Speech Science: Technique, Concept, Theory

Gary Weismer

Department of Communication Sciences and Disorders

UW-Madison

ASHA Zemlin Lecture 2012
Disclosure Statements

Much of the work reviewed here was funded by NIH (Award NIDCD R01 DC03723, G. Weismer, PI)

The presenter has no financial interest related to the content of this presentation

The presenter’s co-authors on various publications and presentations associated with the work presented here: [Drs. Y. Yunusova (U. Toronto), R. Kent (UW-Madison), P. Milenkovic (UW-Madison), K. Bunton (U. Arizona), K. Tjaden (U. Buffalo), J. Westbury (UW-Madison), C. Kuo (U. Buffalo), G. Turner (U. Central Missouri), & Y-J. Kim (Louisiana State U.)] have no responsibility for the content of this presentation, and make no endorsement of it; errors, misstatements of facts, misinterpretations, and other scholarly shortcomings rest solely with G. Weismer
Some Advice I Respect

“I try not to become too regular an addict of any one subculture.”

-- Jonathan Lethem, from a 2004 interview in The Guardian
What is the Central Problem of Today’s Talk?

The Basis of Speech Intelligibility Deficits in Persons with Speech Disorders

What does it have to do with technique, concept, theory??
Technique, Concept, Theory

When the pursuit of technique dominates, a concern with process and steady accumulation is superceded by a means-end worldview.


Example: Identifying all aspects of segmental acoustics will permit them to be matched to the auditory system’s capabilities, and therefore expose the basis of speech intelligibility

Concept is an idea about how something might work, and is generally confined to a global notion

Example: Speech intelligibility is as much in the ear of the listener as it is in the mouth of the speaker
Theory is a set of axioms that explains multiple phenomena, the separate axioms often seeming unrelated; when taken together, however, the axioms account for a broad range of events, also unrelated

1. Speech perception is dependent on listener processes that are geared to access words

2. Speech motor control takes the form of movement-acoustic signal trajectories, learned as interlocked, calibrated actions, which are nonetheless subject to short-term (on-the-fly) revision or long-term recalibration, even in adulthood

3. Speech intelligibility among a group of speakers, whether typical or with a speech disorder, is variable with respect to absolute and relative ranking dependent on both speaker and listener variables
A Great Concept:

Stetson: “Speech as a set of movements made audible”

-- “set”: decomposable constituents

-- “movements”: motions, the things easily measurable, such as displacement (and its time derivatives), and ‘magic moments in time and space’ (and co-occurrence)

→ configurations, tube shapes, transforms of which took a back seat to those things that might reveal mechanism

-- “made audible”: movements happen to make sound
Another Great Concept:

Stetson’s Nemesis Notion: “...a set of sounds produced by movements.”

-- “set”: refers to the products, not the producer(s)

-- “sounds”: implies a contrasting set of acoustic events, or a structuring of such events, that humans find useful for communication

Stetson, like some before him (like Froeschels, for example) did not want to put the cart before the horse
Concept

Fant, Stevens

The speech mechanism is exploited to produce language-relevant signals, via movements recruited in the service of contrasts and other language-relevant structures

➢ The speech signal is a code related only incidentally to the fact that movements produce it; that is not to say the movements are not important, but their own “primitives” (e.g., displacements, speeds) do not determine the structure of the sound set

- Examples:
  - the *segment duration rules* proposed by Klatt (and others)
  - the *variation of vowel systems and consonant contrasts* both across languages and within languages (dialects)
  - the *variation of rhythmic structures* of natural-language utterances
  - the decision to use or not to use *tonal contrasts* to effect meaning

*These are not reducible to “movements made audible”.* They are *produced* by “movements made audible”, but not dependent on them
Keyser & Stevens (2006) got right down to business declaring their theory, the concept subsumed within that theory, and perhaps the technique implied by both (that is, linguistic technique):

“We assume as a starting point the correctness of distinctive feature theory”

-- p. 33

Bold, Clean, Crisp.

Back to Lethem:

“I try not to become too regular an addict of any one subculture.”
EMG
Oromotor nonverbal
Speech movement
Palatographic
Aerodynamic

Segmental Measures
--Vowels
--Transitions
--Consonants

Rate Measures
Voice measures
Rhythmic measures

Intelligibility Measures
--Closed set
--Transcription
--Scaling
--Comparison (ABX)
Error Analysis
Mayo Dimensions

PHYSIOLOGICAL
ACOUSTIC
PERCEPTUAL
The Acoustic Basis of Speech Intelligibility Variation

→ *Vowel Space*

→ *F2 Slope*
Diphthong /ai/

Cumulative Probability

F2 Slope (Hz/ms)

0 2 4 6 8 10 12 14

PD
als vs Col 2
Control
Graph showing cumulative probability against F2 Slope (Hz/ms) for control and dysarthric conditions.
/w/

Cumulative Probability

F2 Slope (Hz/ms)

Control F2
Dysarthric F2
Transition Duration (ms) vs. Transition Extent (Hz) for labial obstruents /aI/.

- **Control**
- **ALS**
- **PD**

Data points for each group are distributed as follows:
- **Control**: Smaller cluster around 50 ms duration and 150 Hz extent.
- **ALS**: Larger cluster with higher transition duration and extent, positioned around 200 ms and 400 Hz.
- **PD**: Data point isolated at higher transition duration and extent, around 250 ms and 800 Hz.

The graph highlights differences in articulatory behavior between the groups, with ALS showing a more extended range compared to control and PD.
The Ultimate Concept of Phonetic Reduction:

**Area of the Vowel Space: Corner Vowels (triangular or quadrilateral)**

We all know that, when compared to properly matched “normal” speakers, vowels spaces are smaller for speakers with:

- Parkinson disease, ALS, Stroke, CP, Cerebellar Disease, TBI,...and:
- Hearing impairment and those with cochlear implants,...and:
- Children with Down syndrome
- Persons who stutter
- Persons who are blind
- Persons with treated oral cancers
- Persons with traumatic brain injury...and even:
- Persons with recurrent nerve paralysis
- Persons with Muscular Tension Dysphonia
- Persons with Laryngectomy
We are also faced with somewhat disconcerting observation that the size of the vowel space can vary a lot within the same person, under different speaking conditions, with no apparent loss of speech intelligibility.

Mean F1-F2 Temporal Midpoint: tgt20

Mean F1-F2 Temporal Midpoint: tgt13

Mean F1-F2 Temporal Midpoint: tgt20

A lip rounding chill-out...or is it??
You're just a bunch of stupid consonants! And no one who plays Scrabble likes any of you!!

Irritable Vowel Syndrome
Where are we?

1. We have an idea that many speakers with neuromotor speech disorders produce articulatory reduction, manifest acoustically in different ways (reduced vowel space, reduced F2 slopes, reduced interquartile ranges for selected formants);

2. We do not know much about the nature of reduction as a *speech movement* phenomenon, e.g., if it is spread more or less uniformly across different parts of the speech mechanism, how much it varies in the “typical” population, and if there is a way to express the relationship, across the bridge of the speech acoustic signal, between movement measures and measures of speech intelligibility;

3. We *suspect* a gross measure of articulatory “working space” is related to movement-acoustic mappings for *specific segments*, but also *suspect* there is some independence between the two types of measures
**Theory**: Direct link between articulation and speech intelligibility

→ overall reduction of articulatory ranges
→ inability to make “fine” adjustments
→ poorly controlled “waxing and waning” of articulatory events

**Concept**: Relate segment-sized movements, global indices of articulation, and possibly measures of time-dependent, articulatory “hotspots” and coordination to speech intelligibility, to determine how much variance in intelligibility measures can be accounted for by the former measures

**Technique**: Speech movements; inferences from the speech acoustic signal; comprehensive perceptual material; multivariate statistical procedures
“To feed the cat one must shoo the dog”
4 neurologically-healthy speakers
“To feed the cat one must shoo the dog”
4 speakers with PD and near-normal word intelligibility
“To feed the cat one must shoo the dog”
4 speakers with ALS and near-normal word intelligibility
Back to the Early Days of Speech Intelligibility

Speech Intelligibility = \((a_1, \ldots, n)\text{SEGMENTALS}_{1, \ldots, n} + (b)\text{RES}\)

where \((a_1, \ldots, n)\) are weights for segments 1, \ldots, n (different vowels, nasals, semivowels, rhotics, obstruents).

*The inclusion of different weights is theoretical—different sound classes might make different contributions to speech intelligibility

RES, at this point, is a wastebasket variable for error variance
Speech Intelligibility = \( (a_1, \ldots, n)_{SEGMENTALS} + (b)_{RES} \)

where \((a_1, \ldots, n)\) are weights for segments 1,\ldots,n (different vowels, nasals, semivowels, rhotics, obstruents).

This just doesn’t work so well:

1) Multivariate models predict too much variance in intelligibility scores with too few acoustic variables—sometimes a *single* variable!

2) Many of the predictors are highly intercorrelated, even when there seems to be no reason for them to be (such as VOT for a specific place of articulation and the F2-F1 difference for a particular vowel)

→ In a chapter we wrote long ago (Weismer & Martin, 1992), these kinds of results made us suspicious about 3\(^{rd}\) variable effects, and the difference between specific acoustic variables as components of speech intelligibility, versus reflections of overall speech severity
Intelligibility = (a) $PWS + (b_{1...n})SEGMENTALS_{1...n} + (c) RES$

FIGURE 7

Correlation coefficients, within groups:
- for 20 control talkers: $r = 0.97$
- for 16 talkers with PD: $r = 0.63$
- for 9 talkers with ALS: $r = 0.93$

Working space Area for coupled T3 (in sq mm) vs. Average Speed for coupled T3 (mm/s)
Figure C-5

$\text{r}_{\text{all}} = 0.90$

$\text{r}_{\text{als}} = 0.90$

$\text{r}_{\text{pd}} = 0.66$

$\text{r}_{\text{c}} = 0.62$
Average speed (mm/s) for T3 marker

Average Sentence-based Intelligibility (scaled DME units)

- r for 20 control talkers (△) = 0.42
- r for 16 talkers with PD (○) = 0.53
- r for 9 talkers with ALS (■) = 0.80
45 speakers, all (PD, ALS, Control) combined

$r = 0.67$

$r^2 = 45\%$
Marker Working Space (mm²)

BLUE: UNCOUPLED
BLACK: COUPLED
Intelligibility = (a) $PWS + (b_{1...n})SEGMENTALS_{1...n} + (c) Boundary Goodness + (d) PROS, + (e) RES
Where are We?

1. Need descriptive data on speech movement

2. Need studies in which manipulations such as rate, stress, loudness are analyzed jointly with movement, acoustics, perception (this “controls” the 3rd variable issue

3. Need to learn more about the flow of speech movements and the flow of acoustic “hotspots”, which presumable are locked onto by speech perception mechanisms