

Vocal Charades

An empirical study of iconic production in the vocal modality



Marcus Perlman

Dept. of Cognitive & Information Sciences
University of California, Merced, CA USA

Where do languages come from?

Signed languages originate from created deictic and iconic gestures, which become conventionalized, ritualized, grammaticalized, etc. (See Armstrong & Wilcox 2007 for review)

- Iconicity erodes (to some degree) with time

Similar process for written systems.

But where do spoken languages come from?

How do spoken gestures originate?

1. Vocal origin folks don't really seem to think about it much
 - Critical mass of innate alarm calls (?)
2. Gesture origin folks assume spoken gestures were primarily boot-strapped on motivated manual gestures
 - Vague hand-waving towards onomatopoeia (esp. animals) and emotional sounds

Tomasello (2008: 228)



Thought experiment:

Two groups of children are each alone on an island, one communicates with gestures, the other with vocalizations...

What happens?

Tomasello (2008: 228)



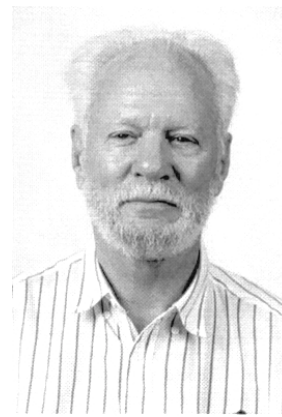
“It is difficult to imagine [the children] inventing on their own vocalizations to refer the attention or imagination of others to the world in meaningful ways—beyond perhaps a few vocalizations tied to emotional situations and/or a few instances of vocal mimicry. Humans have no natural tendencies in the vocal modality—analogous to following gaze directionally in space or interpreting actions as intentional in the gestural/visual modality—to serve as starting points. And so the issue of conventionalizing already meaningful communicative acts never arises.”

Tomasello (2008: 228)



“It is difficult to imagine [the children] inventing on their own vocalizations to refer the attention or imagination of others to the world in meaningful ways—beyond perhaps a few vocalizations tied to emotional situations and/or a few instances of vocal mimicry. Humans have no natural tendencies in the vocal modality—analogous to following gaze directionally in space or interpreting actions as intentional in the gestural/visual modality—to serve as starting points. And so the issue of conventionalizing already meaningful communicative acts never arises.”

Hockett (1978: 274)



“When a representation of some four-dimensional hunk of life has to be compressed into the single dimension of speech, most iconicity is necessarily squeezed out. In one-dimensional projection, an elephant is indistinguishable from a woodshed. Speech perforce is largely arbitrary.”



Pinker & Jackendoff (2005: 209)

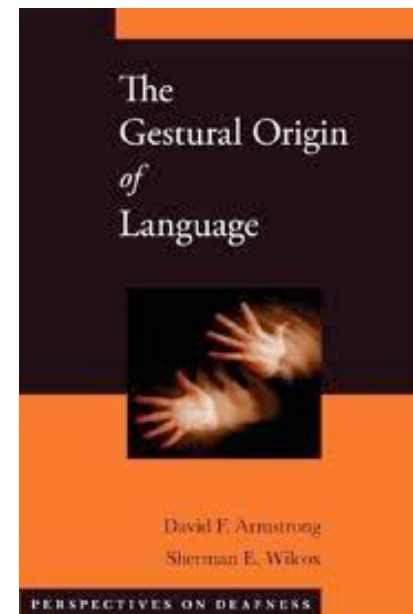


“Humans are not notably talented at vocal imitation in general, only at imitating speech sounds (and perhaps melodies). For example, most humans lack the ability (found in some birds) to convincingly reproduce environmental sounds... Thus ‘capacity for vocal imitation’ in humans might better be described as a capacity to learn to produce speech.”

Armstrong & Wilcox (2007: 123)

“Visual representation can be expected to precede auditory representation because of the vastly greater possibility for iconic productivity in the visual medium.”

How true is this claim?



Hypotheses

Hypothesis 1: The vocal modality has *extremely limited* potential for iconic productivity compared to the manual modality.

Hypothesis 2: The vocal modality has *just as much* potential for iconic productivity compared to the manual modality.

Documented domains of vocal iconicity (to name a few)

Sound symbolism

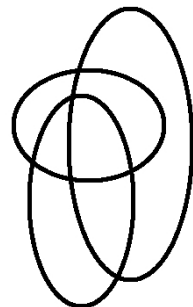
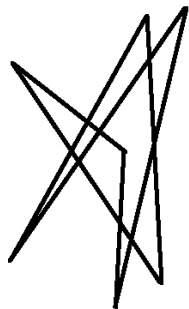
- Size, shape, brightness, gender, distance

Onomatopoeia (prevalent in many non-IE languages)

- Shape, manner of motion, texture, sensation, size, sound, temporal aspect

Iconic prosody

- Emotion, speed, length, size, verticality



Kiki

Bauba

See Perniss et al. 2010
for a review

Vocal Charades

An empirical study of iconic productivity in the vocal modality



Do people have consistent intuitions for producing iconic sounds across different semantic domains?

Method – the rules of Vocal Charades

- Pairs played charades-style game (N = 15 pairs)
- Each player given stack of note cards
 - 30 paired antonymic words mixed up per stack (e.g. bright/dark, big small, up/down)
- Players took turns (10 at a time) using only their voice to express the word on the card as their partner tried to guess
 - No manual gestures or words allowed (partner blindfolded)
 - 20 seconds per turn

30 Word Pairs

Alive / Dead

Antagonistic /
Friendly

Attractive / Ugly

Bad / Good

Big / Small

Bright / Dark

Cold / Hot

Difficult / Easy

Down / Up

Dry / Wet

Dull / Sharp

Fast / Slow

Female / Male

Few / Many

Hard / Soft

Heavy / Light-weight

Here / There

Last year /

Next year

Lift up / Set down

Long / Short

New / Old

No / Yes

Now / Later

Nutritious /

Poisonous

Predator / Prey

Rough / Smooth

Start / Stop

Straight / To the
side

Strong / Weak

Surprising /
Predictable

Analysis

- Used Praat to make acoustic measurements
- Dependent measures included mean pitch, maximum pitch change, absolute pitch change, intensity, duration, harmonics to noise ratio (HNR), repetition rate
- Multiple sounds per turn averaged together
- Compared paired opposite words with paired sample t-tests

Overall results











- 20/30 word pairs significant for at least one acoustic property (after Bonferroni correction on each word pair for number of variables tested)
- 12/20 significant pairs included unique combinations of properties
- Significant pairs included all seven properties

Proportion of word pairs significant by criterion for each dependent measure

Variable	< 0.1	< 0.05	< 0.01	< 0.001
Mean Pitch (30)	0.60 (18)	0.47 (14)	0.30 (9)	0.067 (2)
Max Pitch Change (28)	0.21 (6)	0.14 (4)	0.036 (1)	0.036 (1)
Abs Max Pitch Change (28)	0.46 (13)	0.29 (8)	0.036 (1)	0.036 (1)
Duration (30)	0.40 (12)	0.37 (11)	0.20 (6)	0.067 (2)
Intensity (30)	0.37 (11)	0.27 (8)	0.20 (6)	0.067 (2)
Hnr (30)	0.57 (17)	0.53 (16)	0.33 (10)	0.13 (4)
Repetition Rate (2)	1.0 (2)	1.0 (2)	0.50 (1)	0.00 (0)
Total (178)	0.44 (79)	0.35 (63)	0.19 (34)	0.067 (12)











Word pair characteristics

(bold indicates $p < 0.05$ after Bonferroni correction)

Word Pair	Primary (< 0.001)	Strong (< 0.01)	Moderate (< 0.05)	Marginal (< 0.1)
 Alive	Intensity \uparrow			Abs. pitch chg. \uparrow
 Dead	Hnr \uparrow			
 Antagonistic		Hnr \downarrow		
 Friendly				
 Bad	Hnr \downarrow	Mean pitch \downarrow		
 Good	Abs. pitch chg. \downarrow	Duration \downarrow		
 Big		Mean pitch \downarrow	Duration \uparrow	
 Small		Hnr \downarrow Intensity \uparrow		
 Bright		Mean pitch \uparrow	Intensity \uparrow	Abs. pitch ch. \uparrow
 Dark			Max pitch ch. \uparrow	











Word pair characteristics

(bold indicates $p < 0.05$ after Bonferroni-Holmes correction)

Word Pair	Primary (< 0.001)	Strong (< 0.01)	Moderate (< 0.05)	Marginal (< 0.1)
 Cold  Hot		Hnr ↑	Mean pitch ↓ Duration ↑ Abs. pitch ch. ↓	
 Difficult  Easy	Duration ↑			Mean pitch ↓
 Down  Up	Max pitch ch. ↓			Mean pitch ↓
 Dull  Sharp		Mean pitch ↓		Intensity ↓
 Fast  Slow		Mean pitch ↑	Intensity ↑ Hnr ↓ Rep. rate ↑ Duration ↓	











Word pair characteristics

(bold indicates $p < 0.05$ after Bonferroni-Holmes correction)

Word Pair	Primary (< 0.001)	Strong (< 0.01)	Moderate (< 0.05)	Marginal (< 0.1)
 Female  Male	Mean pitch ↑	Hnr ↑		Abs. pitch ch. ↑
 Few  Many		Rep. rate ↓		
 Long  Short	Duration ↑			Mean pitch ↓
 New  Old		Hnr ↑ Intensity ↑	Duration ↓	Mean pitch ↑ Abs. pitch ch. ↑ Max pitch ch. ↓
 No  Yes		Mean pitch ↓	Abs. pitch ch. ↓ Hnr ↓	

Word pair characteristics

(bold indicates $p < 0.05$ after Bonferroni-Holmes correction)

Word Pair	Primary (< 0.001)	Strong (< 0.01)	Moderate (< 0.05)	Marginal (< 0.1)
 Now  Later		Duration ↓ Intensity ↑ Hnr ↓		
 Nutritious  Poisonous	Hnr ↑		Mean pitch ↓ Intensity ↑	
 Rough  Smooth	Hnr ↓		Max. pitch ch. ↑ Abs. pitch ch. ↓ Mean pitch ↓	
 Strong  Weak	Intensity ↑		Mean pitch ↓	
 Surprising  Predictable	Mean pitch ↑	Intensity ↑	Abs. pitch ch. ↑ Hnr ↓	Duration ↓

Summary of results



- Participants showed consistent intuitions for how to produce iconic vocalizations for a variety of meanings
 - 20/30 pairs
- Participants made use of several acoustic properties (or “dimensions” cf. Hockett) in expressing these meanings
 - All 7 properties, 12 unique combinations

Which hypothesis is correct?

Hypothesis 1: The vocal modality has *extremely limited* potential for iconic productivity compared to the manual modality.

Hypothesis 2: The vocal modality has *just as much* potential for iconic productivity compared to the manual modality.

Which hypothesis is correct?

Hypothesis 1: The vocal modality has *extremely limited* potential for iconic productivity compared to the manual modality.

Hypothesis 2: The vocal modality has *just as much* potential for iconic productivity compared to the manual modality.

Just as much? At least, iconic potential in the vocal modality is vastly underestimated.

Languages are multimodal to the core

- Evidence warrants serious consideration of the role of vocal iconicity in the original and ongoing development of languages
 - The modality “either/or” may be a false dichotomy
- The world is typically visible *and* audible, and in reflection of this, human languages are typically visible and audible too





<https://www.youtube.com/watch?v=TryXVjR7mwY>