Redundancy in Sound Systems

Elizabeth Hume

Rory Turnbull

The Ohio State University

How can information theory inform the study of sound systems?

- In general terms, our research interests include understanding:
  - the physical and cognitive factors that shape sound patterns
  - the connection between the continuous speech stream and the abstract representation of the knowledge that an individual has about sound patterns in their language, i.e. What is the source of the variation in the signal that serves as the basis for phonological patterns?
- It is from this perspective that we consider potential contributions of the concepts and tools of information theory.

One contribution: Redundancy

- It's from our perspectives as phonologists/ phoneticians that we consider how information theory can inform our understanding of sound systems.
- Today we focus on one concept of information theory: redundancy, which we take to mean the recapitulation of elements already present or known.

Shannon's coding theorems
(from Roger Levy's talk yesterday)

Shannon's *source coding theorem* places limits (upper and lower) on lossless compression

In the limit, a message *w* is encodable in $\log \frac{1}{P(w)}$ bits

Low-probability messages require more *resources*

Shannon's *channel coding theorem* places limits (upper and lower) on transmission rates through noise

Ideal coding reaches channel capacity *C*, never higher

Noisy-channel coding requires *redundancy*

Redundancy and phonological theories

- The concept of redundancy has played an key role in phonological theories in a couple of ways:
  - the reduction or elimination of redundancy primarily as a means of achieving representational simplicity.
  - the addition of redundancy in order to maximize the distinctiveness among sounds.

Phonological Theory: Reducing redundancy

- The formalization of redundancy in phonological theory can be traced back at least to work by Roman Jakobson, inspired by work in communication and information theories.
- Cherry, Halle & Jakobson (1953) propose an information theoretic analysis of distinctive features. They show that feature values have different functions and that some are redundant. They formalize the notion of redundancy in phonology by providing a mechanism for reducing the overall specification of sounds, in the interests of representational efficiency (also in Jakobson, Fant & Halle 1952).
- To illustrate, they begin by showing that if 11 distinctive features are used to classify the sound system of Russian, 11 bits of information are needed to classify the system.
- By removing redundant feature values, the number of bits drop to 6.5 per phoneme.

Partial List of Russian Phonemes

	k	k,	g	g,	x	c	ʃ	ʒ	t	t,	d	d,	s	s,	z	z,	ŝ	n	n,	p	p,
VOCALIC	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
CONSONANTAL	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
COMPACT	+	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−
DIFFUSE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GRAVE	+	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+
NASAL	0	0	0	0	0	0	0	0	−	−	−	−	−	−	−	−	−	+	+	−	−
CONTINUANT	−	−	−	−	+	−	+	+	−	−	−	−	+	+	+	+	−	0	0	−	−
VOICED	−	−	+	+	0	0	−	+	−	−	+	+	−	−	+	+	−	−	−	−	−
SHARP	−	+	−	+	0	0	0	0	−	+	−	+	−	+	−	+	0	−	+	−	+
STRIDENT	0	0	0	0	0	0	0	0	−	−	−	−	0	0	0	0	+	0	0	0	0
STRESSED	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Redundant values are expressed as ‘0’

Partial List of Russian Phonemes

	k	k,	g	g,	x	c	ʃ	ʒ	t	t,	d	d,	s	s,	z	z,	ŝ	n	n,	p	p,
VOCALIC	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
CONSONANTAL	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
COMPACT	+	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−
DIFFUSE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GRAVE	+	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+
NASAL	0	0	0	0	0	0	0	0	−	−	−	−	−	−	−	−	−	+	+	−	−
CONTINUANT	−	−	−	−	+	−	+	+	−	−	−	−	+	+	+	+	−	0	0	−	−
VOICED	−	−	+	+	0	0	−	+	−	−	+	+	−	−	+	+	−	−	−	−	−
SHARP	−	+	−	+	0	0	0	0	−	+	−	+	−	+	−	+	0	−	+	−	+
STRIDENT	0	0	0	0	0	0	0	0	−	−	−	−	0	0	0	0	+	0	0	0	0
STRESSED	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Partial List of Russian phonemes

	k	k,	g	g,	x	c	ʃ	ʒ	t	t,	d	d,	s	s,	z	z,	ŝ	n	n,	p	p,
VOCALIC	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
CONSONANTAL	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
COMPACT	+	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−
DIFFUSE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GRAVE	+	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+
NASAL	0	0	0	0	0	0	0	0	−	−	−	−	−	−	−	−	−	+	+	−	−
CONTINUANT	−	−	−	−	+	−	+	+	−	−	−	−	+	+	+	+	−	0	0	−	−
VOICED	−	−	+	+	0	0	−	+	−	−	+	+	−	−	+	+	−	−	−	−	−
SHARP	−	+	−	+	0	0	0	0	−	+	−	+	−	+	−	+	0	−	+	−	+
STRIDENT	0	0	0	0	0	0	0	0	−	−	−	−	0	0	0	0	+	0	0	0	0
STRESSED	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Redundant value of [stressed]:
consonants are predictably unstressed

Partial List of Russian phonemes

	k	k,	g	g,	x	c	ʃ	ʒ	t	t,	d	d,	s	s,	z	z,	ŝ	n	n,	p	p,
VOCALIC	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
CONSONANTAL	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
COMPACT	+	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−
DIFFUSE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GRAVE	+	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+
NASAL	0	0	0	0	0	0	0	0	−	−	−	−	−	−	−	−	−	+	+	−	−
CONTINUANT	−	−	−	−	+	−	+	+	−	−	−	−	+	+	+	+	−	0	0	−	−
VOICED	−	−	+	+	0	0	−	+	−	−	+	+	−	−	+	+	−	−	−	−	−
SHARP	−	+	−	+	0	0	0	0	−	+	−	+	−	+	−	+	0	−	+	−	+
STRIDENT	0	0	0	0	0	0	0	0	−	−	−	−	0	0	0	0	+	0	0	0	0
STRESSED	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Redundant value of [strident]: sibilants are predictably [+strident], given their specification for the feature values [+cont, -grave]

Representational efficiency

This approach formed the basis of a large body of research on underspecification, feature organization, and other methods of achieving representational simplicity, e.g. Archangeli 1984; Archangeli & Pulleyblank 1989; Avery & Rice 1989; Broe 1992; Clements 1985; Kiparsky 1985; Pulleyblank 1986.

↓ Redundancy: Feature economy

Clements (2005: 50)

"languages tend to organize their sound systems by maximizing the use of a small number of highly valued features"

Again, redundancy is minimized to achieve representational simplicity.

Phonological Theory: Redundancy Reduction

- These early ideas laid the groundwork for many fundamental aspects of phonological theory, particularly with regards to reducing redundancy as a means of creating an efficient system, where simplicity is correlated with efficiency.

- Building on the work of, in particular, Lindblom (e.g. 1986) on the theory of hypo- and hyper-articulation, the concept of adding redundancy to representations also became a part of phonological theories.

↑ Redundancy

• Increasing redundancy has generally been studied in the context of maximizing the distinctiveness among sounds at the phonetic level.

- Phonetic enhancement involves introducing an additional articulation

  - dispersion (e.g. Lindblom 1986)

  - trading relations (e.g. Perkell et al 1993)

↑ Redundancy

- Proposals for incorporating ideas about phonetic enhancement into the phonology include:
  - introducing a redundant feature (e.g. Stevens, Keyser & Kawasaki 1986, Stevens & Keyser 1989)
  - positing specific constraint types and rankings (e.g. J. Beckman 1997, Boersma 1997, Flemming 2002, Steriade 1995, 2000)

Connection to Information Theory

- These lines of research highlight some ways in which the concept of redundancy has figured into the study of sound systems, in terms of decreasing it, on the one hand, or increasing it on the other.
- In this talk we build on this previous work on redundancy in phonology and illustrate their close connection to fundamental aspects of communication and information theories.
- It turns out that redundancy plays an even more central role than what we as phonologists have been giving it credit for.

Connection to Information Theory

In what follows, we hope to show that redundancy can be viewed as:

- a unifying concept in the study of sound systems and
- a powerful source of explanation for phonetic variability and the phonological patterns emerging from this variability.

Outline

- communication systems, coding and redundancy
- manipulating redundancy in speech
- redundancy-based accounts of familiar phonological processes

Communication

- The uses of redundancy in phonetics and phonological are largely consistent with the function of redundancy in information theory.
- These ideas follow naturally from the properties of communication systems.

Effective communication

- In communication and information theories, two central properties of an effective communication system are that it is **efficient** and **robust**.
- Efficiency is concerned with representing a message with the fewest symbols.
  - A code that represents its meaning with few symbols is more efficient than one that represents it with many symbols.
  - In terms of efficiency, redundancy is a hindrance.

Robustness

Robustness is the ability of a message to be transmitted correctly: no input output mismatch. In order for this to happen, the signal must be **robust** to interference.

In terms of robustness, redundancy is beneficial.

Efficiency vs. Robustness

meaning	**X**	**Y**	**Z**
dog	ABC	A	A
cat	BCA	B	A
mouse	CAB	C	A

robust
(lower efficiency)

efficient
(lacks robustness)

Errors

- One source of errors is “noise”: anything in the signal that is not relevant to the message that can obscure or alter it.

  e.g. a passing truck, other talkers, yawning, etc.

Redundancy and errors

- Suppose we wish to transmit a message along an electronic channel, through a series of electric impulses representing 1s and 0s.
- We can start with a code that matches letters to sequences of binary digits
- 00000 A
- 00001 B
- 00010 C
- etc...

Errors

- To transmit the message “cat”, we could send the binary digits corresponding to the symbols CAT:

  00010 00000 10100

  C A T

- If there is noise in the channel, we may get an error:

  00010 0**1**000 10100

  C H T

Error Detection

- To increase the likelihood of correct transmission, we can add **redundancy** to the signal by sending each symbol twice:

  CCAATT

- This way, if there are unpaired symbols, we’ll know there is an error:

  C**S**AATT

Error Correction

- If we send each symbols thrice, we can detect and correct errors:

  Intended message: CCCAAATTT

  Transmitted message: CCCAAA**D**TT

- We can recover the correct symbol T given the prior pattern and redundant symbols present.

- By incorporating redundancy, we increase the predictability of the element, i.e. information/plog/ surprisal is lowered.

Reducing Redundancy

Sometimes errors are not very likely due to,

- few possible outcomes (low entropy of event)
- element is predictable based on other factors

> I have a cute, fluffy pet called Snowball.
>
> Whenever I see my CAT, I smile.

Using CCCAAATTT is unnecessary.

With sufficient higher-level predictability, we do not need to depend as much on redundancy in the element itself in order to identify the message. Redundancy can be reduced.

Redundancy in Speech

These principles also apply to speech.

Yet, speech is not a linear sequence of electric pulses.

Rather it is a complex, continuous waveform that allows simultaneity of expression.

Horizontality and Verticality

- We can have **horizontal** redundancy:

CCCAAATTT

- And **vertical** redundancy:

C A T

C A T

C A T

Form and meaning

- Each of these symbols represents a piece of information about the identity of the phoneme.
  - C - “my identity is a voiceless velar stop”
- However, the **form** of these symbols may vary.
  - § - “my identity is an item of furniture that seats more than one person”
  - § may be realized as “couch”, “sofa”, “chesterfield”, etc.

Phonetic Redundancy

- Similarly, each symbol can have a different phonetic realization.
- Thus, **CCC** is not [kkk], but rather each symbol could represent a different phonetic cue to the identity of the phoneme.
- Cues are realized both sequentially (horizontally) and simultaneously (vertically)
- When there are multiple cues to the identity of a phoneme, these cues can be considered redundant (recapitulated elements).

Phonetic Redundancy

[Vt̚] [Vt] [Vtʰ]

less redundant

fewer cues

more redundant

more cues

“The term ‘redundancy’ should not be taken to imply wastefulness; it is a property of speech, and a fact of every system of communication, which serves a most useful purpose. In particular, it helps the hearer to resolve uncertainties introduced by distortion in the signal or by disturbing noises.”

(Cherry, Halle & Jakobson 1953: 39)

Effective communication

"The whole problem of efficient and error-free communication turns out to be that of removing from messages the somewhat inefficient redundancy which they have and then adding redundancy of the right sort in order to allow corrections of errors made in transmission."

Pierce (1961:164)

Predictions

Redundancy in sound systems is manipulated in continuous response to the needs of efficiency and robustness in communicating a message (consistent with Lindblom's 1986 H&H model).

- When robustness is an issue, increasing redundancy is beneficial.

- When robustness is not an issue, decreasing redundancy is beneficial in the interests of increasing efficiency.

Hazan and Simpson (2000)

- Nonsense VCV sequences presented in noise to native English speakers
  - 73.8% consonant intelligibility
- VCV transitions enhanced (amplitude boosted between 2 and 12dB)
  - 82.9% consonant intelligibility
- Cue enhancement – i.e. increasing redundancy – leads to greater intelligibility.

Redundancy as a unifying concept

The recapitulation of elements already present or known.

- can be predicted by other elements in the signal, e.g.
  - multiple phonetic cues, simultaneous or sequential, e.g. [a**tʰ**u]
- or is part of prior knowledge, e.g.
  - frequency of occurrence (higher)
  - pragmatic information (e.g. prior mention)
  - social meaning

Some research questions

- what phonetic form does increased/decreased redundancy take?
- what are the phonological consequences of increased/ decreased redundancy?
- under what conditions is redundancy increased/decreased at the phone/phoneme level?
- over what domains is redundancy calculated? (e.g. segmental, prosodic, lexical, syntactic, pragmatic, social)
- intra-level redundancy: how does the redundancy of elements on the same level interact?
- inter-level redundancy: to what extent does redundancy at a given level influence the manipulation of redundancy at another level?

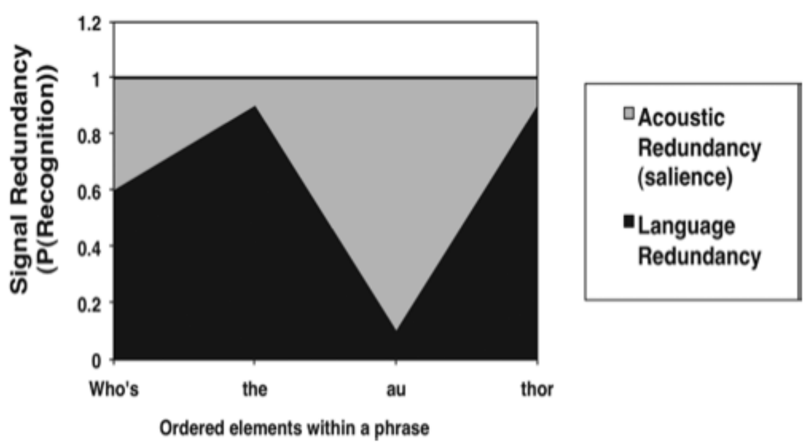
Relation to previous research

- Clearly, there is a considerable body of existing and ongoing research that bears on many of these questions, though not necessarily under the guise of investigating the role redundancy (including the work by many of you in this room!).
- We begin with an example of how redundancy is manipulated at the phonetic level which provides evidence for some of the questions just noted.

Manipulating acoustic redundancy

- Aylett & Turk (2004), phonetic substance (syllable duration) is increased or decreased based on higher level redundancy.
- Higher-level redundancy calculated at multiple levels:
  - word frequency
  - syllable trigram (from British National Corpus): transitional probability of guessing 3rd syllable from two preceding ones.
  - discourse level: givenness, how many times previously measured
- Greater predictability of the syllable based on these factors means that redundancy at the phone level can be decreased.

Acoustic vs. Higher-level Redundancy (from Turk 2010)



See, e.g. Jaeger 2006, Levy & Jaeger 2007 for similar findings regarding syntactic structure.

Phonological consequences of manipulating redundancy

- Given that the manipulation of redundancy can be implemented through phonetic variation, we would expect there to be phonological consequences.
- In fact, taken to the extreme we could view most (all?) phonological patterns as the phonologization of variability resulting from the manipulation of redundancy.

Phonological consequences

The link between decreasing redundancy and phonology processes is well studied.

- reduction, deletion, lenition, neutralization

We focus instead on some of the phonological reflexes of increasing redundancy.

- nasal place assimilation in Japanese
- vowel epenthesis in English
- diachronic chain shift

Japanese nasal place assimilation

Observations:

- A nasal consonant is obligatorily homorganic with a following onset stop (Vance 1987). That is, a nasal consonant assimilates in place to a following stop consonant.
- When not followed by a consonant, the nasal is pronounced as uvular.

saɴ	‘three’
samban	‘number three’
sandan	‘three steps’
saŋko	‘three objects’

from Tsujimura (2007)

Redundancy-based approach

Background:

- The quality of the nasal is constrained. There are 4 phonetic nasals in the language [labial, coronal, dorsal, uvular). In word-final position, only the uvular nasal can occur.
  - The entropy of selecting a word-final consonant in Japanese (uvular or null) is 0.536 bits (based on token frequencies derived from the NTT corpus of Japanese).
- There is greater uncertainty when it comes to identifying among the set of onset stops (b,d,g,p,t,k) with an entropy of 2.429 bits (measured over word-initial stops).

Manipulating redundancy

- Increasing the redundancy of an individual onset consonant makes it more identifiable (more predictable, plog/information/surprisal).
- Adding redundancy in this case takes the form of place assimilation; that is, the onset stop's identity can also be predicted on the basis of the place cues in the preceding nasal.
- Consequence for the nasal: less predictable (no longer always uvular). Information (plog/ surprisal) increases.

Uniform Information Density

Consistent with Florian Jaeger's and Roger Levy's proposal concerning UID.

Information is being spread out more uniformly in Japanese nasal+stop clusters due to place assimilation: information of onset is lowered, information of coda is increased.

How does this help the listener? By adding redundancy on the coda, it gives more information about the onset.

Supported by experimental results.

Processing advantage

- Otake et al. (1996) investigated the role of nasal place assimilation on the processing of place in a following stop consonant by Japanese listeners. They found that their subjects made use of place cues in a nasal consonant to obtain information about the place of articulation of a following stop.
- The authors conclude that ‘place of articulation in a nasal is a reliable source of information about a following stop for Japanese listeners, and they make use of it’ (Otake et al. 1996:3841).
- Dutch listeners, in contrast, ignored place information in a preceding nasal when processing the place identity of a following stop consonant.

Dutch vs. Japanese

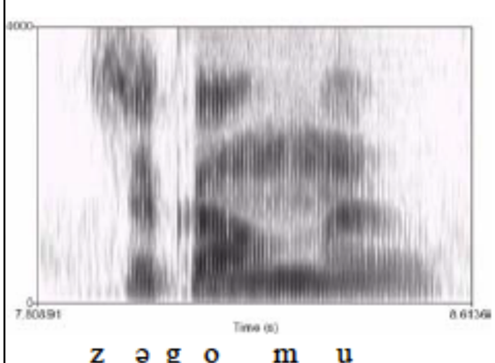
- As the authors point out, these findings reflect the different phonological status accorded place of articulation in preconsonantal nasals in the two languages.
- In Japanese, a nasal is obligatorily homorganic with a following stop. Conversely, while place assimilation does occur between a nasal-stop sequence in Dutch, it does not have the obligatory status it has in Japanese; assimilation fails to occur both within and across word and syllable boundaries.
- We suggest that nasal place assimilation can be viewed as a phonological consequence of increasing the redundancy of the onset stop consonant in order to boost robustness.

Epenthesis

We can take a similar approach to understanding excrescent vowel insertion, which is phonologized in some language contexts as epenthesis.

Davidson & Stone (2003) provide a study of the production of phonotactically illicit consonant clusters by speakers of American English. Stimuli included pseudo-Polish words, e.g. *zbura*, *zgomu*.

Production of *zgomu* as [z**əgomu]**



"the schwa present on the acoustic record follows from the hypothesis that speakers are pulling apart the /z/ and subsequent consonant to prevent their overlap, since they do not have experience with the coordination necessary for the appropriate production of a /zC/ word-initial cluster."(p.12)

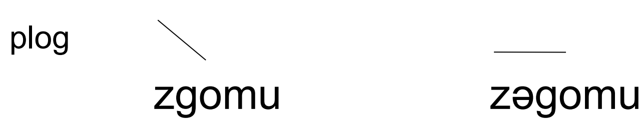
Redudancy approach: Vowel insertion in high information contexts

- #zg is non-occurring in English
  - high information/surprisal/plog
- By pulling apart consonant gestures, redundancy is added to the signal in the form of consonant release and formant structure between the consonants, both of which provide information about the place and manner of the adjacent consonants.
- Adding redundancy lowers the information of the consonant sequence (it is more predictable, conforms to patterns of the language, and is more robust).

Uniform Information Density

Again, this is consistent with Uniform Information Density when we take the following (vowel) context into account.

A rough visualization (not based on real plogs).



Prediction

Given that information is probabilistic, we would expect that the degree to which excrescent vowel insertion is used would depend on the plog of the context.

That is, contexts with higher plog would be more prone to have excrescent vowel insertion than those with lower plog.

(to be tested)

Chain Shifts

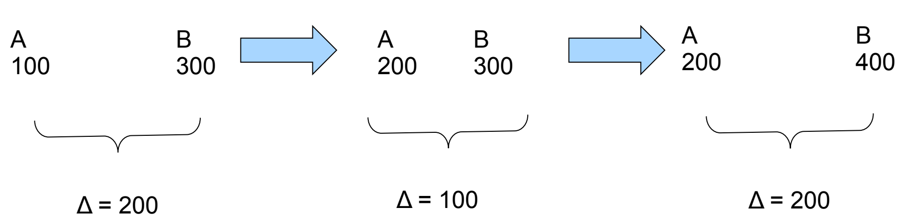
- Another example illustrating how redundancy can be manipulated comes from diachronic chain shifts.
- This may have the effect of keeping redundancy relatively constant among the relevant sounds over time.

jnd and Redundancy

- We start with the assumption that increased phonetic distance between sounds is a manifestation of redundancy.
- We can speak of phonetic distance in terms of *just noticeable difference* (jnd). This is the smallest detectable change between two physical stimuli.
- Suppose that the jnd for duration is 40ms
- Then any duration contrast that is greater than 40ms is redundant.
- The most efficient languages would have vowel length contrasts of 40ms, no more.

An example

- Suppose the phonemes A and B vary along some phonetic parameter space
- The signal redundancy is a function of the difference (Δ) in the parameter space.



Closing thoughts

Concepts from information theory can inform our understanding of language sound systems.

Redundancy can be viewed as a unifying concept.

In fact, grammatical sound patterns can be viewed as the phonologization of variability resulting from the manipulation of redundancy in the signal.

Acknowledgements

Kathleen Hall

Andy Wedel

Adam Ussishkin

Roger Levy