Minimal Pairs and Hyperarticulation of Singleton and Geminate Consonants as Enhancement of Lexical/Pragmatic Contrasts

Shin-ichiro Sano

Keio University
shinichirosano@gmail.com

NELS 48 @ University of Iceland
October 28, 2017
Synopsis

  - Informativity $\Rightarrow$ linguistic behavior

- This project
  - case study of MOP
  - durational contrast of singletons & geminates in spoken Japanese
  - lexical competition induces synchronic, phonetically specific hyperarticulation of phonemic contrasts (Wedel et al. 2013a, b)
  - sub-lexical (within-category) hyperarticulation in minimally contrasting singletons & geminates

- Confirmed that
  - singletons – shorter, geminates – longer
  - hyperarticulation – lexical & pragmatic (non-phonemic) contrasts
  - hyperarticulation $\Leftrightarrow$ informativity (Shannon’s entropy)

- Synchronic hyperarticulation – diachronic maintenance of phonemic contrasts
Synopsis

  - Informativity $\Rightarrow$ linguistic behavior

- **This project**
  - case study of MOP
  - durational contrast of singletons & geminates in spoken Japanese
  - lexical competition induces synchronic, phonetically specific hyperarticulation of phonemic contrasts (Wedel et al. 2013a, b)
  - sub-lexical (within-category) hyperarticulation in minimally contrasting singletons & geminates

- Confirmed that
  - singletons – shorter, geminates – longer
  - hyperarticulation – lexical & pragmatic (non-phonemic) contrasts
  - hyperarticulation $\Leftarrow$ informativity (Shannon’s entropy)

- **Synchronic hyperarticulation – diachronic maintenance of phonemic contrasts**
Synopsis

  - Informativity $\Rightarrow$ linguistic behavior

- This project
  - case study of MOP
  - durational contrast of singletons & geminates in spoken Japanese
  - lexical competition induces synchronic, phonetically specific hyperarticulation of phonemic contrasts (Wedel et al. 2013a, b)
  - sub-lexical (within-category) hyperarticulation in minimally contrasting singletons & geminates

- Confirmed that
  - singletons – shorter, geminates – longer
  - hyperarticulation – lexical & pragmatic (non-phonemic) contrasts
  - hyperarticulation $\Leftarrow$ informativity (Shannon’s entropy)

- Synchronic hyperarticulation – diachronic maintenance of phonemic contrasts
Roadmap

1 Background
   - Research topic
   - Message Oriented Phonology
   - Research objective

2 Method
   - Corpus
   - Data retrieval
   - Dataset
   - Statistical analysis

3 Results
   - Items to be examined
   - Duration of singletons/geminates
   - Patterns in phonetic implementation and informativity

4 Discussion
   - Contrastive hyperarticulation
   - Synchronic pattern – diachronic change
Roadmap

1 Background
- Research topic
  - Message Oriented Phonology
  - Research objective

2 Method
- Corpus
- Data retrieval
- Dataset
- Statistical analysis

3 Results
- Items to be examined
- Duration of singletons/geminates
- Patterns in phonetic implementation and informativity

4 Discussion
- Contrastive hyperarticulation
- Synchronic pattern – diachronic change
**Research topic**

**Length contrast and singleton & geminate**


- In Japanese, length or duration – important role lexically and pragmatically
- Modern Japanese has a variety of length contrasts.

**Vowel length**

<table>
<thead>
<tr>
<th>Short</th>
<th>Long</th>
</tr>
</thead>
<tbody>
<tr>
<td>/obasan/ 'aunt' vs. /obaasan/ 'grandmother'</td>
<td></td>
</tr>
<tr>
<td>/biru/ 'building' vs. /biiru/ 'beer'</td>
<td></td>
</tr>
</tbody>
</table>
Research topic

Length contrast and singleton & geminate

- In Japanese, length or duration – important role lexically and pragmatically
- Modern Japanese has a variety of length contrasts.

Vowel length

<table>
<thead>
<tr>
<th>short</th>
<th>long</th>
</tr>
</thead>
<tbody>
<tr>
<td>/obasan/</td>
<td>/obaasan/</td>
</tr>
<tr>
<td>’aunt’</td>
<td>’grandmother’</td>
</tr>
<tr>
<td>/bīru/</td>
<td>/biiru/</td>
</tr>
<tr>
<td>’building’</td>
<td>’beer’</td>
</tr>
</tbody>
</table>
Modern Japanese has a variety of length contrasts.

**Consonant length**

<table>
<thead>
<tr>
<th>Short</th>
<th>Long</th>
</tr>
</thead>
<tbody>
<tr>
<td>/kata/ 'frame'</td>
<td>/katta/ 'bought'</td>
</tr>
<tr>
<td>/hato/ 'dove'</td>
<td>/hatto/ 'hat'</td>
</tr>
</tbody>
</table>

- Short consonant (e.g. /p, t, k, b, d, g/) – **singleton**
- Long consonant (e.g. /pp, tt, kk, bb, dd, gg/) – **geminate or sokuon**

- Geminates – twice or three times as long as singletons (differs according to place and voicing)

- Pragmatic effect – emphatic lengthening
  
  /sugoi/ 'great' ⇒ vowel /sugooi/ consonant /suggoi/

(cf. Podesva 2004, hyperarticulation of /t/ release in English)
Modern Japanese has a variety of length contrasts.

**Consonant length**

*short*  
/kata/ 'frame'  vs.  /katta/ 'bought'

/hto/ 'dove'  vs.  /htto/ 'hat'

- Short consonant (e.g. /p, t, k, b, d, g/) – *singleton*
- Long consonant (e.g. /pp, tt, kk, bb, dd, gg/) – *geminate or sokuon*

- Geminates – twice or three times as long as singletons (differs according to place and voicing)

- Pragmatic effect – emphatic lengthening
  
  /sugoi/ 'great'  \[\Rightarrow\]  *vowel* /sugoi/  *consonant* /sugoi/

  (cf. Podesva 2004, hyperarticulation of /t/ release in English)
**Research topic**

**Previous studies on singleton/geminate**

- Differences between singletons and geminates
- Identification of cues/factors affecting the choices

- **Phonetic studies**
  - duration, (duration of) preceding/following C/V, intensity, F0, F1

- **Phonological studies**
  - lexical strata (native, Sino-Japanese, mimetics, loanwords), geminates in inflection & compound formation, geminate devoicing & OCP
  (Kawahara & Sano 2013, 2017; Sano 2017)

- Difference in constriction duration – primary acoustic correlate of the singleton-geminate contrast. ⇒ useful test of MOP’s assumption
Previous studies on singleton/geminate

- Differences between singletons and geminates
- Identification of cues/factors affecting the choices

- Phonetic studies
duration, (duration of) preceding/following C/V, intensity, F0, F1
- Phonological studies
lexical strata (native, Sino-Japanese, mimetics, loanwords), geminates in inflection & compound formation, geminate devoicing & OCP (Kawahara & Sano 2013, 2017; Sano 2017)

- Difference in constriction duration – primary acoustic correlate of the singleton-geminate contrast. ⇒ useful test of MOP’s assumption
## Roadmap

1. **Background**
   - Research topic
   - **Message Oriented Phonology**
   - Research objective

2. **Method**
   - Corpus
   - Data retrieval
   - Dataset
   - Statistical analysis

3. **Results**
   - Items to be examined
   - Duration of singletons/geminates
   - Patterns in phonetic implementation and informativity

4. **Discussion**
   - Contrastive hyperarticulation
   - Synchronic pattern – diachronic change
Message Oriented Phonology

**Mathematical theories** (e.g. Information Theory, Bayesian Inference)
- offer a means of mathematically quantify the notion of "efficiency" under the assumption that language is an effective system of message transfer. (e.g. Shannon 1948; Bayes 1763; Laplace 1812)

**MOP**
- applies basic concepts of these theories to phonological research
- Information transfer/message transmission is captured by:
  \[ P(\text{message} | \text{signal}, \text{context}) = P(\text{message} | \text{context}) \times P(\text{signal} | \text{message}) \]
  posterior probability \hspace{1cm} predictability \hspace{1cm} signal specificity

- Posterior probability of a message given a phonological form (signal) in context is a *multiplicative function* of
  1. Predictability of the message in the context
  2. Signal specificity (the degree to which signal differentiates the intended message from competitors)
- Predictability \(\rightarrow\) Signal specificity: high \(\Rightarrow\) low, and vice versa
**Message Oriented Phonology**

**Mathematical theories** (e.g. Information Theory, Bayesian Inference)
- offer a means of mathematically quantify the notion of "efficiency" under the assumption that language is an effective system of message transfer. (e.g. Shannon 1948; Bayes 1763; Laplace 1812)

**MOP**
- applies basic concepts of these theories to phonological research
- Information transfer/message transmission is captured by:

\[ P(\text{message} \mid \text{signal}, \text{context}) = P(\text{message} \mid \text{context}) \times P(\text{signal} \mid \text{message}) \]

  - posterior probability
  - predictability
  - signal specificity

- Posterior probability of a message given a phonological form (signal) in context is a *multiplicative function* of
  1. Predictability of the message in the context
  2. Signal specificity (the degree to which signal differentiates the intended message from competitors)
- Predictability – high \(\Rightarrow\) Signal specificity – low, and vice versa
Message Oriented Phonology

Mathematical theories (e.g. Information Theory, Bayesian Inference)
- offer a means of mathematically quantify the notion of "efficiency" under the assumption that language is an effective system of message transfer. (e.g. Shannon 1948; Bayes 1763; Laplace 1812)

MOP
- applies basic concepts of these theories to phonological research
- Information transfer/message transmission is captured by:

$$P(\text{message} \mid \text{signal, context}) = P(\text{message} \mid \text{context}) \times P(\text{signal} \mid \text{message})$$

- Posterior probability of a message given a phonological form (signal) in context is a multiplicative function of
  1. Predictability of the message in the context
  2. Signal specificity (the degree to which signal differentiates the intended message from competitors)

- Predictability – high ⇒ Signal specificity – low, and vice versa
Application of informativity in linguistics
(e.g. Aylett & Turk 2004; Hume & Bromberg 2005; Bell et al. 2009; Jäger 2010; Cohen-Priva 2012; Wedel et al. 2013a, b; Shaw et al. 2014; Kawahara 2016; Nelson & Wedel 2017)

- Contrastive hyperarticulation of VOT in English (Nelson & Wedel 2017)
- Functional load in diachronic change (Wedel et al. 2013a, b)
- Predictability of vowels in English (Aylett & Turk 2004)
- Predictability of content words in English (Bell et al. 2009)
- Informativity & truncation in Chinese compounds (Shaw et al. 2014)
- Quality of epenthetic V in English & French (Hume & Bromberg 2005)
- Informativity & syntactic patterns (Jaeger 2010)
- Informativity & geminate devoicing in Japanese (Kawahara 2016)

Supporting evidence that informativity plays an important role in linguistic behavior.
Message Oriented Phonology

Application of informativity in linguistics
(e.g. Aylett & Turk 2004; Hume & Bromberg 2005; Bell et al. 2009; Jäger 2010; Cohen-Priva 2012; Wedel et al. 2013a, b; Shaw et al. 2014; Kawahara 2016; Nelson & Wedel 2017)

- Contrastive hyperarticulation of VOT in English (Nelson & Wedel 2017)
- Functional load in diachronic change (Wedel et al. 2013a, b)
- Predictability of vowels in English (Aylett & Turk 2004)
- Predictability of content words in English (Bell et al. 2009)
- Informativity & truncation in Chinese compounds (Shaw et al. 2014)
- Quality of epenthetic V in English & French (Hume & Bromberg 2005)
- Informativity & syntactic patterns (Jaeger 2010)
- Informativity & geminate devoicing in Japanese (Kawahara 2016)

Supporting evidence that informativity plays an important role in linguistic behavior.
Roadmap

1. Background
   - Research topic
   - Message Oriented Phonology
   - Research objective

2. Method
   - Corpus
   - Data retrieval
   - Dataset
   - Statistical analysis

3. Results
   - Items to be examined
   - Duration of singletons/geminates
   - Patterns in phonetic implementation and informativity

4. Discussion
   - Contrastive hyperarticulation
   - Synchronic pattern – diachronic change
This study

- applies equation (p.9), and examines Japanese gemination patterns.
  - Japanese gemination exhibits some information-related biased patterns.

- **message** – geminacy
- **signal** – phonological form (singleton/geminate)
- **context** – whether segment is minimally contrastive or not
- **predictability** – probability of singleton/geminate (prior expectation)
  - (negative value of Shannon’s entropy)
  - (how likely singletons/geminates are to be realized by the current signal)
- **signal specificity** – duration or SG ratio
  - (how clearly the speech signal conveys geminacy in message transmission)
  1. SG ratio: mean duration of geminates/mean duration of singletons
  2. Shannon’s entropy \( H(x) \): \[ H(x) = -\sum P(x) \log_2(P(x)) \]

\( \Rightarrow \) operationalized in the analysis
Research objective

Hypothesis

- the durations of minimally contrasting singletons & geminates are hyperarticulated (singletons – shorter, geminates – longer) to provide more information to distinguish their host words from their minimal pair counterparts (competitors).
**Roadmap**

1. **Background**
   - Research topic
   - Message Oriented Phonology
   - Research objective

2. **Method**
   - Corpus
   - Data retrieval
   - Dataset
   - Statistical analysis

3. **Results**
   - Items to be examined
   - Duration of singletons/geminates
   - Patterns in phonetic implementation and informativity

4. **Discussion**
   - Contrastive hyperarticulation
   - Synchronic pattern – diachronic change
CSJ-RDB (NINJAL 2012)
(Corpus of Spontaneous Japanese–Relational Database)

- A part of CSJ ("Core" with rich annotation)
- Size: 201 speech samples (45 hours of speech)
- Various kinds of annotations are linked together.
  - phonetic/phonological information
    e.g. segment, mora, word, phrase, accent, intonation, time (onset/end) → duration
  - morphological information
    e.g. grammatical category, inflectional form
  
  Detailed data retrieval (specify the target)

- Organization: APS (formal) / SPS (casual)
  → study the difference in register
Corpus

**CSJ-RDB** (NINJAL 2012)
(Corpus of Spontaneous Japanese–Relational Database)

- A part of CSJ ("Core" with rich annotation)
- Size: 201 speech samples (45 hours of speech)
- Various kinds of annotations are linked together.
  - **Phonetic/phonological information**
    - e.g. segment, mora, word, phrase, accent, intonation, time (onset/end) → duration
  - **Morphological information**
    - e.g. grammatical category, inflectional form
  
  ⬇

  Detailed data retrieval (specify the target)

- Organization: APS (formal) / SPS (casual)
  → study the difference in register
Corpus

CSJ-RDB (NINJAL 2012)
(Corpus of Spontaneous Japanese–Relational Database)

- A part of CSJ ("Core" with rich annotation)
- Size: 201 speech samples (45 hours of speech)
- Various kinds of annotations are linked together.
  - phonetic/phonological information
    - e.g. segment, mora, word, phrase, accent, intonation, time (onset/end) → duration
  - morphological information
    - e.g. grammatical category, inflectional form
- Detailed data retrieval (specify the target)
- Organization: APS (formal) / SPS (casual)
  → study the difference in register
# Roadmap

1. **Background**
   - Research topic
   - Message Oriented Phonology
   - Research objective

2. **Method**
   - Corpus
   - Data retrieval
   - Dataset
   - Statistical analysis

3. **Results**
   - Items to be examined
   - Duration of singletons/geminates
   - Patterns in phonetic implementation and informativity

4. **Discussion**
   - Contrastive hyperarticulation
   - Synchronic pattern – diachronic change
I retrieved the target data from the CSJ-RDB.

12 speech samples in the CSJ-RDB

Target segments: singletons & geminates
   (& word, phrase that contained the target)

Using MySQL (implementing the programming language SQL)
   http://www.navicat.com

Search formula in SQL

Employing phonetic/phonological and morphological information
Data retrieval

Filtering

- Tokens were excluded from the dataset if targeted segments were a part of filled pauses or word fragments (a morpheme marked with (D )).
- Tokens were included if targeted segments possessed a non-standard pronunciation (a morpheme marked with (W )).

Annotation

- Duration was calculated based on the annotation in the CSJ-RDB (end time – onset time)
- Labels regarding minimal pair categories.
Data retrieval

- Manually annotated in an item-by-item manner, with reference to the Japanese dictionary.
- If a member of a pair is a proper noun, jargon, an archaic form, a dialectal form, or one that differs from its counterpart in accent and/or grammatical category, the pair is not regarded as a minimal pair.
- Three categories
  - lexically contrastive minimal pairs
  - pragmatically contrastive minimal pairs
  - absence of minimal pairs

*pragmatically contrastive: gemination due to emphasis (e.g. /sugoi/ vs. /suggoi/ 'great') or allophonic pairs due to style/register (e.g. /mina/ vs. /minna/ 'everyone,' /fakusu/ vs. /fakkusu/ 'fax')
Data retrieval

- Manually annotated in an item-by-item manner, with reference to the Japanese dictionary.
- If a member of a pair is a proper noun, jargon, an archaic form, a dialectal form, or one that differs from its counterpart in accent and/or grammatical category, the pair is not regarded as a minimal pair.
- **Three categories**
  1. lexically contrastive minimal pairs
  2. pragmatically contrastive minimal pairs
  3. absence of minimal pairs

  *pragmatically contrastive: gemination due to emphasis (e.g. /sugoi/ vs. /suggoi/ 'great') or allophonic pairs due to style/register (e.g. /mina/ vs. /minna/ 'everyone,' /fakusu/ vs. /fakkusu/ 'fax')*
Roadmap

1. **Background**
   - Research topic
   - Message Oriented Phonology
   - Research objective

2. **Method**
   - Corpus
   - Data retrieval
   - **Dataset**
   - Statistical analysis

3. **Results**
   - Items to be examined
   - Duration of singletons/geminates
   - Patterns in phonetic implementation and informativity

4. **Discussion**
   - Contrastive hyperarticulation
   - Synchronic pattern – diachronic change
Dataset

- Exhaustive search & filtering $\Rightarrow$ 12,583 tokens

Table 1. The frequency distribution of singleton/geminate in the CSJ-RDB

<table>
<thead>
<tr>
<th></th>
<th>frequency</th>
<th>ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>singleton</td>
<td>10,717</td>
<td>0.852</td>
</tr>
<tr>
<td>geminate</td>
<td>1,866</td>
<td>0.148</td>
</tr>
</tbody>
</table>

Table 2. The duration of singleton/geminate in the CSJ-RDB

<table>
<thead>
<tr>
<th></th>
<th>mean duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>singleton</td>
<td>34.4 msec</td>
</tr>
<tr>
<td>geminate</td>
<td>88.1 msec</td>
</tr>
<tr>
<td>SG ratio</td>
<td>2.56</td>
</tr>
</tbody>
</table>

(SG ratio: mean duration of geminates/mean duration of singletons)
Table 3. The frequency distribution of singleton/geminate by the presence/absence of minimal pairs

<table>
<thead>
<tr>
<th></th>
<th>lexical</th>
<th>pragmatic</th>
<th>absence</th>
</tr>
</thead>
<tbody>
<tr>
<td>singleton</td>
<td>162</td>
<td>148</td>
<td>10,407</td>
</tr>
<tr>
<td>geminate</td>
<td>161</td>
<td>31</td>
<td>1,674</td>
</tr>
</tbody>
</table>

Each token – subjected to the analysis
Roadmap

1 Background
   • Research topic
   • Message Oriented Phonology
   • Research objective

2 Method
   • Corpus
   • Data retrieval
   • Dataset
   • Statistical analysis

3 Results
   • Items to be examined
   • Duration of singletons/geminates
   • Patterns in phonetic implementation and informativity

4 Discussion
   • Contrastive hyperarticulation
   • Synchronic pattern – diachronic change
Statistical analysis

*Distributional skews were tested by*

- Linear mixed-effects (hierarchical generalized linear) model
  (Barr 2013; Bar et al. 2013)
- *lmer* in *R* (R development Core Team 1993-2017)
- **Variables:**
  - Dependent variable: duration of singleton/geminate
  - Fixed effect: presence/absence of minimal pairs (lexically or pragmatically contrastive)
  - Random effects (grouping variables): speakers and items
    *Random slope and random intercepts were included in the model to have maximal random effects structure.*
  - Post-hoc test: multiple comparisons with Tukey’s method
Distributional skews were tested by

- Linear mixed-effects (hierarchical generalized linear) model (Barr 2013; Bar et al. 2013)
- `lmer` in R (R development Core Team 1993-2017)

**Variables:**

- Dependent variable: duration of singleton/geminate
- Fixed effect: presence/absence of minimal pairs (lexically or pragmatically contrastive)
- Random effects (grouping variables): speakers and items
  *Random slope and random intercepts were included in the model to have maximal random effects structure.*

- Post-hoc test: multiple comparisons with Tukey’s method
Roadmap

1. **Background**
   - Research topic
   - Message Oriented Phonology
   - Research objective

2. **Method**
   - Corpus
   - Data retrieval
   - Dataset
   - Statistical analysis

3. **Results**
   - Items to be examined
   - Duration of singletons/geminates
   - Patterns in phonetic implementation and informativity

4. **Discussion**
   - Contrastive hyperarticulation
   - Synchronic pattern – diachronic change
Items to be examined

Hypothesis

- Contrastive hyperarticulation: phonetic implementation is exaggerated to enhance the contrast between singletons & geminates

1. If singletons contrast in a minimal pair ⇒ is their duration shorter than that do not?
2. If geminates contrast in a minimal pair ⇒ is their duration longer than that do not?
3. How about pragmatic contrast (similar to lexical contrast or absence)?
Roadmap

1. Background
   - Research topic
   - Message Oriented Phonology
   - Research objective

2. Method
   - Corpus
   - Data retrieval
   - Dataset
   - Statistical analysis

3. Results
   - Items to be examined
   - Duration of singletons/geminates
   - Patterns in phonetic implementation and informativity

4. Discussion
   - Contrastive hyperarticulation
   - Synchronic pattern – diachronic change
Duration of singletons

Figure 1. The duration of singletons in three categories

$\text{t}=2.315, \ p<0.05$

- **Duration**: lexical, pragmatic $<$ absence ($p<0.01$)
**Duration of geminates**

Figure 2. The duration of geminates in three categories

\[ t = -2.364, \ p < 0.05 \]

- **Duration**: lexical, pragmatic < absence \((p<0.05)\)
Summary of the results

**Observed pattern**

- **singleton**: lexical, pragmatic < absence
- **geminate**: lexical, pragmatic > absence

1. √ If singletons contrast in a minimal pair ⇒ their duration is shorter than that do not.
2. √ If geminates contrast in a minimal pair ⇒ their duration is longer than that do not.
3. Pragmatic contrast – similar to lexical contrast (minimal pair ⇒ singletons – shorter, geminates longer)
Summary of the results

**Observed pattern**

- singleton: lexical, pragmatic < absence
- geminate: lexical, pragmatic > absence

1. √ If singletons contrast in a minimal pair ⇒ their duration is shorter than that do not.
2. √ If geminates contrast in a minimal pair ⇒ their duration is longer than that do not.
3. Pragmatic contrast – similar to lexical contrast (minimal pair ⇒ singletons – shorter, geminates longer)
# Roadmap

1. **Background**
   - Research topic
   - Message Oriented Phonology
   - Research objective

2. **Method**
   - Corpus
   - Data retrieval
   - Dataset
   - Statistical analysis

3. **Results**
   - Items to be examined
   - Duration of singletons/geminates
   - Patterns in phonetic implementation and informativity

4. **Discussion**
   - Contrastive hyperarticulation
   - Synchronic pattern – diachronic change
Patterns in phonetic implementation and informativity

- How the results (phonetic implementation) relates to informativity?
- SG ratio: degree of contrast, distance between singleton & geminate
- Shannon’s entropy $H(x)$ represents informativity
- Entropy values were calculated for three categories.
  
  e.g. Lexical contrast

The conditional probabilities of singleton and geminate are:

$$P(\text{singleton}) = \frac{162}{162+161} = 0.502$$

$$P(\text{singleton}) = \frac{161}{162+161} = 0.498$$

The information content of each segment is:

$$\text{Inf}(\text{singleton}) = -\log_2(0.502) = 0.996$$

$$\text{Inf}(\text{geminate}) = -\log_2(0.498) = 1.004$$

The entropy of a singleton-geminate contrast (token frequency: 162 and 161) in lexical contrast is:

$$H(\text{singleton \, geminate}) = -\sum P(x) \log_2(P(x)) = 0.502 \times 0.498 + 0.996 \times 1.004 = 1.00$$
Patterns in phonetic implementation and informativity

- How the results (phonetic implementation) relates to informativity?
- SG ratio: degree of contrast, distance between singleton & geminate
- **Shannon’s entropy** $H(x)$ represents informativity
- Entropy values were calculated for three categories.
  e.g. Lexical contrast

The conditional probabilities of singleton and geminate are:

- $P(\text{singleton}) = 162/ (162+161) = 0.502$
- $P(\text{singleton}) = 161/ (162+161) = 0.498$

The information content of each segment is:

- $\text{Inf}(\text{singleton}) = -\log_2(0.502) = 0.996$
- $\text{Inf}(\text{geminate}) = -\log_2(0.498) = 1.004$

The entropy of a singleton-geminate contrast (token frequency: 162 and 161) in lexical contrast is:

$$H(\text{singleton geminate}) = -\sum P(x) \times \log_2(P(x)) = 0.502 \times 0.498 + 0.996 \times 1.004 = 1.00$$
Patterns in phonetic implementation and informativity

Table 4. SG ratio and Shannon's entropy in three categories

<table>
<thead>
<tr>
<th></th>
<th>lexical</th>
<th>pragmatic</th>
<th>absence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG ratio</td>
<td>4.45</td>
<td>3.62</td>
<td>2.48</td>
</tr>
<tr>
<td>Shannon's entropy</td>
<td>1</td>
<td>0.66</td>
<td>0.58</td>
</tr>
</tbody>
</table>

- **SG ratio**: lexical > pragmatic > absence
- **Shannon’s entropy**: lexical > pragmatic > absence

SG ratio follows informativity represented by Shannon’s entropy
Patterns in phonetic implementation and informativity

Table 4. SG ratio and Shannon’s entropy in three categories

<table>
<thead>
<tr>
<th></th>
<th>lexical</th>
<th>pragmatic</th>
<th>absence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG ratio</td>
<td>4.45</td>
<td>3.62</td>
<td>2.48</td>
</tr>
<tr>
<td>Shannon’s entropy</td>
<td>1</td>
<td>0.66</td>
<td>0.58</td>
</tr>
</tbody>
</table>

- **SG ratio**: lexical > pragmatic > absence
- **Shannon’s entropy**: lexical > pragmatic > absence

SG ratio follows informativity represented by Shannon’s entropy
Roadmap

1. Background
   - Research topic
   - Message Oriented Phonology
   - Research objective

2. Method
   - Corpus
   - Data retrieval
   - Dataset
   - Statistical analysis

3. Results
   - Items to be examined
   - Duration of singletons/geminates
   - Patterns in phonetic implementation and informativity

4. Discussion
   - Contrastive hyperarticulation
   - Synchronic pattern – diachronic change
Contrastive hyperarticulation

- Hypothosis – borne out:
  Durations of singletons & geminates in a minimal pair show the effects of hyperarticulation to provide more information to distinguish their host word from its minimal pair competitor (cf. Nelson & Wedel 2017).

- Presence of a minimal pair competitor
  \[ \Rightarrow \] predictability with which a target segment is identified lower
  \[ \Rightarrow \] requires the signal specificity to be more informative/salient to differentiate the target from other competitors
  \[ \Rightarrow \] hyperarticulation of phonetic cues that provide more information to distinguish their host word from its minimal pair competitor.

\[ P(message \mid signal, context) = P(message \mid context) \times P(signal \mid message) \]

- Posterior probability
- Predictability
- Signal specificity

\[ \uparrow \] low

\[ \uparrow \] high

\[ \uparrow \] constant
Contrastive hyperarticulation

- Hypothesis – borne out:
  Durations of singletons & geminates in a minimal pair show the effects of hyperarticulation to provide more information to distinguish their host word from its minimal pair competitor (cf. Nelson & Wedel 2017).

- Presence of a minimal pair competitor
  ⇒ predictability with which a target segment is identified lower
  ⇒ requires the signal specificity to be more informative/salient to differentiate the target from other competitors
  ⇒ hyperarticulation of phonetic cues that provide more information to distinguish their host word from its minimal pair competitor.

\[ P(message \mid signal, context) = P(message \mid context) * P(signal \mid message) \]

- posterior probability
- predictability
- signal specificity

\[ \text{constant} \quad \text{low} \quad \text{high} \]
Contrastive hyperarticulation

- Hypothesis – borne out:
  Durations of singletons & geminates in a minimal pair show the effects of hyperarticulation to provide more information to distinguish their host word from its minimal pair competitor (cf. Nelson & Wedel 2017).

- Presence of a minimal pair competitor
  ⇒ predictability with which a target segment is identified lower
  ⇒ requires the signal specificity to be more informative/salient to differentiate the target from other competitors
  ⇒ hyperarticulation of phonetic cues that provide more information to distinguish their host word from its minimal pair competitor.

\[
P(\text{message} \mid \text{signal, context}) = P(\text{message} \mid \text{context}) \times P(\text{signal} \mid \text{message})
\]

posterior probability predictability signal specificity

\[\uparrow\] low
Contrastive hyperarticulation

- Hypothesis – borne out:
  Durations of singletons & geminates in a minimal pair show the effects of hyperarticulation to provide more information to distinguish their host word from its minimal pair competitor (cf. Nelson & Wedel 2017).

- Presence of a minimal pair competitor
  ⇒ predictability with which a target segment is identified lower
  ⇒ requires the signal specificity to be more informative/salient to differentiate the target from other competitors
  ⇒ hyperarticulation of phonetic cues that provide more information to distinguish their host word from its minimal pair competitor.

\[ P(message \mid signal, context) = P(message \mid context) \times P(signal \mid message) \]

- Posterior probability
- Predictability
- Signal specificity
- Low
- Constant
- High
Contrastive hyperarticulation

- **Degree of durational contrast** between singletons & geminates (or hyperarticulation) reflected in the SG ratio follows from the informativity of singleton-geminate contrasts represented by Shannons entropy.

- Hyperarticulation of durational contrast is observed both in **lexical** minimal pairs and **pragmatic** minimal pairs.

  $\Rightarrow$ role of non-lexical/non-phonemic information in linguistic behavior.
## Roadmap

1. **Background**
   - Research topic
   - Message Oriented Phonology
   - Research objective

2. **Method**
   - Corpus
   - Data retrieval
   - Dataset
   - Statistical analysis

3. **Results**
   - Items to be examined
   - Duration of singletons/geminates
   - Patterns in phonetic implementation and informativity

4. **Discussion**
   - Contrastive hyperarticulation
   - Synchronic pattern – diachronic change
Synchronic pattern – diachronic change

**Functional load hypothesis**

- Phonological contrasts that carry high functional loads (more minimal pairs) are less likely to neutralize (Martinet 1952; Hockett 1967; Surendran & Niyogi 2006; Wedel et al. 2013a,b).\(^1\)

- Wedel et al. (2013a, b): demonstrate using a large database (9 languages/dialects) that the number of lexical minimal pairs distinguished by a phoneme opposition was a strong predictor of merger probability. (high functional load $\Rightarrow$ contrast – maintained)

- Prediction: frequent enhancement of a phonetic cue to a lexical category will become reflected in its long-term phonetic representation.

---

\(^1\)Prior to 2013, the functional load hypothesis has been tested by many different ways, which produced mixed results (see Hockett 1967, and Surendran and Niyogi 2006).
This project – supporting evidence

- The results predict that synchronically informative contrasts are hyperarticulated, and thus their phonetic implementation is perceptually salient.
  ⇒ This results in diachronic stability, whereby informative contrasts tend to be preserved.

- Thus, the hyperarticulation of individual sounds induced by lexical (pragmatic) competition can influence long-term change in the system of phonemic contrasts (cf. Baese & Goldrick 2009; Peramunage et al. 2011).
Thank you!

This project was supported by JSPS KAKENHI Grant #16K16831, #26284059, and #16H03426.
## References

References


References


References


References


