
  – phonetic detail not adequately explained, e.g. /usu/ may show less lip-rounding on /s/ than in vowels
Coarticulation Resistance
(Bladon and Al-Bamerni, 1976)

• Some segments are more resistant to coarticulation than others
  – English apical consonants /l/ and /n/ become laminal in the context of the laminal fricatives /s/ and /z/,
    but /s/ and /z/ don’t become apical in an apical context
  – In RP, dark [l] more resistant to coarticulation than clear [l]

• **Gradient approach** is necessary

• Allocate
  – higher CR value features of /s/ and /z/ than /l/ and /n/
  – higher CR value to features of dark [l] than clear [l]
Coarticulation Resistance
(Bladon and Al-Bamerni, 1976)

• Each feature of each allophone has a coarticulation resistance (CR) index
  e.g. [3 CR], [10 CR]…

• Anticipatory and carryover coarticulation is allowed to spread freely until it is inhibited by a high CR value on some segment

• CRs possibly different for each language/accent
Coarticulation Resistance  
(Bladon and Al-Bamerni, 1976)

• Difficulties?
  – rate and style
  – structural constraints
  – descriptive, not explanatory?
Window model (Keating 1988, 1990)

• Coarticulation can be either:
  * **phonological**
    - caused by rules of feature spreading
    - successive segments share an attribute fully
  * **phonetic**
    - more gradual in time and space
    - affects portions of segments to varying degrees
Window model (Keating 1988, 1990)

Phonetic realisation:

– conversion of feature combinations into spatio-temporal targets
– each feature is associated with a range of values called a ‘window’
– the window represents the overall contextual variability
– the wider the window, the more coarticulation possible for that segment on that feature
Window model (Keating 1988, 1990)

- Phonetic coarticulation results through finding the most efficient pathway through the windows of successive segments

Farnetani (1997: Fig. 12.8)
Window model  (Keating 1988, 1990)

• Resolves e.g. /usu/

• But:
  – rate and style
  – structural constraints
  – more descriptive than explanatory
‘Coproduction’ theories
(e.g. Fowler; Bell-Berti and Harris)

• Hypothesis: underlying invariance
Reality: surface variability
→ reconcile by redefining primitives of speech production as dynamically specified ‘gestures’

• Coarticulation is the variability due to overlapping of gestures in time
A sequence of three overlapping gestures based on Farnetani (1997: Fig. 12.9)
A sequence of three overlapping gestures based on Farnetani (1997: Fig. 12.9)

gesture 2 prominent, i.e. has maximal influence on VT shape;
lesser influences from gestures 1 and 3
A sequence of three overlapping gestures

Based on Farnetani (1997: Fig. 12.9)
‘Coproduction’ theories

• Gestures have their own *temporal structure*
• Gestures can *overlap in time*
• Degree of overlap controlled at plan stage

• Gestures are not altered/influenced by other gestures – just overlap with adjacent gestures, i.e. are coproduced
‘Coproduction’ theories

• Amount of overlap depends on the extent to which articulators are shared

e.g. VbV
  bilabial closure – jaw, upper lip, lower lip
  jaw = only articulator required in gestures for adjacent V goals

cf. VdV or VgV
  alveolar contact / velar contact
  tongue involved in gestures for V goals and consonant
  → greater interference/blending of gestures

• Consistent with patterns of coarticulation observed
‘Coproduction’ theories

Problems:

– How are these gestures learned…
  …from (principally) auditory input?
  …given their contextual variation?

– Extremely reduced speech – does it involve the same gestures as citation forms?
Summary & next lecture…

• Speech production – highly complex activity
• Not clear how speech production targets are represented – phonemes, feature matrices, words, gestures…?
• Coarticulation offers insights into the units and organisational structures involved in speech production
• Next week: models of speech production – planning of utterance → production