Papers on Japanese Imperfect Puns

Shigeto Kawahara

Jan, 2010

Preface

Shigeto Kawahara
Rutgers University
kawahara@rci.rutgers.edu

1 The rationale

This volume "Papers on Japanese puns" is a collection of papers on Japanese imperfect puns that I wrote between 2007 and 2009 (often in collaboration with Professor Kazuko Shinohara, the Tokyo University of Agriculture and Technology), although most of the papers appear in print in 2009 and 2010. There are several reasons to put together a volume like this. First, I believe that I have done a wide range of analyses on Japanese imperfect puns, and it is convenient to have a collection of relevant papers. Second, although as much as I liked—and still do like—the project, I feel that I have done enough and am ready to move on to other projects. This collection would thus serve as a closure on my side (though I plan to build my future research partially on this project). Third, related to the first two reasons, I hope that building on these works, somebody interested in this project can take up some remaining issues, and that this volume serves as a useful reference.[1]

2 A brief history

Here is a somewhat personal history of how the research was developed. I was first inspired by a colloquium talk by Professor Donca Steriade at the University of Massachusetts in 2003, and started working on Japanese rap rhyming, the result of which is included as the final paper of this volume. Professor Kazuko Shinohara and I applied the same methodology to Japanese imperfect puns, and wrote the third paper in this volume. Then, we decided that imperfect puns has lots of potential, providing a rich resource for linguistic analyses.

In developing this research program, we also realized its pedagogical values because of its accessibility. For example, the sixth paper included in this volume was based on a BA thesis study by Nobuhiro Yoshida at the Tokyo University of Agriculture and Technology under the supervision of Professor Shinohara. Two of my undergraduate students at Rutgers, Ayanna Beatie and Allan Schwade, worked on Korean tongue twisters from a similar perspective and they presented their work at the undergraduate linguistic conferences at Harvard. Moreover, the materials attract the interest of students in introductory classes, and also the interest of audience outside the field—I have been asked to give my talk on Japanese rap rhymes and puns to various non-professional audiences.

[1] Interested readers are also referred to my website for further information as well, which I will try to update frequently.

3 The structure of this volume

Although readers are welcome to start with any paper in the volume, I ordered them in such a way that readers will see how the research is developed and why I did what I did.

The first two papers are “overview papers”.

The first paper “Faithfulness, correspondence, and perceptual similarity: Hypotheses and experiments” explains why people—including myself—are working on verbal art patterns. In the paper I attempt to tie this general research program with the notion of faithfulness constraints in Optimality Theory. Theoretically-informed readers may start from here.

The second paper “Probing knowledge of similarity through puns” is based on a lecture that I gave at Sophia University in summer of 2008, and provides an overview of this research program. It should be noted that some contents in sections 3 and 4 are a bit outdated, and superseded by the subsequent papers in this volume.

The next four papers present more concrete analyses.

The third paper “The role of psychoacoustic similarity in Japanese puns” reports an analysis of consonant mismatches in Japanese puns.

The fourth paper “Calculating vocalic similarity through puns” does the same with vowels.

The fifth paper “Syllable intrusion in Japanese puns, dajare” analyzes cases in which one phrase internally contains an extra syllable which does not appear in the other phrase.

The sixth paper “Phonetic and psycholinguistic prominences in pun formation: Experimental evidence for positional faithfulness” investigates positional effects. We address whether Japanese speakers care about positions of mismatches in puns. We show that they do and discuss why that is interesting.

At the end, I include my (relatively old) paper on half rhymes in Japanese rap songs, because this paper was a precedent of this general research program.

4 A disclaimer

Putting together this volume does not constitute publication; it is equivalent to circulating offprints. In other words, copyrights belong to whoever has published/will publish each paper. The reference information is given below.

5 References

1. Kawahara, Shigeto (2009) Faithfulness, correspondence, perceptual similarity: Hypotheses and experiments. Journal of the Phonetic Society of Japan. 13.2: 52-61. (A special volume

"Development and elaboration of OT in various domains").

2. Kawahara, Shigeto (2009) Probing knowledge of similarity through puns. In T. Shinya (ed.) Proceedings of Sophia University Linguistic Society 23. Tokyo: Sophia University Linguistics Society. pp.111-138.

3. Kawahara, Shigeto and Kazuko Shinohara (2009) The role of psychoacoustic similarity in Japanese puns: A corpus study. Journal of Linguistics 45.1: 111-138.

4. Kawahara, Shigeto and Kazuko Shinohara (2010) Calculating vocalic similarity through puns. Journal of Phonetic Society of Japan 14.

5. Shinohara, Kazuko and Shigeto Kawahara (2010) Syllable intrusion in Japanese puns, dajare. Proceedings of the 10th meeting of Japan Cognitive Linguistic Association.

6. Kawahara, Shigeto and Kazuko Shinohara (to appear) Phonetic and psycholinguistic prominences in pun formation. In M. den Dikken and W. McClure (eds.) Japanese/Korean Linguistics 18. CSLI.

7. Kawahara, Shigeto (2007) Half rhymes in Japanese rap lyrics and knowledge of similarity. Journal of East Asian Linguistics 16.2: 113-144.

6 Acknowledgements

This research is supported by a Research Council Grant from Rutgers University. Although I received assistance from so many people in writing these papers, I won't repeat them here—you will find their names in each of the papers that follow this preface. However, I must thanks Donca Steriade again here for her initial inspiration as well as her constant encouragement, and Kazuko Shinohara for her continuing patience in collaborating with me.

Shigeto Kawahara
January 2010

1. Kawahara, Shigeto (2009) Faithfulness, correspondence, perceptual similarity: Hypotheses and experiments. Journal of the Phonetic Society of Japan. 13.2: 52-61. (A special volume "Development and elaboration of OT in various domains").

Faithfulness, Correspondence, and Perceptual Similarity: Hypotheses and Experiments*

Shigeto Kawahara, Rutgers University

Keywords: faithfulness, correspondence, perceptibility, similarity, laboratory phonology

1 Introduction

The principle of faithfulness—one of the many characteristics specific to Optimality Theory (OT Prince and Smolensky, 1993/2004)—has opened up new lines of research. The questions addressed in these lines of research have been difficult or even impossible to address in previous theories of phonology. This paper discusses how the principle of faithfulness sheds light on the issues surrounding the phonetics-phonology interface. After reviewing new theories and hypotheses that were made possible to address by principle of faithfulness, this paper reports two similarity judgment experiments that test some of the premises of these theories.

OT has two types of constraints: markedness constraints and faithfulness constraints. OT liberates output markedness problems from solutions by which those problems are solved, whereas a rule-based theory packages markedness problems and solutions into one format. This characteristic of OT has allowed grammar to encode phonetic reasons behind phonological alternations in the formulation of markedness constraints (Hayes and Steriade, 2004; Myers, 1997).[1] Since the issue of encoding phonetic naturalness via markedness constraints has been extensively discussed in the literature already, this paper instead discusses how faithfulness constraints have provided us with a novel way to investigate the relation between phonetics and phonology.

This paper starts with brief discussion of faithfulness constraints in OT in section 2, and moves on to discuss how the principle of faithfulness allows grammar to directly express perceptibility effects in phonology (section 3). Section 4 discusses Correspondence Theory, which has provided a fresh view on the parallel between phonology and verbal art. Section 5 and 6 report similarity judgment experiments that test some premises of the hypotheses discussed in Section 3 and 4. Throughout this paper, I touch on many current issues in OT, but some of these may distract the main flow of the paper—I thus leave the discussion of those current debates in footnotes.[2]

*Some issues that I discuss in this paper came to my attention during my graduate training at the University of Massachusetts, for which I am grateful to John Kingston, John McCarthy, Joe Pater, and Lisa Selkirk. I am also grateful to the participants of my seminar in Spring 2009 at Rutgers University who helped me organize my thoughts on the related issues. Many thanks to Kazu Kurisu, Dan Mash, Maki Shimotani, Kazuko Shinohara, Donca Steriade, Kyoko Yamaguchi, and the reviewer Haruka Fukazawa for comments on earlier versions of this paper. This project is partly funded by a Research Council Grant from Rutgers University. All remaining errors are mine.

[1] OT for this reason has brought about a renewed interest in research on phonetically-driven phonology. However, OT itself is a theory of constraint interaction, which has nothing to do with phonetic naturalness in phonology. Therefore it is mistaken to extend one's argument against phonetically-driven phonology to an argument against OT in general.

[2] Footnotes are great places to find research topics in general (McCarthy, 2008a, p.163).

2 Faithfulness constraints in Optimality Theory

Faithfulness constraints in OT militate against changes from underlying forms (input) to surface forms (output) (see de Lacy, to appear, for a recent overview). As intuitive as this principle sounds, previous rule-based theories of phonology did not explicitly recognize this principle. In fact Halle (1995, p.28) asserts that "... the existence of phonology in every language shows that Faithfulness is at best an ineffective principle that might well be done without." The principle of faithfulness is crucial in OT (see Prince 1997 for a reply to Halle 1995), because otherwise all input forms would be reduced to the most unmarked form of that language, say [ba] (Chomsky, 1995, p.380). One of the characteristics of OT, therefore, is that it explicitly recognizes the role of faithfulness in its theory. This principle was first formulated as Containment Theory (Prince and Smolensky, 1993/2004). In this theory, no segments are literally deleted or added; instead deletion is captured as unparsing to a higher prosodic level and epenthesis as an unfilled prosodic position. However Containment Theory of faithfulness has largely been replaced by the by-now standard theory in the current literature, Correspondence Theory (McCarthy and Prince, 1995).

Correspondence Theory of faithfulness has two characteristics: (i) any two representations can stand in correspondence, and (ii) faithfulness constraints prohibit disparity between the two representations, i.e., maximize the similarity between them. The following discussion illustrates how these two characteristics have allowed us to express certain generalizations in our linguistic behavior that previous theories of phonology failed to capture.

3 Maximizing perceptual similarity

Acknowledging the principle of faithfulness in phonological theory has allowed us to entertain and pursue the hypothesis that speakers maximize the *psychoacoustic* or *perceptual* similarity between inputs and outputs. Perhaps the most influential—and also provocative—formulation of this hypothesis is the P-map hypothesis proposed by Steriade (2001/2008; 2001a). The problem that she tackles is the observation that languages resolve coda voiced obstruents by devoicing them, but not by any other means (say epenthesis or nasalization). Kenstowicz (2003) likewise observes that when speakers whose language lacks voiced stops borrow words with voiced stops, they borrow them as voiceless stops, never as nasal stops. This example is a tip of an iceberg, a problem that has come to be known as a "too-many-repairs/solutions" problem.[3] Steriade argues that languages prefer coda devoicing over other phonological resolutions because devoicing involves a less perceptible change than other phonological changes would. Given the principle of P-map—i.e. speakers maximize the perceptual similarity between input and output—speakers should resort to devoicing, to the extent that devoicing involves the least perceptible change.[4] In its most general form, the more perceptible the change a phonological alternation involves, the higher the rank of the faithfulness constraint it violates.

[3] It would be mistaken to blame OT for predicting too many solutions for particular markedness problems. On the contrary, OT has allowed us to see that there is a problem to be solved. OT in its original formulation does predict that any markedness problem can in principle be resolved by multiple phonological means, while in actuality we observe certain limited ways in which some markedness problems can be solved. However, a rule-based theory of phonology makes the same prediction; this too-many-solutions problem is an issue that any adequate theory of phonology must account for (Blumenfeld 2006; Lombardi 2001; McCarthy 2008a; Steriade 2001/2008). Some proposals regarding the too-many-solutions problem within and out of OT include the fixed-ranking approach based on P-map (Steriade, 2001/2008), OT with Candidate Chains (OT-CC) (McCarthy, 2008b), Targeted constraints (Wilson, 2001), procedural markedness constraints (Blumenfeld, 2006), MAX feature constraints (Lombardi, 2001), and restrictions on diachronic paths leading to phonologization (Myers, 2002).

[4] The original P-map hypothesis predicts that languages would always choose one phonological change for a particular markedness problem, the one chosen being the one that is the least perceptible. However, some markedness problems are solved by various phonological alternations. A typical example is nasal-voiceless stop clusters, which can be resolved by post-nasal voicing, nasalization of stop, denasalization of nasal, etc (Pater, 1999) (see also Zuraw and Lu 2009). One emerging theory to address this problem is to say that constraint rankings projected from P-map are default rankings rather than fixed rankings (Steriade, 2001b; Wilson, 2006). The prediction of this amendment is that novel, emerging phonological patterns follow the ranking predicted by the P-map.

This principle may seem like a simple restatement of the principle of faithfulness, but it is innovative in that it reformulates the notion of similarity as perceptual similarity.[5]

Another example in which perception-based faithfulness constraints have proven useful is the observation that nasal consonants are more likely to assimilate in place than oral consonants (Mohanan, 1993). There are no languages in which only oral consonants assimilate in place, but there are languages in which only nasal consonants assimilate (e.g. Malayalam: Mohanan 1993). In standard phonological feature theories, the asymmetry remains a puzzle because [place] in nasal and [place] in oral consonants are not distinct, and in fact should not be distinct to the extent that homorganic nasal-stop clusters share the same [place] feature, as assumed in standard autosegmental phonology (Goldsmith, 1976). So where does the asymmetry come from?

Jun (2004) argues that the asymmetry comes from the perceptibility of [place] in nasal and oral consonants. He argues that [place] in nasals is less perceptible than [place] in oral consonants and that speakers rank the faithfulness constraint for oral [place] higher than the faithfulness constraint for nasal [place]. The difference in perceptibility of [place] in nasal and oral consonants is supported by some phonetic considerations. Jun (2004), following Malécot (1956), argues that the place contrast in nasals is obscured due to coarticulatory nasalization. Weaker perceptibility of place in nasals finds some evidence in previous psycholinguistic studies. A similarity judgement task shows that speakers judge nasal minimal pairs as more similar to each other than oral consonant minimal pairs (Mohr and Wang, 1968). Pols (1983) shows that Dutch speakers perceive the place contrast more accurately in oral consonants than in nasal consonants (though see the experiments reported below for complications).

Not only has the P-map hypothesis provided insights into cross-linguistic patterning of phonology, it has provided explanations of novel phonological patterns as well. The native phonology of Japanese does not allow voiced obstruent geminates. However, when Japanese speakers borrow words from other languages, mainly from English, they geminate (some) word-final consonants (Shirai, 2002), which resulted in voiced obstruent geminates (e.g. *doggu* 'dog' and *eggu* 'egg'). Having borrowed these words, Japanese speakers have spontaneously started devoicing voiced geminates when they appear with another voiced consonant (Nishimura, 2003), as in (1).

(1) Geminates can devoice if they co-occur with another voiced obstruent

a. ba**dd**o → ba**tt**o 'bad'
b. ba**gg**u → ba**kk**u 'bag'
c. do**gg**u → do**kk**u 'dog'

(2) Singletons do not devoice in the same environment

a. gi**b**u → *gi**p**u 'give'
b. ba**g**u → *ba**k**u 'bug'
c. da**g**u → *da**k**u 'Doug'

The devoicing in (1) takes place due to a dissimilative constraint against two voiced obstruents within the same stem, a constraint known as Lyman's Law in the native phonology of Japanese (Itô and Mester, 1986). However, two voiced singleton obstruents do not devoice in loanwords, as in (2).

Kawahara (2006) argues that this difference between singletons and geminates arises from the ranking IDENT(VOI)-SING $\gg$ OCP(VOI) $\gg$ IDENT(VOI)-GEM, where IDENT(VOI)-SING and IDENT(VOI)-GEM are faithfulness constraints for voicing for singletons and geminates, and OCP(VOI) is a constraint against two voiced obstruents within the same stem.[6] Speakers would project the ranking IDENT(VOI)-SING $\gg$ IDENT(VOI)-GEM if a voicing contrast is less perceptible in geminates than in singletons. Kawahara (2006)

[5] Other proposals encode phonetic perceptibility in markedness constraints by prohibiting non-perceptible contrasts (Flemming, 1995). However, the maximization of perceptual similarity between two corresponding representations can be only formulated in terms of faithfulness constraints, because markedness constraints evaluate the wellformedness of a structure at one-level of representation (Kawahara and Shinohara, to appear).

[6] A markedness based approach is undesirable because the relevant markedness constraint would have to penalize voiced geminates only when they also violate OCP(VOI), a constraint like *[VOICEOBSGEM&OCP(VOI)]$_{stem}$ (Nishimura, 2003) (see Kawahara 2006, sec. 3.3). Pater (to appear) develops a reanalysis of (1) and (2) using Harmonic Phonology with weighted, rather than ranked, constraints, which dispenses with such a complicated markedness constraint. See Tesar (2007) for a reply.

supports the premise about the perceptual asymmetry between voicing in singletons and voicing in geminates in acoustic and perception experiments. Voiced geminates in Japanese are semi-devoiced because of their aerodynamic difficulty (Ohala, 1983), and the semi-devoicing leads to a lower perceptibility of the voicing contrast in geminates. This case shows that the perceptibility of a phonological contrast can shape a novel phonological pattern.[7] In summary, the principle of faithfulness provides a bridge between phonetic perceptibility and phonological grammar.[8]

4 Generalizing Correspondence theory

The principle of maximization of similarity has brought about the formulation of the P-map hypothesis, which has interesting—and controvertial—consequences. Another way in which faithfulness constraints have opened up a new line of research is the study of verbal art including rhyming and puns.

Correspondence Theory in principle allows any two representations to stand in correspondence. In their original proposal, McCarthy and Prince (1995) argue that correspondence holds not only between inputs and outputs, but also between base and reduplicants. Ever since then, the correspondence relation has been argued to hold in many dimensions e.g. between based and derived words (Benua, 1997) (see de Lacy to appear for a recent review). This generality of Correspondence Theory has also resulted in the renewed interests in the study of verbal art (Holtman, 1996; Steriade, 2003).

To discuss an example that relates to the previous discussion on place assimilation, Kawahara (2007) and Kawahara and Shinohara (2009) found that when Japanese speakers pair two consonants in rap rhyming and punning, they are far more likely to pair [m]-[n] than [p]-[t]. Thus there exists an interesting parallel between this observation and the phonological pattern discussed in section 3. Correspondence Theory allows us to capture the parallel in a straightforward manner: both in input-output correspondence and word-word correspondence in rhyming and puns, speakers are more comfortable having the [m]-[n] pair in correspodence than the [p]-[t] pair in correspondence, because the former pair involves more perceptually similar consonants.

We find another interesting parallel between phonology and verbal art. Recall that Steriade (2001/2008) has argued that a voicing contrast is least perceptible among those contrasts made by spectral continuity; that is, speakers neutralize the voicing contrast more than any other contrasts, because minimal pairs differing in voicing are most perceptually similar to each other. This hypothesis finds independent support from rhyme and pun patterns in Japanese; speakers are most willing to pair consonants that differ only in voicing, arguably because they are perceptually similar.

Yet another way in which Correspondence Theory reveals an interesting parallel between verbal art and phonology is positional effects. In phonology speakers avoid making changes in initial syllables (Beckman, 1998), perhaps because initial syllables are psycholinguistically prominent, and such changes would consequently make lexical access difficult (Hawkins and Cutler, 1988). For example in Sino-Japanese, initial syllables allow many consonants but non-initial syllables allow only [t] and [k] (Tateishi, 1990). Assuming

[7] One debate concerning the general issue of phonetic naturalness in phonology is whether such perceptibility effects are encoded in synchronic grammar or result from diachronic changes. The first position, which has been implicitly assumed here, asserts that speakers possess knowledge of perceptibility and have the principle of minimization of perceptual disparity between the corresponding segments. An alternative is to say that listeners simply misperceive contrasts that are not perceptible, which result in a sound change (Blevins, 2004; Myers, 2002). In this theory speakers do not need to have explicit knowledge of perceptibility. However, some studies have argued that when speakers innovate novel phonological patterns, they show phonetically natural patterns, even when historical misperceptions are not at issue (Kawahara, 2006; Wilson, 2006; Zuraw, 2007). To the extent that historical changes can bring about unnatural phonological patterns, it would be crucial to look at novel, emergent phonological patterns which speakers spontaneously create in order to support the thesis of phonetic naturalness in phonology.

[8] There is potentially a chicken-and-egg problem here, because our linguistic knowledge affects our speech perception as well (Massaro and Cohen, 1983; Moreton, 2002): Does speech perception affect phonology first? Or does phonology affect speech perception first? The answer would probably be that the influence is bi-directional. The challenge therefore is how to model this bi-directionality (Boersma, 2006; Hume and Johnson, 2001).

the Richness of the Base (Prince and Smolensky, 1993/2004), speakers need to map an input like /sasu/ to [satu] in Sino-Japanese, as in (3a).

(3) The parallel between phonological mapping and pun pairing

a. Phonological input-output mapping

$$\begin{array}{lcccccc} \text{Input} & / & \mathbf{s}_i & \text{a}_j & \mathbf{s}_k & \text{u}_l & / \\ & & | & | & | & | & \\ \text{Output} & [& \mathbf{s}_i & \text{a}_j & \mathbf{t}_k & \text{u}_l &] \end{array}$$

b. Pun pairing

$$\begin{array}{lcccccc} \text{Word 1} & [& \mathbf{s}_i & \text{a}_j & \mathbf{s}_k & \text{u}_l &] \\ & & | & | & | & | & \\ \text{Word 2} & [& \mathbf{s}_i & \text{a}_j & \mathbf{t}_k & \text{u}_l &] \end{array}$$

Kawahara and Shinohara (to appear) have shown via a wellformedness judgment experiment that the same pattern—the avoidance of disparity in initial segments—is observed in puns. We have found that speakers judge a pun involving an initial mismatch (e.g. **s***asetsu-ni* **z***asetsu* ‘I gave up turning left’) as less wellformed than a pun involving an internal mismatch (e.g. *hi***s***ashi-ni hi***z***ashi* ‘Sunlight on the sun roof’). Correspondence Theory allows us to generalize the two observations, as in (3): speakers avoid having a mismatched correspondence pair in initial positions, more so than having a mismatched pair in internal positions. In other words, Correspondence Theory models two separate patterns—resistance of initial syllables being changed in phonology and the wellformedness judgment pattern in puns—using a single principle.

To summarize, Correspondence Theory formalizes the parallel between phonology and verbal art.[9] Furthermore, this finding has given rise to a new research program: to the extent that the same principle governs both phonology and verbal art, we can investigate our phonological knowledge through verbal art. See Kawahara and Shinohara (2009) for references, and Kawahara (2009) as well as the author’s website for suggestions for future research regarding Japanese puns.

5 Testing some premises: Experiment 1

The maximization principle of similarity incorporates the effect of perceptibility in phonology. The generality of Correspondence Theory captures the parallel between phonology and verbal art. In addition to some open questions that I outlined in the preceding footnotes, one important line for future research is to test hypotheses about perceptual grounding of phonology by experiments. There have been several studies that specifically test such hypotheses (see Kawahara, to appear, for a review), but there are several hypotheses that remain to be tested. For example, Winter (2003) points out that the evidence for lower perceptibility of [place] in nasal is weak, and he himself did not find convincing evidence for a perceptibility difference between nasal [place] and oral [place] in a difference magnitude estimation task or an AX discrimination task. This debate shows that it is important to test the premises for perception-based explanations of phonological patterns. To this end I report (admittedly preliminary) similarity judgment experiments that attempt to test the assumptions about perceptual similarity discussed in the preceding sections.

5.1 Method

The first experiment was a paper-based forced-choice similarity judgment task. The experiment addressed two hypotheses: (i) nasal minimal pairs are more similar to each other than oral consonant minimal pairs (Jun, 2004) (ii) pairs differing in voicing are more similar than pairs differing in other manner features (Kawahara, 2007; Kawahara and Shinohara, 2009; Kenstowicz, 2003; Steriade, 2001/2008). The stimuli

[9] The observation about the parallel between phonology and verbal art is not new, explicitly noticed at least as early as Kiparksy (1973) and Zwicky (1976) (in fact, the origin of literary linguistics dates back to even older time: Fabb 1997). However, Correspondence Theory has provided a tool with which to formulate the parallel explicitly. Another point that is worth mentioning is that some proposals have claimed for a return to Containment Theory (e.g. van Oostendorp, 2008). As far as I can see, it is impossible to capture the parallel between phonology and verbal art in this theory, because Containment Theory does not provide a general mechanism to relate two representations.

consisted of two pairs of consonants (e.g. [ba]-[pa] vs. [ba]-[ma]) to allow participants to judge which pair involved more similar consonants. The task of the participants was thus analogous to comparing (the similarity of) two input-output pairs. The list of the stimuli is given in (4). In addition to these 8 target comparisons, 12 filler dummy comparisons were added. The order of two pairs was counterbalanced by preparing two types of questionnaire.

(4) The stimuli used to address the two hypotheses

a. The weaker perceptibility of nasal [place]: [ma]-[na] vs. [ba]-[da], [ma]-[na] vs. [pa]-[ta]

b. The weaker perceptibility of [voice]: [ba]-[pa] vs. [ba]-[va], [ba]-[pa] vs. [ba]-[ma], [da]-[ta] vs. [da]-[za], [da]-[ta] vs. [da]-[na], [za]-[sa] vs. [za]-[na], [za]-[sa] vs. [za]-[da]

The participants were students taking an introductory psychology class at Rutgers University and two graduate students in the linguistics department. None of them were familiar with the related P-map hypotheses tested in this study. They were encouraged to read the stimuli silently before responding to each question and base their judgment on auditory quality rather than orthographic similarity. The entire process took about 20 minutes. No compensation was given to the participants. Excluding non-native speakers of English, the data from 34 speakers entered into the subsequent statistical analysis. To statistically assess the obtained data, after excluding non-responses, a binomial test was run against the null hypothesis that the participants' responses were random (that is, the probability of one particular pair to be chosen as more similar is .5). The alpha-level was adjusted to 0.05/8=0.006 by a Bonferroni adjustment.

5.2 Result

Table 1 tallies the results. "Expected responses" are those that are expected from the two hypotheses being tested: nasal minimal pairs are more similar to each other than oral consonant minimal pairs, and minimal pairs differing in voicing are more similar to minimal pairs differing in other manner features (nasality and continuany). Statistical significance is signaled by an asterisk.

Table 1: The number of expected, unexpected, and no-responses in Experiment 1.

	Pairs			Expected responses	Unexpected responses	No responses	p
a.	[ma]-[na]	vs.	[pa]-[ta]	25 (74%)	9	0	$= .003$*
b.	[ma]-[na]	vs.	[ba]-[da]	26 (79%)	7	1	$< .001$*
c.	[ba]-[pa]	vs.	[ba]-[ma]	27 (79%)	7	0	$< .001$*
d.	[da]-[ta]	vs.	[da]-[na]	22 (65%)	12	0	$n.s.$
e.	[za]-[sa]	vs.	[za]-[na]	30 (88%)	4	0	$< .001$*
f.	[ba]-[pa]	vs.	[ba]-[va]	24 (71%)	10	0	$n.s.$
g.	[da]-[ta]	vs.	[da]-[za]	25 (74%)	9	0	$= .003$*
h.	[za]-[sa]	vs.	[za]-[da]	29 (85%)	5	0	$< .001$*

5.3 Discussion

As observed in the first two rows in Table 1, the first hypothesis is statistically confirmed—participants judged the nasal minimal pair more similar to each other than oral consonant minimal pairs at a more than chance frequency. The rows (c, e) statistically show that speaker judge the minimal pair differing in voicing more similar than the minimal pair differing in nasality, and the comparison (d) shows the same tendency ($p = .03$). Finally, the rows (f, g, h) show that speakers tend to judge the minimal pair differing in voicing more similar than the minimal pair differing in continuancy, although the comparison in (f) did not reach significance ($p = .007$).

6 Experiment 2

6.1 Introduction

Although Experiment 1 supports the two hypotheses about the phonetic grounding of phonological patterns—weaker perceptibility of [place] in nasal and weaker perceptibility of [voice]—the results could have been affected by orthographic similarity, although the participants were encouraged to use auditory impression. Furthermore, the phonological alternations under question—nasal place assimilation and coda devoicing—occur in coda, and therefore we should test the hypotheses about perceptibility in coda as well. Therefore the second experiment tested these hypotheses both in onset and coda using auditory stimuli.

In addition to the two hypotheses tested in Experiment 1, this experiment tested two more hypotheses, summarized in (5).

(5) The four hypotheses tested in Experiment 2

a. The [place] contrast is weaker in nasals than oral consonants.

b. The [voice] contrast is weaker than [nas] contrast.[10]

c. The redundant [+voice] feature in sonorants promote their similarity with voiced obstruents.

d. Glides and [h] are not highly audible and hence similar to ϕ whereas a strident like [s] is highly audible and not similar to ϕ.

The hypothesis in (5c) was proposed to explain the observation that in languages that avoid similar consonants in adjacent positions, sonorants are considered to be more similar to voiced obstruents than to voiceless obstruents (Frisch et al., 2004; Kawahara and Shinohara, 2009). Hypothesis (5d) explains why languages prefer to use glottal consonants and glides for epenthesis while no languages epenthesize [s]: speakers prefer consonants that are most similar to ϕ for epenthesis, and [s] is too different from ϕ (Kawahara and Shinohara, 2009; Steriade, 2001a). The hypothesis (5d) also explains why [s] is unlikely to be deleted in loanword adaptation (Steriade, 2001a), again because [s] is too different from ϕ. The high audibility of [s] also explains why [s] can violate the sonority sequencing requirement in English onset clusters (Wright, 2004). The format of the experiment is the same as Experiment 1; speakers were presented with two pairs of sounds within each comparison and asked to judge which pair involved more similar consonants.

6.2 Method

The stimuli consist of 12 pairs to test the four hypotheses in (6).

(6) The stimuli in Experiment 2

a. Hypothesis (5a): [ma]-[na] vs. [pa]-[ta], [ma]-[na] vs. [ba]-[da] , [am]-[an] vs. [ap]-[at], [am]-[an] vs. [ab]-[ad]

b. Hypothesis (5b): [ba]-[pa] vs. [ba]-[ma], [da]-[ta] vs. [da]-[na], [ab]-[ap] vs. [ab]-[am], [ad]-[at] vs. [ad]-[an]

c. Hypothesis (5c): [ba]-[ma] vs. [pa]-[ma], [da]-[na] vs. [ta]-[na],

[10] Experiment 2 did not compare [voice] and [cont], because this comparison is not relevant to the P-map hypothesis. Recall that the hypothesis addresses why languages only resort to devoicing to resolve coda voiced obstruents; however, spirantization would not eliminate coda voiced obstruents, because voiced fricatives are still voiced obstruents.

d. Hypothesis (5d): [wa]-[a] vs. [sa]-[a], [ha]-[a] vs. [sa]-[a]

.

In order to create the stimuli, a native speaker of English pronounced all the stimulus syllables in a frame sentence 'Please say the word X again'. Each syllable was written on a separate index card, and the order was randomized. His speech was recorded through an AT4040 Cardioid Capacitor microphone with a pop filter in a sound-attenuated recording booth and amplified through an ART TubeMP microphone pre-amplifier (JVC RX 554V). The speech was digitized with 44k sampling rate upon recording using GoldWave. After the recording, the syllables were extracted from the frame sentence at zero crossing. Since the speaker did not assign a uniform pitch contour to all syllables, the pitch contour was artificially made uniform by imposing a flat contour at 110Hz using PSOLA in Praat (Boersma and Weenink, 1999-2009). Their amplitude was also made uniform at the peak of 0.6. The syllables were then windowed with on- and off-ramps of 0.005 ms. The resynthesized syllables were then combined to form pairs with 100 ms of silence in between. Two pairs were finally combined with 500 ms in between. The ordering of pairs was controlled by preparing two orderings.

The participants were students of introductory linguistics classes at Rutgers University. The stimuli were played through HK 195 multimedia speaker systems from a Macintosh computer in quiet rooms, and they were asked to choose which pair sounded more similar to each other. The inter trial interval was 5 seconds, although the participants were encouraged to use their first auditory impression. In order to avoid the effect of orthography, the answer sheet did not provide the orthographic representations of the stimuli. The overall experiment took about 15 minutes. They were paid one dollar for their time.

The data from non-native speakers of English were excluded from the analysis. Also, data from two participants who chose the first pair as more similar in all but one comparison were excluded. As a result, data from 36 participants entered into the statistical analysis. The statistical significance of the results was assessed via a binomial test. The alpha-level was set at .01 for the following reason; some results were expected from Experiment 1, so that a drastic Bonferronization would increase Type 2 error (Myers and Well, 2003, p.243-244); on the other hand, since there were 12 comparisons, not adjusting the alpha-level may result in the inflation of Type 1 error.

6.3 Results

Table 2 tallies responses that are expected from the hypotheses in (5). Unlike Experiment 1 there were no non-responses. Again asterisks signal statistical significance. The asterisks in parentheses show significant results that are opposite from expectation.

6.4 Discussion

The first hypothesis, the weaker perceptibility of [place] in nasal, is not supported by the results. Only the onset comparison (a) supports it, but the other three pairs (c-d) did not support the hypothesis. Surprisingly, given the comparison (d) in coda, English speakers judged the oral consonant pair as more similar than the nasal consonant pair. The second hypothesis, the weaker perceptibility of [voice] compared to [nasal], is observed only in coda pairs. The onset results (e, f) are not compatible with the results in Experiment 1 or with Japanese pun or rhyme patterns (Kawahara, 2007; Kawahara and Shinohara, 2009). However, the coda results (g, h) are consistent with the idea that speakers prefer coda devoicing to coda nasalization because the former involves smaller perceptual changes (Steriade, 2001/2008). The third hypothesis that voicing in sonorant promotes their similarity with voiced obstruents was not supported—neither pairs showed statistically significant skew. The fourth hypothesis that glides and [h] are closer to ϕ than [s] is supported.

In summary, the experiment supported only a subset of phonetically-based hypotheses about phonological patterns. The results, however, were not conclusive and further experimentation is warranted. First, since one comparison ([A]-[B] vs. [C]-[D]) was presented only once, the listeners may have had difficulty

Table 2: The number of expected and unexpected responses in Experiment 2.

	Pairs			Expected responses	Unexpected responses	p
a.	[ma]-[na]	vs.	[ba]-[da]	29 (81%)	7	$<$ .001*
b.	[ma]-[na]	vs.	[pa]-[ta]	19 (53%)	17	$n.s.$
c.	[am]-[an]	vs.	[ab]-[ad]	16 (44%)	20	$n.s.$
d.	[am]-[an]	vs.	[ap]-[at]	8 (22%)	28	$<$ .001(*)
e.	[ba]-[pa]	vs.	[ba]-[ma]	11 (31%)	25	$<$ .01(*)
f.	[da]-[ta]	vs.	[da]-[na]	15 (42%)	21	$n.s.$
g.	[ab]-[ap]	vs.	[ab]-[am]	29 (81%)	7	$<$ .001*
h.	[ad]-[at]	vs.	[ad]-[an]	26 (72%)	10	$<$ .01*
i.	[ba]-[ma]	vs.	[pa]-[ma]	23 (64%)	13	$n.s.$
j.	[da]-[na]	vs.	[ta]-[na]	15 (42%)	21	$n.s.$
k.	[wa]-[a]	vs.	[sa]-[a]	30 (83%)	6	$<$ .001*
l.	[wa]-[a]	vs.	[sa]-[a]	31 (86%)	5	$<$ .001*

in remembering the first pair by the time they heard the second pair. Therefore, a follow-up experiment is planned in which the same comparison will be repeated multiple times.

Second, the current experiment is based on one token of each comparison, and in order to further verify the generality of the results, it would be desirable to prepare multiple tokens of the same contrast pairs pronounced by multiple speakers. In particular, the speaker recorded for the current experiment did not release the word-final stops. The lack of release may be responsible for low perceptibility of oral [place] contrasts, because release bursts convey place distinctions (Stevens and Blumstein, 1978). The result is in fact compatible with what Winter (2003) found—when speakers were asked to estimate differences of minimal pairs, if stop minimal pairs do not have audible release, they were considered as similar as nasal pairs. It would therefore be interesting to test the perceptibility of both released and unreleased oral consonants.

Third, since place assimilation takes place in preconsonantal positions rather than in word-final positions, it would be interesting to include comparisons like [amka]-[anka] vs. [abka]-[adka]. However, since English has nasal place assimilation, and since this property affects English speakers' speech perception patterns (Darcy et al., 2009), this comparison needs to be tested in languages which do not show any place assimilation.

More generally, the studies reported here are admittedly preliminary, and we need to test listeners from other languages to investigate the robustness of the perceptual asymmetries under question. We also need to test the hypotheses about perceptibility of different contrasts in other experimental methods such as identification/discrimination experiments under noise and magnitude estimation tasks. In summary, the experiments support only a subset of proposed hypotheses but open up possibilities for further experimentation.

7 Conclusion

Optimality Theory has allowed us to address issues that have been hitherto impossible to ask. The principle of faithfulness has opened up the possibility that phonology may encode phonetic perceptibility in phonology. Correspondence Theory's formalization of faithfulness captures both our quotidian speech behavior and verbal art patterns. While these research programs have produced interesting results, the hypotheses on the phonetic grounding of phonological patterns should be tested experimentally.

References

Beckman, J. (1998) *Positional Faithfulness*. Doctoral dissertation, University of Massachusetts, Amherst.

Benua, L. (1997) *Transderivational Identity: Phonological Relations between Words*. Doctoral dissertation, University of Massachusetts, Amherst.

Blevins, J. (2004) *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge: Cambridge University Press.

Blumenfeld, L. (2006) *Constraints on Phonological Interactions*. Doctoral dissertation, Stanford University.

Boersma, P. (2006) "A programme for bidirectional phonology and phonetic and their acquisition and evolution." Ms. University of Amsterdam.

Boersma, P. and Weenink, D. (1999-2009) "Praat: doing phonetics by computer." A software.

Chomsky, N. (1995) *The Minimalist Program*. Cambridge, MA: MIT Press.

Darcy, I., Ramus, F., Christophe, A., Kinzler, K. and Dupoux, E. (2009) "Phonological knowledge in compensation for native and non-native assimilation." In Frank, K., Féry, C. and van de Vijver, R. (eds.) *Variation and Gradience in Phonetics and Phonology*. Berlin & New York: Mouton de Gruyter.

de Lacy, P. (to appear) "Markendess and faithfulness constraints." In van Oostendorp, M., Ewen, C. J., Hume, E. and Rice, K. (eds.) *The Blackwell companion to phonological theory*. Oxford: Blackwell-Weiley.

Fabb, N. (1997) *Linguistics and Literature: Language in the Verbal Arts of the World*. Oxford: Basil Blackwell.

Flemming, E. (1995) *Auditory Representations in Phonology*. Doctoral dissertation, UCLA.

Frisch, S., Pierrehumbert, J. and Broe, M. (2004) "Similarity avoidance and the OCP." Natural Language and Linguistic Theory 22, pp. 179–228.

Goldsmith, J. (1976) *Autosegmental Phonology*. Doctoral dissertation, MIT. Published by Garland Press, New York , 1979.

Halle, M. (1995) "Letter commenting on Burzio's "the rise of optimality thoery"." Glot International 1(9/10), pp. 27–28.

Hawkins, J. and Cutler, A. (1988) "Psycholinguistic factors in morphological asymmetry." In Hawkins, J. A. (ed.) *Explaining Language Universals*., (pp. 280–317). Oxford: Basil Blackwell.

Hayes, B. and Steriade, D. (2004) "Introduction: The phonetic bases of phonological markedness." In Hayes, B., Kirchner, R. and Steriade, D. (eds.) *Phonetically Based Phonology*., (pp. 1–33). Cambridge: Cambridge University Press.

Holtman, A. (1996) *A Generative Theory of Rhyme: An Optimality Approach*. Doctoral dissertation, Utrecht Institute of Linguistics.

Hume, E. and Johnson, K. (2001) "A model of the interplay of speech perception and phonology." In Hume, E. and Johnson, K. (eds.) *The role of speech perception in phonology*. San Diego: Academic Press.

Itô, J. and Mester, A. (1986) "The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility." Linguistic Inquiry 17, pp. 49–73.

Jun, J. (2004) "Place assimilation." In Hayes, B., Kirchner, R. and Steriade, D. (eds.) *Phonetically-based Phonology*, (pp. 58–86). Cambridge: Cambridge University Press.

Kawahara, S. (2006) "A faithfulness ranking projected from a perceptibility scale: The case of voicing in Japanese." Language 82, pp. 536–574.

Kawahara, S. (2007) "Half-rhymes in Japanese rap lyrics and knowledge of similarity." Journal of East Asian Linguistics 16, pp. 113–144.

Kawahara, S. (2009) "Probing knowledge of similarity through puns." In Shinya, T. (ed.) *Proceedings of Sophia Linguistics Society*, volume 23, (pp. 110–137). Tokyo: Sophia University Linguistics Society.

Kawahara, S. (to appear) "Experimental approaches in generative phonology." In van Oostendorp, M., Ewen, C. J., Hume, E. and Rice, K. (eds.) *The Blackwell companion to phonological theory*. Oxford: Blackwell-Weiley.

Kawahara, S. and Shinohara, K. (2009) “The role of psychoacoustic similarity in Japanese puns: A corpus study.” Journal of Linguistics 45, pp. 111–138.

Kawahara, S. and Shinohara, K. (to appear) “Phonetic and psycholinguistic prominences in pun formation: Experimental evidence for positional faithfulness.” In McClure, W. and den Dikken, M. (eds.) *Japanese/Korean Linguistics 18*. Stanford: CSLI.

Kenstowicz, M. (2003) “Salience and similarity in loanword adaptation: A case study from Fijian.” Ms. MIT.

Kiparsky, P. (1973) “The role of linguistics in a theory of poetry.” In Freeman, D. (ed.) *Essays in Modern Stylistics*, (pp. 9–23). London: Methuen.

Lombardi, L. (2001) “Why place and voice are different: Constraint-specific alternations in Optimality Theory.” In Lombardi, L. (ed.) *Segmental Phonology in Optimality Theory*, (pp. 13–45). Cambridge, UK: Cambridge University Press.

Malécot, A. (1956) “Acoustic cues for nasal consonants: An experimental study involving a tape-splicing technique.” Language 32, pp. 274–84.

Massaro, D. and Cohen, M. (1983) “Phonological context in speech perception.” Perception & Psychophysics 34, pp. 338–348.

McCarthy, J. J. (2008a) *Doing Optimality Theory*. Oxford: Blackwell-Weiley.

McCarthy, J. J. (2008b) “The serial interaction of stress and syncope.” Natural Language and Linguistic Theory 26, pp. 499–546.

McCarthy, J. J. and Prince, A. (1995) “Faithfulness and reduplicative identity.” In Beckman, J., Walsh Dickey, L. and Urbanczyk, S. (eds.) *University of Massachusetts Occasional Papers in Linguistics 18*, (pp. 249–384). Amherst: GLSA.

Mohanan, K. P. (1993) “Fields of attraction in phonology.” In Goldsmith, J. (ed.) *The Last Phonological Rule: Reflections on Constraints and Derivations*, (pp. 61–116). Chicago: University of Chicago Press.

Mohr, B. and Wang, W. S. (1968) “Perceptual distance and the specification of phonological features.” Phonetica 18, pp. 31–45.

Moreton, E. (2002) “Structural constraints in the perception of English stop-sonorant clusters.” Cognition 84, pp. 55–71.

Myers, J. and Well, A. (2003) *Research Design and Statistical Analysis*. Mahwah: Lawrence Erlbaum Associates Publishers. Mahwah.

Myers, S. (1997) “Expressing phonetic naturalness in phonology.” In Roca, I. (ed.) *Derivations and Constraints in Phonology*, (pp. 125–152). Oxford: Oxford University Press.

Myers, S. (2002) “Gaps in factorial typology: The voicing in consonant clusters.” Ms. University of Texas.

Nishimura, K. (2003) *Lyman’s Law in loanwords*. MA thesis, Nagoya University.

Ohala, J. J. (1983) “The origin of sound patterns in vocal tract constraints.” In MacNeilage, P. (ed.) *The Production of Speech*, (pp. 189–216). New York: Springer-Verlag.

Pater, J. (1999) “Austronesian nasal substitution and other NC effects.” In Kager, R., van der Hulst, H. and Zonneveld, W. (eds.) *The Prosody-Morphology Interface*, (pp. 310–343). Cambridge: Cambridge University Press.

Pater, J. (to appear) “Weighted constraints in generative linguistics.” Cognitive Science .

Pols, L. (1983) “Three mode principle component analysis of confusion matrices, based on the identification of Dutch consonants, under various conditions of noise and reverberation.” Speech Communication 2, pp. 275–293.

Prince, A. (1997) “Elsewhere & otherwise.” Ms. Rutgers University.

Prince, A. and Smolensky, P. (1993/2004) *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden and Oxford: Blackwell.

Shirai, S. (2002) *Gemination in Loans from English to Japanese*. MA thesis, University of Washington.

Steriade, D. (2001a) “Directional asymmetries in place assimilation: A perceptual account.” In Hume, E. and Johnson, K. (eds.) *The Role of Speech Perception in Phonology*, (pp. 219–250). New York: Academic

Press.

Steriade, D. (2001b) "What to expect from a phonological analysis." Talk presented at LSA 77, Washington, D.C.

Steriade, D. (2001/2008) "The phonology of perceptibility effect: The p-map and its consequences for constraint organization." In Hanson, K. and Inkelas, S. (eds.) *The nature of the word*, (pp. 151–179). Cambridge: MIT Press. Originally circulated in 2001.

Steriade, D. (2003) "Knowledge of similarity and narrow lexical override." In Nowak, P. M., Yoquelet, C. and Mortensen, D. (eds.) *Proceedings of the twenty-ninth annual meeting of the Berkeley Linguistics Society*, (pp. 583–598). Berkeley: BLS.

Stevens, K. and Blumstein, S. (1978) "Invariant cues for place of articulation in stop consonants." Journal of Acoustical Society of America 64, pp. 1358–1368.

Tateishi, K. (1990) "Phonology of Sino-Japanese morphemes." In Lamontagne, G. and Taub, A. (eds.) *University of Massachusetts Occasional Papers in Linguistics 13*, (pp. 209–235). Amherst: GLSA.

Tesar, B. (2007) "A comparison of lexicographic and linear numeric optimization using violation difference ratios." Ms. Rutgers University.

van Oostendorp, M. (2008) "Incomplete devoicing in formal phonology." Lingua 118, pp. 1362–1374.

Wilson, C. (2001) "Consonant cluster neutralization and targeted constraints." Phonology 18, pp. 147–197.

Wilson, C. (2006) "Learning phonology with substantive bias: An experimental and computational study of velar palatalization." Cognitive Science 30, pp. 945–980.

Winter, S. (2003) *Empirical Investigations into the Perceptual and Articulatory Origins of Cross-Linguistic Asymmetries in Place Assimilation*. Doctoral dissertation, Ohio State University.

Wright, R. (2004) "A review of perceptual cues and cue robustness." In Hayes, B., Kirchner, R. and Steriade, D. (eds.) *Phonetically Based Phonology*, (pp. 34–57). Cambridge: Cambrdige University Press.

Zuraw, K. (2007) "The role of phonetic knowledge in phonological patterning: Corpus and survey evidence from Tagalog reduplication." Language 83, pp. 277–316.

Zuraw, K. and Lu, Y.-A. (2009) "Diverse repairs for multiple labial consonants." Natural Language and Linguistic Theory 27, pp. 197–224.

Zwicky, A. (1976) "This rock-and-roll has got to stop: Junior's head is hard as a rock." In Mufwene, S., Walker, C. and Steever, S. (eds.) *Proceedings of Chicago Linguistic Society 12*, (pp. 676–697). Chicago: CLS.

2. Kawahara, Shigeto (2009) Probing knowledge of similarity through puns. In T. Shinya (ed.) Proceedings of Sophia University Linguistic Society 23. Tokyo: Sophia University Linguistics Society. pp. 111-138.

Probing knowledge of similarity through puns*

Shigeto Kawahara
Rutgers University, the State University of New Jersey
kawahara@rci.rutgers.edu

Abstract

This paper outlines the aims, results and future prospects of a general research program which investigates knowledge of similarity through the investigation of Japanese imperfect puns (*dajare*). I argue that speakers attempt to maximize the similarity between corresponding segments in composing puns, just as in phonology where speakers maximize the similarity between, for example, inputs and outputs. In this sense, we find non-trivial parallels between phonology and pun patterns. I further argue that we can take advantage of these parallels, and use puns to investigate our linguistic knowledge of similarity. To develop these arguments, I start with an overview of the results of some recent projects, and follow that with patterns that provide interesting lines of future research. I hope that this paper stimulates further research in this area, which is much understudied in the linguistic literature.

*Acknowledgements: I would like to thank Donca Steriade, whose colloquium talk at UMass, Amherst (Steriade, 2002) has inspired this general project. I would also like to thank the audiences at Sophia University (July 19th, 2008) and the Conference on Language, Communication & Cognition (LCC) at University of Brighton (Aug 4th, 2008). Kazuko Shinohara is a principle co-investigator of this general project. Section 2 is based on Kawahara & Shinohara (2009). The first experiment in section 3 builds on a BA thesis research by Nobuhiro Yoshida and was presented at LCC. I am grateful to Yuuya Takahashi for his technical assistance, as well as to Kazu Kurisu, Kazuko Shinohara, Takahito Shinya, Kyoko Yamaguchi, and Betsy Wang for comments on earlier versions of this draft. Some of the results reported below are preliminary—those who are interested in taking up any of the projects outlined below should feel free to contact the author at kawahara@rci.rutgers.edu.

1 Introduction

The notion of "*similarity*" plays an important role in recent phonological theories. For example, speakers seem to maximize the similarity between inputs and outputs, and this observation is expressed as a general notion of faithfulness constraints in Optimality Theory (Prince & Smolensky, 1993/2004). Speakers are also known to maximize the similarity between morphologically related words (Benua, 1997; Kenstowicz, 1996; Steriade, 2000). However, a question remains as to *what kind of similarity* speakers deploy to shape phonological patterns. Some recent proposals argue that speakers make use of *psychoacoustic* or *perceptual* similarity (Fleischhacker, 2005; Kawahara, 2006; Steriade, 2000, 2001; Zuraw, 2007, among others) while others argue for the importance of more abstract, phonological similarity (see Bailey & Hahn, 2005; Frisch, Pierrehumbert, & Broe, 2004; Kawahara, 2007; LaCharité & Paradis, 2005, for various models of similarity).

Against this background, this paper provides an overview of a general research project, which aims to investigate our linguistic knowledge of similarity through the study of Japanese imperfect puns. The first goal of this project is to describe and analyze the patterns of Japanese imperfect puns. The second goal of this project is to show that speakers minimize the differences between corresponding elements in puns, just as in phonology where speakers minimize the differences between corresponding segments. The third goal is, based on that premise, to investigate our knowledge of similarity through the analysis of imperfect puns. To achieve these goals, in this paper I summarize two major projects that have been completed. In section 2, I show that speakers attempt to maximize psychoacoustic similarity between corresponding segments in puns, and in section 3, I show that speakers avoid disparities in phonetically and psycholinguistically strong positions. It is my hope that this paper stimulates further research in this area, which has been understudied in the literature. To this end, in section 4 of this paper, I outline some patterns of imperfect puns that are yet to be investigated.

In the rest of this introductory section, I briefly introduce how Japanese speakers create puns. Japanese puns, or *dajare*, are very common. Speakers compose puns by creating sentences using identical/similar words or phrases, as in (1) and (2).[1]

[1]Speakers can also change an underlying form to achieve better resemblance with the corresponding word. For example, in *Hokkaidoo-wa dekkai do* 'Hokkaido is big.', speakers change the final conversational particle /zo/ to [do] to make /dekkai zo/ more similar to [hokkaido]. Speakers can also replace a part of proper names, clichés, or famous phrases with a similar sounding

(1) Arumikan-no ue-ni aru mikan.
Aluminum can-GEN top-LOC exist orange
'An orange on an aluminum can.'

(2) Aizu-san-no aisu.
Aizu-from-GEN ice cream
'Ice cream from Aizu.'

The first example in (1) involves an identical sequence of sounds, [arumikan].[2] The second example on the other hand involves a pair of two similar phrases, *aizu* and *aisu*. I will henceforth refer to the examples that involve non-identical matching sequences as "imperfect puns". Intuitively, we can already see that speakers attempt to maximize the similarity between two corresponding elements in puns. In the first example, the corresponding sound sequences are identical, and in the second example, the two words are almost identical except for the [z]-[s] pair, which involves nevertheless similar consonants (see Fleischhacker, 2005; Zwicky & Zwicky, 1986, for a similar observation about English imperfect puns). However, two questions still remain: (i) is the intuition supported by independent evidence beyond our intuition? (ii) if so, what kind of similarity is important in shaping pun patterns? Section 2 addresses these questions.

2 Consonant correspondence

This section[3] reviews a previous analysis of consonant correspondence in imperfect puns by Kawahara & Shinohara (2009). The aim of this general project was to answer the following question: what kind of similarity do speakers use to make

word. For instance, one finds a pun like *Maccho-ga uri-no shoojo* 'A girl who's proud to be a macho', which is based on *Macchi uri-no shoojo* 'The Little Match Girl'—*macchi* is replaced with *maccho-ga*. We have not analyzed these patterns yet, but they are certainly worth further investigations.

[2]Accents differ, however. The first [arumikan] is unaccented, whereas the second [arumikan] has accents on the first [a] and [i]. It would be interesting to investigate how accentual mismatches affect the formation of puns.

[3]Due to space limitation, the following discussion is brief. For further methodological details, discussions and references, see Kawahara & Shinohara (2009) (Kawahara, Shigeto & Kazuko Shinohara "The role of psychoacoustic similarity in Japanese imperfect puns", *Journal of Linguistics*, 45(1), ©Cambridge University Press, forthcoming (March 2009)). I am grateful to Cambridge University Press for allowing me to reproduce the data and Figure 1 from the paper.

puns? Before moving on, a remark on our methodology is in order. A traditional, 'introspection-based' approach to generative phonology seems inappropriate to address this question, because the distinction between, say, featural similarity and psychoacoustic similarity can be subtle, and one can hardly be unbiased about this sort of judgments (see Schütze, 1996, for critical discussions on a purely introspection-based approach to linguistics). Instead we took on a corpus-based approach to find general, statistical patterns in pun pairing.

2.1 Method

To investigate how similarity influences the formation of Japanese imperfect puns, Kawahara & Shinohara (2009) collected 2371 examples of imperfect puns from online websites and consultations with native Japanese speakers. We defined corresponding domains in imperfect puns as sequences of moras in which corresponding vowels are identical.[4] For example, in *Aizu-san no aisu*, the corresponding domains include the first three moras (*aizu* and *aisu*). In the domains defined as such, we compiled pairs of corresponding consonants between the two words. We ignored identical pairs of consonants, because we were interested in similarity, not identity. Therefore, from the example *Aizu-san no aisu*, for instance, we extracted the [z-s] pair. We coded the consonant pairs based on surface forms, instead of phonemic forms. For instance, [ʃi] is arguably derived from /si/, but we considered its onset consonant as [ʃ] (in this regard, we followed Kawahara, Ono & Sudo (2006) and Walter (2007) who suggest that similarity is based on surface forms rather than phonemic forms). We counted only onset consonants, because Japanese coda consonants place-assimilate to the following consonants (Itô, 1986). We ignored singleton/geminate differences, because we were focusing on segmental similarity.

The aim of this project was to investigate what kinds of consonant pairs speakers like to use in imperfect puns. For this purpose, one needs a measure of combinability between two segments. Absolute frequencies are of little use (Trubetzkoy, 1969). Consider the following analogy. The probability of 'coming to school' and 'having breakfast' is perhaps higher than the probability of 'going to a supermarket' and 'being hit by lightning'. However, just because the former combination

[4]There are examples like *Haideggaa-no zense-wa hae dekka?* 'Was Heidegger a fly in his previous life?' in which the corresponding words have one internal vowel mismatch. At the time of data collection, we did not find many examples of this kind, so we took the definition as defined above—however, see section 4.1 for the analysis of examples with vowel mismatches.

has a higher probability, it does not mean that 'coming to school' and 'having breakfast' are better combined than 'going to a supermarket' and 'being hit by lightning': the probability of 'being hit by lightning' is low in the first place. Instead of absolute frequencies, therefore, we need a measure of the frequencies of two combined events relativized with respect to their individual frequency.

For this reason, we used O/E ratios as a measure of how well two consonants combine. O/E ratios are ratios between how often a pair is observed (O-values) and how often it is expected to occur if two elements are combined at random (E-values) (e.g. Frisch et al., 2004). Mathematically, the O-value of a sound [A] is how many times [A] occurs in the corpus, and the E-value of a pair [A-B] is $P(A) \times P(B) \times N$ (where $P(X)$ = the probability of the sound [X] to occur in the corpus; N = the total number of consonants). The higher the O/E ratios, the more likely the two consonants are combined. In performing an O/E analysis, we excluded consonants whose O-values are less than 20 because including them yielded exceedingly high O/E ratios: combining two rare consonants yields a very low E-value, and any observed pair of that type would result an artificially high O/E ratio (e.g. the O/E ratio of the [p^j-b^j] pair is 1121.7). Therefore, we excluded [ts], [dʒ], [j], [n^j], [r^j], and all non-coronal palatalized consonants. As a result, 535 pairs of consonants were left for the subsequent analysis.

2.2 Featural similarity

To statistically verify the correlation between combinability in puns and similarity, as a first approximation, Kawahara & Shinohara (2009) estimated the similarity of consonant pairs in terms of how many feature specifications they share in common (e.g. Bailey & Hahn, 2005). (I later discuss the importance of psychoacoustic similarity in section 2.3 and argue that featural similarity fails to account for some detailed aspects of the pun patterns. I nevertheless start with this simple version of featural similarity in order to first statistically establish the correlation between similarity and combinability.)

To estimate the numbers of shared feature specifications among the set of distinctive consonants in Japanese, Kawahara & Shinohara (2009) deployed eight features: [sonorant], [consonantal], [continuant], [nasal], [strident], [voice], [palatalized], and [place]. According to this system, for example, [s] and [ʃ] are considered to be highly similar because they share seven distinctive features, whereas a pair like [ʃ]-[m] agrees only in [cons], and is considered as a highly non-similar pair.

Figure 1 shows the correlation between featural similarity and combinability in

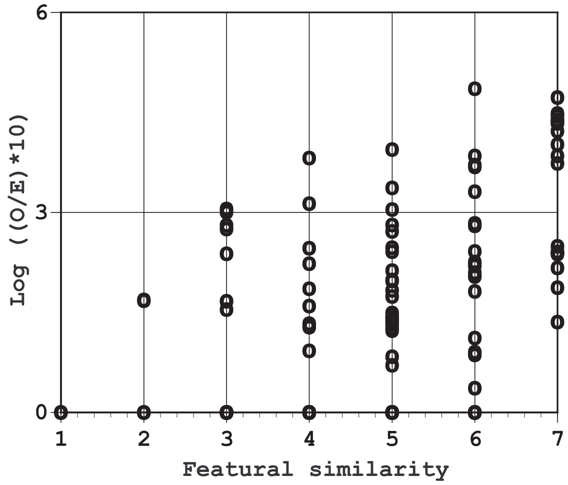

Figure 1: The correlation between featural similarity (the number of shared feature specifications) and combinability ($\log_e((O/E) * 10)$).

puns. For each consonant pair, it plots the number of shared feature specifications on the x-axis and plots the natural log-transformed O/E ratios multiplied by 10 ($\log_e((O/E)*10)$) on the y-axis. We applied log-transformation so as to fit all the data points in a small graph. Since log(0) is undefined, we replaced $\log_e(0)$ with 0. However, the log of O/E ratios smaller than 1 are negative, and hence would be incorrectly treated as values smaller than zero (e.g. $.5 > 0$, but $\log_e(.5) < 0$). Therefore, we multiplied O/E ratios by 10 before log-transformation. The plot excludes pairs in which one member is null (e.g. [ϕa]-[ta]) since it is difficult to define ϕ in terms of distinctive features; I discuss these pairs in section 2.3.

Figure 1 shows the positive correlation between combinability and featural similarity: the more feature specifications a pair shares, the higher the corresponding O/E ratio is ($\rho = .497, t(134) = 6.63, p < .001$). This statistically significant correlation shows that speakers tend pair similar consonants when they compose puns.

2.3 Psychoacoustic similarity

The analysis in section 2.2 used distinctive features to estimate similarity. Kawahara & Shinohara (2009) argue however that featural similarity ultimately fails to

Table 1: The O/E ratios of minimal pairs differing in place.

m-n:	8.85	b-d:	1.09	p-t:	1.11
		b-g:	.65	p-k:	1.08
		d-g:	.39	t-k:	.87

capture some of the detailed aspects of pun pairings. Rather Japanese speakers use psychoacoustic similarity or *perceived similarity between sounds, which refers to acoustic details*. One example that supports the role of psychoacoustic similarity is the [r-d] pair, whose O/E ratio is 3.99. According to the featural similarity measure, the pair agrees in six features (all but [son] and [cont]), but the average O/E ratio of other consonant pairs agreeing in six distinctive feature is 1.49. The 95% confidence interval of the O/E ratios of such pairs is .39–2.59, and 3.99 is outside of the interval. Thus the high O/E ratio of the [r-d] pair is unexpected in terms of featural similarity, but makes sense from a psychoacoustic perspective. Japanese [r] is a flap which involves a ballistic constriction (Nakamura, 2002). [r] and [d] auditorily resemble each other in that they are both voiced consonants with short constrictions (Price, 1981).

Below I present four more kinds of arguments for pun composers exploiting psychoacoustic—rather than featural—similarity. First, speakers treat [place] in oral consonants and [place] in nasal consonants differently (2.3.1). Second, speakers consider a [voice] mismatch less serious than a mismatch in other features (2.3.2). Third, speakers are sensitive to similarity contributed by voicing in sonorants, a phonologically inert feature (2.3.3). Finally, pun-makers pair ϕ with consonants that sound similar to ϕ (2.3.4).

2.3.1 Sensitivity to context-dependent salience of [place]

The first piece of evidence for the importance of psychoacoustic similarity is the fact that speakers treat [place] in oral consonants and [place] in nasal consonants differently. Table 1 lists the O/E ratios of minimal pairs of consonants differing in place.

In Table 1, the [m-n] pair has a higher O/E ratio than any other minimal pair (by a binomial-test, $p = .5^6 = .034$). This pattern shows that Japanese speakers treat the [m-n] pair as more similar than any minimal pairs of oral consonants:

they treat the place distinction in nasal consonants as less salient than in oral consonants.

We find support for the lower perceptibility of [place] in nasals in some previous phonetic and psycholinguistic studies. First, a similarity judgment experiment shows that nasal minimal pairs are considered perceptually more similar to each other than oral consonant minimal pairs (Mohr & Wang, 1968). Second, it has been shown that Dutch speakers perceive [place] in nasals less accurately than [place] in oral consonants (Pols, 1983). Place cues are less salient in nasal consonants than in oral consonants for two reasons: (i) formant transitions into and out of the neighboring vowels are obscured by coarticulatory nasalization and (ii) burst spectra play an important role in distinguishing different places of articulation (Stevens & Blumstein, 1978), but nasals have weak or no bursts (see Jun, 1995, 2004, for relevant discussion).

In short, speakers take into account the lower perceptibility of [place] in nasals, and the lower perceptibility of [place] in nasals has a psychoacoustic root: the blurring of formant transitions and weak bursts. The data in Table 1 therefore shows that speakers use psychoacoustic similarity in composing puns.

2.3.2 Sensitivity to different saliency of different features

The second argument that speakers deploy psychoacoustic rather than featural similarity comes from the fact that they treat a mismatch in [voice] as less disruptive than a mismatch in other manner features. This patterning accords well with previous claims about the perceptibility of [voice] and its phonological behavior. The weaker perceptibility of [voice] is supported by previous psycholinguistic findings, such as Multi-Dimensional Scaling (Peters, 1963; Walden & Montgomery, 1975), a similarity judgment task (Broecke, 1976; Greenberg & Jenkins, 1964), and an identification experiment under noise (Wang & Bilger, 1973). Steriade (2001), building on this psycholinguistic work, argues that [voice] is phonologically more likely to neutralize than other manner features because the change in [voice] is less perceptible. Kenstowicz (2003) further observes that in loanword adaptation, when the recipient language has only a voiceless stop and a nasal, the original voiced stops always map to the stop—i.e. [d] is always borrowed as [t] but not as [n]. Again, this observation follows if the change in [voice] is less perceptible than the change in other features, assuming that speakers attempt to minimize the perceptual changes caused by their phonology.

Given the relatively low perceptibility of [voice], if speakers are sensitive to

Table 2: The O/E ratios of minimal pairs differing in three manner features.

cont		nasal		voice	
p-ɸ:	5.58	b-m:	4.68	p-b:	8.51
t-s:	.90	d-n:	1.12	t-d:	7.64
d-z:	1.68			k-g:	8.03
				s-z:	11.3
				ʃ-ʒ:	6.81

psychoacoustic similarity, we expect that pairs of consonants that are minimally different in [voice] should combine more frequently than minimal pairs that disagree in other manner features. Table 2 compares the O/E ratios for the minimal pairs defined by [cont], [nasal], and [voice].

The O/E ratios of the minimal pairs that differ in [voice] are significantly larger than those of the minimal pairs that differ in [cont] or [nasal] ($Wilcoxon\ W = 15, z = 2.61, p < .01$). Speakers treat minimal pairs that differ only in [voice] as very similar, which indicates that they know that a disagreement in [voice] does not disrupt perceptual similarity as much as other manner features.[5] In other words, Japanese speakers know the varying perceptual salience of different features.

2.3.3 Sensitivity to similarity contributed by voicing in sonorants

The third piece of evidence that speakers resort to psychoacoustic similarity is that speakers are sensitive to similarity contributed by voicing in sonorants—

[5] The following question was raised during my talk at Sophia University: wouldn't we be able to explain the high O/E ratios of the minimal pairs differing in [voice] by saying that they are orthographically similar? Indeed, Japanese orthography distinguishes these minimal pairs by a diacritic (two raised dots). While orthography may play a role, the high combinability of minimal pairs differing in [voice] is attested in pun and rhyming patterns of other languages (Eekman, 1974; Steriade, 2003; Zwicky, 1976; Zwicky & Zwicky, 1986), suggesting that orthography is not the only factor. Moreover, in Japanese orthography, [h] and [b] are distinguished by the same diacritic, but the O/E ratio of this pair is 2.6, which is lower than the values of pairs differing in voicing in Table 2. I thus conclude that orthography cannot be the only reason for the high combinability of consonants differing in [voice].

Table 3: The combinability of voiced and voiceless obstruents with sonorants.

	Voiced obstruents	Voiceless obstruents
Paired with sonorants	63 (**18.2%**)	30 (**6.0%**)
Total	346	497

a phonologically inert/redundant feature. Table 3 shows how often Japanese speakers combine sonorants with voiced obstruents and voiceless obstruents in our database.

63 out of 346 tokens of voiced obstruents are paired with sonorants, whereas 30 out of 497 tokens of voiceless obstruents are paired with sonorants. The probability of voiced obstruents corresponding with sonorants (.18; *s.e.*=.02) is higher than the probability of voiceless obstruents corresponding with sonorants (.06; *s.e.* = .01) ($z = 5.22, p < .001$). Table 3 thus suggests that Japanese speakers treat sonorants as being more similar to voiced obstruents than to voiceless obstruents

The claim that speakers use psychoacoustic similarity when forming puns correctly predicts sonorants are closer to voiced obstruents than to voiceless obstruents. First, both sonorants and voiced obstruents have low frequency energy during the constriction, but voiceless obstruents do not. Second, Japanese voiced stops, especially [g], are often lenited intervocalically, resulting in formant continuity, much like sonorants (see Kawahara, 2006; Kawahara & Shinohara, 2009, for acoustic evidence). For these reasons, voiced obstruents are acoustically more similar to sonorants than voiceless obstruents are.

On the other hand, if speakers were using featural, rather than psychoacoustic, similarity, the pattern in Table 3 is not predicted, given the behavior of [+voice] in Japanese sonorants. Phonologically, voicing in Japanese sonorants behaves differently from voicing in obstruents: Japanese requires that there be no more than one "voiced segment" within a stem, but only voiced obstruents, not voiced sonorants, count as "voiced segments". Previous studies have thus proposed that in Japanese, [+voice] in sonorants is underspecified (Itô & Mester, 1986), sonorants do not bear the [voice] feature at all (Mester & Itô, 1989), or sonorants and obstruents bear different [voice] features (Rice, 1993). No matter how we featurally differentiate voicing in sonorants and voicing in obstruents, sonorants and voiced obstruents do not share the same phonological feature for voicing. Therefore, the featural similarity view—augmented with underspecification or structural differences between sonorant voicing and obstruent voicing—makes an incorrect prediction that sonorants are equidistant from voiceless obstruents and voiced obstruents.

2.3.4 Sensitivity to similarity to ϕ

As a final piece of evidence for the importance of psychoacoustic similarity, Kawahara & Shinohara (2009) show that consonants that correspond with ϕ are those that are psychoacoustically similar to ϕ. In some imperfect puns, consonants in one phrase do not have a corresponding consonant in the other phrase (i.e. one syllable is onsetless), as in *Hayamatte ϕayamatta* 'I apologized without thinking carefully'. (3) lists the set of consonants whose O/E ratios with ϕ are larger than 1 (recall that these are consonants which are paired with ϕ more frequently than expected).

(3) [w]: 6.25, [r]: 4.59, [h]: 3.72, [m]: 2.54, [n]: 1.49, [k]: 1.39

Kawahara & Shinohara (2009) argue that the consonants listed in (3) sound similar to ϕ. First, since [w] is a glide, the transition between [w] and the following vowel is not clear-cut, which makes the presence of [w] hard to detect. Myers & Hanssen (2005) demonstrate that given a sequence of two vocoids, listeners may misattribute the transitional portion to the second vocoid, which effectively shortens the percept of the first vocoid. Thus, due to the blurry boundaries and consequent misparsing, the presence of [w] may be perceptually hard to detect.

Next, Japanese [r] is a flap, which involves a ballistic constriction (Nakamura, 2002). The short constriction makes [r] sound similar to ϕ. Third, the propensity of [h] to correspond with ϕ is expected because [h] lacks a superlaryngeal constriction and hence its spectral properties assimilate to neighboring vowels (Keating, 1988), making the presence of [h] difficult to detect.

Fourth, for [m] and [n], Kawahara & Shinohara (2009) speculate that the edges of these consonants with flanking vowels are blurry due to coarticulatory nasalization, causing them to be interpreted as belonging to the neighboring vowels. Downing (2005) argues that the transitions between vowels and nasals can be misparsed due to their blurry transitions, and that the misparsing effectively lengthens the percept of the vowel. As a result of misparsing, the perceived duration of nasals may become shortened.

Finally, [k] extensively coarticulates with adjacent vowels in terms of tongue backness (de Lacy & Kingston, 2006). As a result it fades into its environment and becomes perceptually similar to ϕ.

To summarize, speakers pair ϕ with segments that sound similar to ϕ—especially those that fade into their environments—but not with consonants whose presence is highly perceptible. Stridents like [s] and [ʃ], for example, do not often correspond with ϕ because of their salient long duration and great intensity of the noise spectra (Wright, 1996). Coronal stops coarticulate least with surrounding vowels (de Lacy & Kingston, 2006) and hence they perceptually stand out from their environments. These consonants are unlikely to correspond with ϕ.

On the other hand, featural similarity does not offer a straightforward explanation for the set of consonants in (3). One may point out that (3) includes sonorants, with the exception of [h] and [k], but there is no sense in which sonorous consonants are similar to ϕ phonologically (Kirchner, 1998). We should in fact not consider ϕ as being sonorous; although languages prefer to have sonorous segments in syllable nuclei (Dell & Elmedlaoui, 1985; Prince & Smolensky, 1993/2004), no languages prefer to have ϕ nuclei. Sonority thus fails to explain why speakers prefer to pair ϕ with the set of consonants in (3). Rather the list of consonants in (3) includes sonorants because their phonetic properties make them sound like ϕ.

2.4 Summary and discussion

Japanese speakers perceive similarity between sounds based on acoustic information, and use that knowledge of psychoacoustic similarity to compose imperfect puns. I have presented four kinds of arguments: (i) context-dependent salience of the same feature, (ii) relative salience of different features, (iii) similarity contributed by a phonologically-inert feature, and (iv) similarity between consonants and ϕ.[6]

Finally, let me discuss non-trivial parallels which we found between the imperfect pun patterns and phonological patterns. For example, in Japanese imperfect puns, [m] and [n] are more likely to correspond with each other than [p] and [t] are, just as in other languages' phonology in which [m] and [n] alternate with each other (i.e. nasal place assimilation) while [p] and [t] do not (Jun, 2004; Mohanan,

[6]The next step would be to obtain a matrix of perceived similarity through a psycholinguistic experiment (e.g. a similarity judgment task, an identification/discrimination experiment under noise) and use that matrix to analyze the pun patterns. This is one of the open lines of future research, which an interested reader is more than welcome to take on.

1993). In both verbal art and phonology, nasal pairs are more likely to correspond with each other than oral consonant pairs.

Similarly, in the pun pairing patterns, minimal pairs that differ in voicing are more likely to correspond with each other than other types of minimal pairs that differ in nasality and continuency. This patterning parallels cross-linguistic observation that a voicing contrast is more easily neutralized than other manner features (Kenstowicz, 2003; Steriade, 2001). In both cases, a voicing mismatch between two corresponding elements is tolerated, more so than a mismatch in other features.

Third, voicing in sonorants promotes similarity with voiced obstruents, and this finds its parallel in linguistic similarity avoidance patterns. Arabic, for example, avoids pairs of similar adjacent consonants, and sonorants are treated as more similar to voiced obstruents than to voiceless obstruents (Frisch et al., 2004). Finally, when consonants correspond with ϕ, speakers choose those that fade into their environment. Again, this pattern finds a parallel in phonology: when speakers epenthesize a segment, they typically choose segments that are perceptually least intrusive (usually glottal consonants for consonants and schwa or high vowels for vowels) (Kwon, 2005; Steriade, 2001; Walter, 2004). Based on these results we can conclude that non-trivial parallels exist between verbal art patterns and phonological patterns.

3 Positional effects

In the previous section, I hope to have established two points: (i) speakers minimize the difference between corresponding elements in puns as well as in phonology, and (ii) the similarity is defined in terms of psychoacoustic features. We now turn to two experiments that address the issue of *positional effects* in imperfect puns (Kawahara, Shinohara, & Yoshida, 2008). The analysis in section 2 does not take into consideration *where* in the punning words mismatches occur. However, we would think that positions of mismatches should matter, because in phonology positions of changes matter (Beckman, 1998; Steriade, 1994). In particular, speakers avoid making changes in phonetically and psycholinguistically prominent positions, such as stressed syllables, long vowels, word-initial syllables, and root segments (Beckman, 1998). The question that arises is, therefore, "In puns, would speakers avoid mismatches in phonetically and psycholinguistically strong positions?" By addressing this question, this project aims to further advance the claim that non-trivial parallels exist between pun patterns and phonological patterns. Concretely, we will observe that the principle of position faithfulness—avoidance of disparities in prominent positions—exists both in phonology and in puns.

3.1 Initial positions

3.1.1 Introduction

I start with the discussion of initial syllables. It is well known that speakers maintain more contrasts in initial syllables than in non-initial syllables (Beckman, 1998). For example, in Sino-Japanese, while initial syllables can contain a variety of consonants, second syllables only allow [t] and [k] (Kawahara, Nishimura, & Ono, 2002; Tateishi, 1990). Similarly, mid vowels are found only in initial syllables in Tamil (Beckman, 1998). Assuming that all possible contrasts are allowed in the input *as per* Richness of the Base (Prince & Smolensky, 1993/2004), we can say that speakers avoid changing input specifications particularly in initial syllables in phonology. In Sino-Japanese, if there were an underlying form like /sasu/, then speakers avoid changing the initial [s] but not the final [s] (resulting in, say, [satu]). The theory of *positional faithfulness* directly captures this observation by positing that underlying contrasts in initial syllables are protected by special positional faithfulness constraints (Beckman, 1998).

The privilege of initial syllables may be due to their psycholinguistic salience (see Beckman, 1998; Hawkins & Cutler, 1988; Smith, 2002, for an overview). Initial syllables play an important role in lexical access (word recognition); for example, hearing initial portions of words help listeners to retrieve the whole words (Horowitz, Chilian, & Dunnigan, 1969; Horowitz, White, & Atwood, 1968; Nooteboom, 1981). In "tip-of-the-tongue" phenomena, speakers can guess what the first sound is more accurately than non-initial sounds (A. Brown, 1991; R. Brown & MacNeill, 1966), and word-initial sounds help retrieve the whole word efficiently (Freedman & Landauer, 1966). Listeners detect mispronunciations in non-initial positions faster than those in initial syllables (Cole, 1973; Cole & Jakimik, 1980)—once listeners hear initial syllables, they anticipate what is coming next, so that they can detect mispronunciations quickly. Because of their psycholinguistic prominence, speakers avoid disparities in initial syllables, which are noticeable.[7]

[7]It has also been shown that the effects of sound symbolism—particular images associated with particular sounds—are stronger in initial syllables than in non-initial syllables (Kawahara,

Table 4: A correspondence theoretic illustration of the parallel between phonology and pun formation. The left figure (a)=phonological input-output correspondence. The right figure (b)=surface-to-surface correspondence in pun.

Input	/	s_i	a_j	s_k	u_l	/	Word 1	[	s_i	a_j	s_k	u_l	]
		|	|	|	|				|	|	|	|	
Output	[	s_i	a_j	t_k	u_l	]	Word 2	[	s_i	a_j	t_k	u_l	]
a. IO mapping							b. Pun pairing						

Due to their psycholinguistic prominence, speakers avoid mismatches in initial syllables in phonology. If the same principle governs pun formation, speakers should avoid mismatches in initial syllables in imperfect puns as well. Correspondence Theory (McCarthy & Prince, 1995) allows us to illustrate the parallel in a particularly clear manner, as in Table 4.[8] In this theory, segments at two levels of representations (e.g. input and output) stand in correspondence, and their identity is required by a set of faithfulness constraints. As the left-hand figure (a) shows, speakers can allow a disparity word-internally (i.e. /s_k/-[t_k]) but not word-initially (i.e /s_i/-[s_i]), as is the case in Sino-Japanese. Likewise, as in figure (b), in composing puns speakers may avoid having an non-identical correspondence word-initially much more so than word-internally. What we will observe is precisely this.

3.1.2 Method

In order to control for factors other than positional effects, we performed well-formedness judgment experiments. The first experiment tested whether speakers avoid mismatches in initial positions. The stimuli were pairs of words that contain a pair of sounds that minimally differ in voicing ([t-d], [d-t], [k-g], [g-k], [s-z], [z-

Shinohara, & Uchimoto, 2008).

[8]Several authors have proposed a correspondence theory of rhymes (Hayes & MacEachern, 1998; Holtman, 1996; Steriade, 2003; Steriade & Zhang, 2001; Yip, 1999).

s]).[9] To control for the phonological distance between the punning constituents, the stimuli all had the same structure, [X-particle Y]. The punning constituents X and Y were all three syllables long. In one condition the mismatch occurred in the initial syllables (e.g. **s***asetsu-ni* **z***asetsu* 'I gave up making a left turn.'), and in another condition, the mismatch occurred in the second syllables (*hi***s***ashi-ni hi***z***ashi* 'sunlight on the sun root.').[10] Additional filler items were interwoven with the target questions. The participants rated both the funniness and the acceptability of each pun pair on a 1-to-4 scale for both questions. We used the two different ratings to separate perceptions of wellformedness from comedic value. The questionnaire started with two sample questions, with one example which is clearly an example of a Japanese pun (*Arumikan-no ue-ni aru mikan* 'An orange on an alminum can.') and one example which clearly is not (*Hana-yori dango* 'Foods are more important than love.'). A total of 37 speakers participated in this study, but we excluded eight of them because they did not consider the first example *Arumikan-no ue-ni aru mikan* as a good pun or considered *Hana-yori dango* as a perfect pun.

3.1.3 Results and discussion

Figure 2 illustrates the result of the rating experiment—it plots the average wellformedness ratings for the puns that have initial syllable mismatches and those that have internal syllable mismatches. A within-subject t-test reveals that speakers judged mismatches in initial syllables less acceptable than those in non-initial syllables ($t(28) = 2.69, p < .05$). In other words, speakers avoid mismatches in a psycholinguistically prominent position. Therefore, the principle of positional faithfulness is observed both in puns and in phonology: the parallel illustrated in Table 4 is experimentally supported.

[9] We could not control for matches in accents for two reasons. First, we could not find enough minimal pairs if we controlled accents. Second, lexical accents are subject to high interspeaker variability and therefore it was impossible to find minimal pairs that match in accents for all speakers.

[10] We also included pairs which contained mismatches in final syllables in our experiment (Kawahara, Shinohara, & Yoshida, 2008). These syllables behaved just like initial syllables. This result is a bit unexpected because it is known that initial syllables are psycholinguistically more prominent than final syllables (see in particular Nooteboom, 1981). It may be that recency effects (Gupta, 2005; Gupta, Lipinski, Abbs, & Lin, 2005) are playing a role here—speakers remember the final syllables of the first word most vividly when they find the second punning word, and therefore they avoid mismatches in final syllables because of the vivid memory.

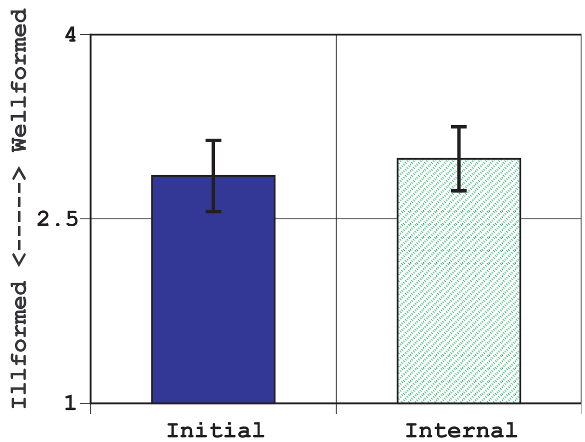

Figure 2: Wellformedness of puns with initial mismatches and those with internal mismatches. The error bars represent 95% confidence intervals across the 29 participants.

3.2 Long vowels

3.2.1 Introduction

We have observed that speakers avoid mismatches in initial positions. We can thus conclude that the principle of positional effects is in effect in the formation of imperfect puns. This conclusion can be reinforced if we observe a different type of positional effect. We investigated whether speakers avoid disparities in long vowels, because they do so in phonology (Kirchner, 1998; Steriade, 1994). Hindi for example allows a surface nasality contrast in long vowels, but not in short vowels (see Steriade, 1994, for references and other cases). A hypothetical underlying /tããtã/ would map to [tããta], as illustrated in the left-hand figure (a) in Table 5. In general, we observe in phonology that speakers avoid making changes—or neutralizing contrasts—in long vowels more often than in short vowels. The question is thus whether we observe the same principle in pun pairings, as in the right-hand figure (b) in Table 5.

Table 5: A correspondence theoretic illustration of the parallel between phonology and pun formation. The left figure (a)=phonological input-output correspondence. The right figure (b)=surface-to-surface correspondence in pun.

Input	/	t_i	$\tilde{a}\tilde{a}_j$	t_k	$\tilde{a}_l$	/		Word 1	[	t_i	$\tilde{a}\tilde{a}_j$	t_k	$\tilde{a}_l$	]
		|	|	|	|					|	|	|	|	
Output	[	t_i	$\tilde{a}\tilde{a}_j$	t_k	a_l	]		Word 2	[	t_i	$\tilde{a}\tilde{a}_j$	t_k	a_l	]
a. IO mapping								b. Pun pairing						

3.2.2 Method

The method is almost identical to the previous experiment, except that we had four practice questions (in addition to the two examples used in the previous experiment, *Manjuu-o mittsu moratta Akechi Mitsuhide-ga "A, kechi, mittsu hidee"* 'Akechi Mitsuhide was given three pieces of manjuu, and said "that's mean, only three?"'—an example of a good pun—and *Dakara, kore-wa zura dewa arimasen* 'I am telling you that this is not a wig.'—an example of non-pun). The design had three fully crossed factors: 10 vowel combinations ([a-i], [a-u], [a-e], [a-o], [i-u], [i-e], [i-o], [u-e], [u-o], [e-o]) × 2 orders (e.g. [a-i] vs. [i-a]) × 2 lengths (short vs. long). An example of a crucial pair was: *jook***uu***-no jook***aa** 'A joker in the sky.' vs. *ripp***u***-ga ripp***a** 'The lips are fine.'. Additional fillers were added and interwoven with the target items. 26 speakers participated in the study. All the participants judged the sample good puns as good puns and sample non-puns as non-puns, and hence the data from all the participants were included in the analysis.

3.2.3 Results

Figure 3 compares the wellformedness of puns with long vowel mismatches and those with short vowel mismatches. As expected, speakers rated those with long mismatches as worse than those with short mismatches ($t(25) = 3.83, p < .001$). In this regard, speakers avoid mismatches in long vowels more than mismatches in short vowels. Mismatches in long vowels are perceptually salient because of their long duration (Steriade, 2003), and hence avoided by the participants.

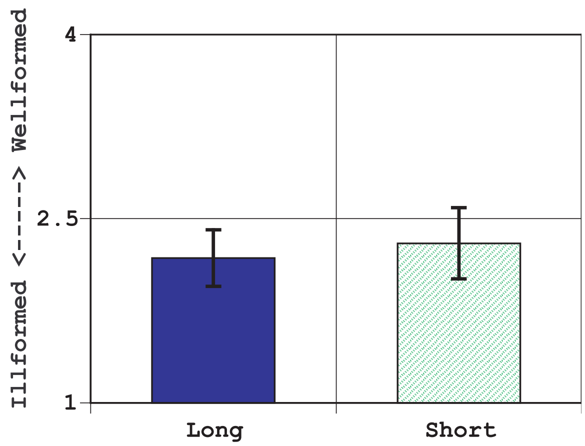

Figure 3: Wellformedness of puns with long vowel mismatches and short vowel mismatches. The error bars represent 95% confidence intervals across the 26 participants.

3.3 Summary and discussion

The two experiments reported in this section show that speakers avoid mismatches in phonetically and psycholinguistically prominent positions in Japanese puns, just as they do so in phonological patterns. This finding further supports the claim that phonological and pun patterns are governed by the same principles.

In addition to finding that positional effects do exist in puns, our results bear on one current theoretical debate: the issue of positional faithfulness theory versus positional markedness theory. To account for patterns of positional neutralization patterns, in which some contrasts only surface in certain positions, the positional faithfulness theory postulates that speakers avoid changes in strong positions (Beckman, 1998): "Avoid input-output mismatches in strong positions". On the other hand, the positional markedness theory postulates that speakers avoid having contrasts in weak positions (Zoll, 1998): "Do not have contrasts in weak positions". As I discussed above, the principle of positional faithfulness can explain our results because we observe that speakers avoid mismatches in strong positions. On the other hand, the positional markedness theory has nothing to say about the results because it evaluates the wellformedness of one form only, but not the relation between two forms.

4 Further issues: other patterns

In addition to the two projects summarized above, there are other types of puns that are worth investigating. In this section, I provide an overview of other interesting types of pun patterns, which can and should be pursued in future research. I strongly hope that this discussion stimulates further researches on imperfect puns. If readers are interested, please contact the author at `kawahara@rci.rutgers.edu`.

4.1 Vowel mismatches

Speakers can combine two words minimally different in one vowel to create a pun sentence, as in (4).

(4) Haideggaa-no zense-wa hae dekka?
Heidegar-GEN previous life-TOP fly copula
'Is Heidegger a fly in his previous life?'

In this example, *Haideggaa* and *haedekka* are paired, even though the second vowels are not identical ([i] vs. [e]). We collected 547 examples of this sort—those that involve pairing of two words that minimally differ in one (short) vowel—and calculated the O/E ratios for each pair. We then took their reciprocals as distances between the five vowels, and created a distance map using an MDS (Multi-Dimensional Scaling) analysis. The results appear in Figure 4.

In Figure 4, the five vowels are configured in a way that closely resembles the vowel space. Moreover, as argued in section 2, speakers seem to use psychoacoustic similarity—rather than featural similarity—to create puns. The distance matrix obtained from our analysis in fact statistically correlates with a distance matrix created based on euclidian distances using F1 and F2 Bark values reported in Hirahara & Kato (1992) ($r = .647, t(8) = 2.40, p = .043$).

There are two primary reasons to believe that psychoacoustic similarity, rather than featural similarity, shapes the pun patterns here. First, in Figure 4, [a] is closer to [o] than [e], although [a] differs from both [o] and [e] in terms of two features (the [a-o] pair disagrees in [low] and [round]; the [a-e] pair disagrees in [low] and [back]). The proximity of [a] to [o] receives a straightforward explanation from a psychoacoustic point of view, because both [a] and [o] have low F2 (Hirahara & Kato, 1992; Keating & Huffman, 1984). Moreover, in Hirahara & Kato's experiment, where they tested the influence of F0 on vowel perception,

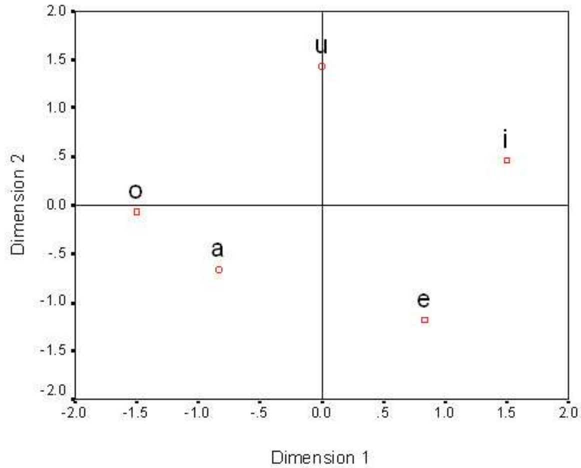

Figure 4: An MDS analysis of the vowel distance matrix computed based on combinability in puns. The distances between the five vowels are calculated as reciprocals of O/E ratios.

they found that raising the F0 of original [a] can result in the percept of [o], but not that of [e].

Second, in Figure 4, [i] and [u] are much closer to each other than [e] and [o] are—this pattern again receives a straightforward explanation because in Japanese [u] is less rounded than [o] is, and hence has higher F2, making it closer to front vowels.[11] Again, in Hirahara & Kato's experiment, raising the F0 of the original [i] could result in the percept of [u], whereas applying the same operation to [e] did not result in the percept of [o].

While we seem to have good evidence that speakers use psychoacoustic similarity when they compute similarity between vowels, in order to further strengthen this point, the matrix should be compared with a perceptual map obtained by some method (a similarity judgment task, a confusion experiment under noise etc).

[11]One may argue that Japanese [u] is phonologically [-rounded] and thus [u] is closer to [i] than [o] is to [e] in terms of distinctive features. However, Japanese [u] is arguably [+rounded] phonologically, because [u] turns preceding /h/ into a rounded, bilabial fricative [ɸ].

4.2 Metathesis

There are a few more other patterns that are yet to be analyzed. For example, some examples of puns can involve metathesis of moras, as in (5) where the order of [ma] and [ne] is reversed:

(5) Shimane-no shinema.
Shimane-GEN cinema
'A cinema in Shimane.'

Given examples like (5), we should investigate what kind of moras/syllables can metathesize with each other. If our hypothesis that speakers minimize the perceptual disparities between corresponding segments is correct, then we should find examples where [ABC] and [ACB] are perceptually similar ([A], [B], [C] stand for CV-moras) (cf. Blevins & Garrett, 1998; Hume, 1998). We have collected some examples that involve metathesis—ideally, one should collect more examples and see what kind of metathesis patterns are common.

4.3 Syllable intrusion

Finally, sometimes speakers combine two words in which one word contains one mora/syllable which is not contained in the other word, as in (6)

(6) Bundoki-o bundottoki.
protractor-ACC take away
'Take away a protractor (from him).'

We would like to know what kind of syllables can be 'inserted'. We have collected examples of this sort, and it seems to us that the majority of the examples fit in one of the three following categories: (i) the intruded syllable has a high vowel (e.g. *Fujiki Naoto-wa fushigi-na o***hi***to* 'Fujiki Naoto is a strange person.'), (ii) the intruded syllable has a vowel identical to the preceding or following vowel (e.g. *Kamigata-ga kami ga***kat***ta* 'My hair-do was divinely done.', and (iii) the intruded syllable is an affix or a particle (e.g. *Kurisumasu-wa kuri***de***sumasu* 'I will have chestnuts for Christmas'). Out of the 149 examples we have collected, 144 fit in one of these patterns. The data are yet to be analyzed to see if these observations are robust, and moreover if the principle of the maximization of psychoacoustic similarity is responsible for shaping these patterns (high vowels, copy vowels, and affixal vowels are similar to ϕ?).

5 General summary

Speakers minimize perceptual disparities between corresponding segments in puns. In this sense, we find non-trivial parallels between pun pairing patterns and phonological patterns. In this paper, I hope to have demonstrated that we can probe our knowledge of similarity through studying puns. Pun patterns are understudied—they thus provide an untapped resource for future research.

References

Bailey, T., & Hahn, U. (2005). Phoneme similarity and confusability. *Journal of Memory and Language*, *52*, 339-362.

Beckman, J. (1998). *Positional faithfulness*. Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Benua, L. (1997). *Transderivational identity: Phonological relations between words*. Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Blevins, J., & Garrett, A. (1998). The origins of consonant-vowel metathesis. *Language*, *74*, 508-56.

Broecke, M. P. R. van den. (1976). *Hierarchies and rank orders in distinctive features*. Amsterdam: Van Gorcum, Assen.

Brown, A. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin*, *109*(2), 204-223.

Brown, R., & MacNeill, D. (1966). The 'tip of the tongue' phenomenon. *Journal of Verbal Learning and Verbal Behavior*, *5*, 325-337.

Cole, R. (1973). Listening for mispronunciations: A measure of what we hear during speech. *Perception & Psychophysics*, *13*, 153-156.

Cole, R., & Jakimik, J. (1980). How are syllables used to recognize words? *Journal of Acoustical Society of America*, *67*, 965-970.

de Lacy, P., & Kingston, J. (2006). *Synchronic explanation.* (ms. University of Massachusetts and Rutgers University)

Dell, F., & Elmedlaoui, M. (1985). Syllabic consonants and syllabification in Imdlawn Tashlhiyt Berber. *Journal of African Languages and Linguistics*, *7*, 105-130.

Downing, L. (2005). On the ambiguous segmental status of nasal in homorganic nc sequences. In M. van Oostendorp & J. M. van der Weijer (Eds.), *The* *internal organization of phonological segments* (p. 183-216). Berlin: Mouton de Gruyter.

Eekman, T. (1974). *The realm of rime: A study of rime in the poety of the Slavs.* Amsterdam: Adolf M. Hakkert.

Fleischhacker, H. (2005). *Similarity in phonology: Evidence from reduplication and loan adaptation.* Unpublished doctoral dissertation, UCLA.

Freedman, J., & Landauer, T. (1966). Retrieval of long-term memory: "tip-of-the-tongue" phenomenon. *Psychonomic Science, 4*, 309-310.

Frisch, S., Pierrehumbert, J., & Broe, M. (2004). Similarity avoidance and the OCP. *Natural Language and Linguistic Theory, 22*, 179-228.

Greenberg, J., & Jenkins, J. (1964). Studies in the psychological correlates of sound system of american English. *Word, 20*, 157-177.

Gupta, P. (2005). Primacy and recency in nonword repetition. *Memory, 13*(3), 318-324.

Gupta, P., Lipinski, J., Abbs, B., & Lin, P.-H. (2005). Serial position effects in nonword repetition. *Journal of Memory and Language, 53*, 141-162.

Hawkins, J., & Cutler, A. (1988). Psycholinguistic factors in morphological asymmetry. In J. A. Hawkins (Ed.), *Explaining language universals.* (p. 280-317). Oxford: Basil Blackwell.

Hayes, B., & MacEachern, M. (1998). Quatrain form in English folk verse. *Language, 64*, 473-507.

Hirahara, T., & Kato, H. (1992). The effect of F0 on vowel identification. In Y. Tohkura, E. V. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production and linguistic structure* (p. 89-112). Tokyo: Ohmsha.

Holtman, A. (1996). *A generative theory of rhyme: An optimality approach.* Unpublished doctoral dissertation, Utrecht Institute of Linguistics.

Horowitz, L., Chilian, P., & Dunnigan, K. (1969). Word fragments and their reintegrative powers. *Journal of Experimental Psychology, 80*, 392-394.

Horowitz, L., White, M., & Atwood, D. (1968). Word fragments as aids to recall: The organization of a word. *Journal of Experimental Psychology, 76*, 219-226.

Hume, E. (1998). Metathesis in phonological theory: The case of Leti. *Lingua, 104*, 147-186.

Itô, J. (1986). *Syllable theory in prosodic phonology.* Unpublished doctoral dissertation, University of Massachusetts, Amherst. (Published 1988. Outstanding Dissertations in Linguistics series. New York: Garland)

Itô, J., & Mester, A. (1986). The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry*, *17*, 49-73.

Jun, J. (1995). *Perceptual and articulatory factors in place assimilation: An optimality theoretic approach*. Unpublished doctoral dissertation, University of California, Los Angeles.

Jun, J. (2004). Place assimilation. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically-based phonology* (p. 58-86). Cambridge: Cambridge University Press.

Kawahara, S. (2006). A faithfulness ranking projected from a perceptibility scale: The case of voicing in Japanese. *Language*, *82*, 536-574.

Kawahara, S. (2007). Half-rhymes in Japanese rap lyrics and knowledge of similarity. *Journal of East Asian Linguistics*, *16*(2), 113-144.

Kawahara, S., Nishimura, K., & Ono, H. (2002). Unveiling the unmarkedness of Sino-Japanese. In W. McClure (Ed.), *Japanese/Korean linguistics 12* (Vol. 12, p. 140-151). Stanford: CSLI.

Kawahara, S., Ono, H., & Sudo, K. (2006). Consonant co-occurrence restrictions in Yamato Japanese. In T. Vance & K. Jones (Eds.), *Japanese/Korean linguistics 14* (Vol. 14, p. 27-38). Stanford: CSLI.

Kawahara, S., & Shinohara, K. (2009). The role of psychoacoustic similarity in Japanese imperfect puns: A corpus study. *Journal of Linguistics*, *45*(1).

Kawahara, S., Shinohara, K., & Uchimoto, Y. (2008). A positional effect in sound symbolism: An experimental study. *Proceedings of Japan Cognitive Linguistics Association*, 417-427.

Kawahara, S., Shinohara, K., & Yoshida, N. (2008). *Positional effects in Japanese imperfect puns.* (A talk presented at Language, Communication, and Cognition (Brighton Univresity, August 4th, 2008))

Keating, P. (1988). Underspecification in phonetics. *Phonology*, *5*, 275-292.

Keating, P., & Huffman, M. (1984). Vowel variation in Japanese. *Phonetica*, *41*, 191-207.

Kenstowicz, M. (1996). Base identity and uniform exponence: alternatives to cyclicity. In J. Durand & B. Laks (Eds.), *Current trends in phonology: models and methods* (Vol. 1, p. 363-393). Salford, Manchester: European studies research institute.

Kenstowicz, M. (2003). *Salience and similarity in loanword adaptation: A case study from Fujian.* (ms. MIT)

Kirchner, R. (1998). *An effort-based approach to consonant lenition*. Unpublished doctoral dissertation, University of California, Los Angeles.

Kwon, B.-Y. (2005). The patterns of vowel insertion in IL phonology: The p-map account. *Studies in Phonetics, Phonology and Morphology*, *11*(2), 21-49.

LaCharité, D., & Paradis, C. (2005). Category preservation and proximity versus phonetic approximation in loanword adaptation. *Linguistic Inquiry*, *36*, 223-258.

McCarthy, J., & Prince, A. (1995). Faithfulness and reduplicative identity. In J. Beckman, L. Walsh Dickey, & S. Urbanczyk (Eds.), *University of Massachusetts occasional papers in linguistics 18* (p. 249-384). Amherst: GLSA.

Mester, A., & Itô, J. (1989). Feature predictability and underspecification: Palatal prosody in Japanese mimetics. *Language*, *65*, 258-93.

Mohanan, K. P. (1993). Fields of attraction in phonology. In J. Goldsmith (Ed.), *The last phonological rule: Reflections on constraints and derivations* (p. 61-116). Chicago: University of Chicago Press.

Mohr, B., & Wang, W. S. (1968). Perceptual distance and the specification of phonological features. *Phonetica*, *18*, 31-45.

Myers, S., & Hanssen, B. (2005). The origin of vowel-length neutralization in vocoid sequences. *Phonology*, *22*, 317-344.

Nakamura, M. (2002). The articulation of the Japanese /r/ and some implications for phonological acquisition. *Phonological Studies*, *5*, 55-62.

Nooteboom, S. (1981). Lexical retrieval from fragments of spoken words: Beginnings vs. endings. *Journal of Phonetics*, *9*, 407-424.

Peters, R. (1963). Dimensions of perception for consonants. *Journal of the Acoustical Society of America*, *35*, 1985-1989.

Pols, L. (1983). Three mode principle component analysis of confusion matrices, based on the identification of Dutch consonants, under various conditions of noise and reverberation. *Speech Communication*, *2*, 275-293.

Price, P. (1981). *A cross-linguistic study of flaps in Japanese and in American English*. Unpublished doctoral dissertation, University of Pennsylvania.

Prince, A., & Smolensky, P. (1993/2004). *Optimality Theory: Constraint interaction in generative grammar*. Malden and Oxford: Blackwell.

Rice, K. (1993). A reexamination of the feature [sonorant]: The status of sonorant obstruents. *Language*, *69*, 308-344.

Schütze, C. (1996). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. University of Chicago Press.

Smith, J. (2002). *Phonological augmentation in prominent positions*. Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Steriade, D. (1994). *Positional neutralization and the expression of contrast.* (ms. University of California, Los Angeles)

Steriade, D. (2000). Paradigm uniformity and the phonetics-phonology boundary. In M. B. Broe & J. B. Pierrehumbert (Eds.), *Papers in laboratory phonology V: Acquisition and the lexicon* (p. 313-334). Cambridge: Cambridge University Press.

Steriade, D. (2001). *The phonology of perceptibility effect: The p-map and its consequences for constraint organization.* (ms. MIT)

Steriade, D. (2002). *Semi-rhymes, relative similarity and phonological knowledge.* (A colloquium talk presented at University of Massachusetts, Amherst, December 2002)

Steriade, D. (2003). Knowledge of similarity and narrow lexical override. In P. M. Nowak, C. Yoquelet, & D. Mortensen (Eds.), *Proceedings of Berkeley Linguistics Society 29* (p. 583-598). Berkeley: BLS.

Steriade, D., & Zhang, J. (2001). *Context-dependent similarity* (Tech. Rep.).

Stevens, K., & Blumstein, S. (1978). Invariant cues for place of articulation in stop consonants. *Journal of Acoustical Society of America*, *64*, 1358-1368.

Tateishi, K. (1990). Phonology of Sino-Japanese morphemes. In G. Lamontagne & A. Taub (Eds.), *University of Massachusetts Occasional Papers in Linguistics 13* (p. 209-235). Amherst: GLSA.

Trubetzkoy, N. (1969). *Principles of phonology.* Berkeley: University of California Press.

Walden, B., & Montgomery, A. (1975). Dimensions of consonant perception in normal and hearing-impaired listeners. *Journal of Speech and Hearing Research*, *18*, 445-455.

Walter, M.-A. (2004). Loan adaptation in Zazaki Kurdish. In M. Kenstowicz (Ed.), *Working papers in endangered and less familiar languages 6* (p. 97-106). Cambridge: MITWPL.

Walter, M.-A. (2007). *Repetition avoidance in human language.* Unpublished doctoral dissertation, MIT.

Wang, M., & Bilger, R. (1973). Consonant confusions in noise: A study of perceptual features. *Journal of Acoustical Society of America*, *54*, 1248-1266.

Wright, R. (1996). *Consonant clusters and cue preservation in Tsou.* Unpublished doctoral dissertation, University of California, Los Angeles.

Yip, M. (1999). Reduplication as alliteration and rhyme. *Glot International*, *4*(8), 1-7.

Zoll, C. (1998). *Positional asymmetries and licensing.* (ms. MIT)

Zuraw, K. (2007). The role of phonetic knowledge in phonological patterning: Corpus and survey evidence from tagalog reduplication. *Language*, *83*(2), 277-316.

Zwicky, A. (1976). This rock-and-roll has got to stop: Junior's head is hard as a rock. In S. Mufwene, C. Walker, & S. Steever (Eds.), *Proceedings of Chicago Linguistic Society 12* (p. 676-697). Chicago: CLS.

Zwicky, A., & Zwicky, E. (1986). Imperfect puns, markedness, and phonological similarity: With fronds like these, who needs anemones? *Folia Linguistica*, *20*, 493-503.

3. Kawahara, Shigeto and Kazuko Shinohara (2009) The role of psychoacoustic similarity in Japanese puns: A corpus study. Journal of Linguistics 45.1: 111-138.

J. Linguistics **45** (2009), 111–138.
doi:10.1017/S0022226708005537 Printed in the United Kingdom

The role of psychoacoustic similarity in Japanese puns: A corpus study[1]

SHIGETO KAWAHARA

Rutgers University, the State University of New Jersey

KAZUKO SHINOHARA

Tokyo University of Agriculture and Technology

(Received 19 July 2007; revised 16 May 2008)

A growing body of recent work on the phonetics–phonology interface argues that many phonological patterns refer to psychoacoustic similarity – perceived similarity between sounds based on detailed acoustic information. In particular, two corresponding elements in phonology (e.g. inputs and outputs) are required to be as psychoacoustically similar as possible (Steriade 2001a, b, 2003; Fleischhacker 2005; Kawahara 2006; Zuraw 2007). Using a corpus of Japanese imperfect puns, this paper lends further support to this claim. Our corpus-based study shows that when Japanese speakers compose puns, they require two corresponding consonants to be as similar as possible, and the measure of similarity rests on psychoacoustic information. The result supports the hypothesis that speakers possess a rich knowledge of psychoacoustic similarity and deploy that knowledge in shaping verbal art patterns.

1. Introduction

A number of recent studies on the phonetics–phonology interface argue that many phonological patterns refer to PSYCHOACOUSTIC SIMILARITY – perceived similarity between sounds based on detailed acoustic information. Expanding on the general notion of faithfulness constraints in Optimality Theory (Prince & Smolensky 1993/2004, McCarthy & Prince 1995), several authors have proposed that speakers attempt to maximize the psychoacoustic similarity between inputs and outputs (Boersma 1998; Steriade 2001a, b, 2003; Côté 2004; Jun 2004; Fleischhacker 2005; Kawahara 2006;

[1] We are grateful to our informants for composing ample examples of Japanese puns for us. We would also like to thank Gary Baker, Michael Becker, John Kingston, Kazu Kurisu, Dan Mash, Lisa Shiozaki, Betsy Wang, and two anonymous *JL* referees for their comments on earlier drafts of this paper. Remaining errors are ours. An earlier version of this paper appeared as Kawahara & Shinohara (2007). Both the data and analyses presented in this version supersede those in Kawahara & Shinohara (2007).

Wilson 2006; Zuraw 2007). For example, several researchers have observed that nasals are more likely to undergo place assimilation than oral consonants (Cho 1990, Mohanan 1993, Boersma 1998, Jun 2004). Steriade (1994, 2001a), Boersma (1998), and Jun (2004) have proposed that this asymmetry arises because a place contrast is harder to perceive in nasals than in oral consonants, and therefore speakers can tolerate a change of articulation in nasal consonants but not in oral consonants: nasal pairs differing in [place] are 'psychoacoustically similar enough', whereas pairs of oral consonants differing in [place] are 'psychoacoustically too different' (see also Malécot 1956, Mohr & Wang 1968, Pols 1983, Hura, Lindblom & Diehl 1992, Kurowski & Blumstein 1993, Ohala & Ohala 1993, Davis & MacNeilage 2000 for the low perceptibility of [place] in nasals; see Kohler 1990, Hura et al. 1992, Huang 2001, Johnson 2003, Kawahara 2006 for the notion of 'perceptually tolerated articulatory simplification').

Another example of the role of psychoacoustic similarity in phonology comes from the observation that a voicing contrast is more likely to be neutralized than other manner contrasts. Steriade (2001b) points out that underlying voiced obstruents readily undergo devoicing in coda positions in many languages such as German, Polish, Russian and Turkish, but coda voiced obstruents never become nasalized in any languages. To explain this observation, drawing on previous psycholinguistic studies (Peters 1963, Greenberg & Jenkins 1964, Walden & Montgomery 1975, van den Broecke 1976), she proposes that a voicing contrast is less perceptible than other manner contrasts, and that speakers tolerate a change in voicing because the change is psychoacoustically the least salient. These proposals have led to – and have expanded on – the general idea of the P-map (for 'Perceptual'-map) (Steriade 2001a, b, 2003), which states that speakers know the perceptual distances between two phonological elements, and that based on this knowledge, they attempt to minimize the perceptual disparity between two corresponding elements in phonology (most commonly inputs and outputs).[2]

This paper contributes to the body of literature which argues for the importance of psychoacoustic similarity, but from a slightly different angle, namely from the perspective of verbal art. A sizable amount of work has shown that psychoacoustic similarity plays an important role in shaping phonological patterns, including verbal art patterns. Previous studies have revealed that speakers require psychoacoustic similarity between two

[2] The principle of minimization of perceptual disparity has been proposed in other correspondence relations: between base and derived words (Steriade 2000, Zuraw 2007), between base and reduplicant (Fleischhacker 2005), between source and loanwords (Steriade 2001a, Kang 2003, Kenstowicz 2003; cf. LaCharité & Paradis 2005), and between rhyming or alliterating elements (see below).

rhyming or alliterating segments, and this claim has been demonstrated in Middle English alliteration (Minkova 2003), English imperfect puns (Fleischhacker 2005), Japanese rap rhymes (Kawahara 2007), and Romanian half-rhymes (Steriade 2003). Building on this body of work, our corpus-based analysis demonstrates that Japanese speakers require two corresponding elements in puns to be as similar as possible, and the measure of similarity rests on psychoacoustic information. Our results show that both phonological patterns and verbal art patterns are governed by the same principle: the maximization of psychoacoustic similarity between two corresponding elements.

To make the case that speakers use psychoacoustic similarity in composing verbal art patterns, this paper analyzes *dajare*, which is a linguistic word game – more specifically, the creation of imperfect puns – in Japanese. In constructing *dajare*, speakers make a sentence out of two words or phrases that are identical or near-identical. An illustrative example is *huton-ga huttonda* 'the futon was blown away', which involves the juxtaposition of two similar words: *huton* and *huttonda*. An analogous English example would be something like 'Titanic in panic'. In many languages, similar sounds tend to form sound pairs in a range of verbal art patterns, including rhymes, alliterations, and imperfect puns (see Kawahara 2007 for a recent overview). The role of similarity in imperfect puns has been explored in English (Lagerquist 1980, Zwicky & Zwicky 1986, Sobkowiak 1991, Hempelmann 2003, Fleischhacker 2005) as well as in Japanese (Otake & Cutler 2001, Cutler & Otake 2002, Shinohara 2004). All of these studies agree that *some* kind of similarity plays a role in the formation of imperfect puns. We argue that it is specifically the PSYCHOACOUSTIC SIMILARITY that underlies imperfect pun parings.

Before closing this introductory section, a remark on our methodology is in order. In addition to lending further support to the role of psychoacoustic similarity in verbal art, our methodology is also innovative; that is, we undertake a statistical approach to the analysis of Japanese imperfect puns. The statistical approach allows us to uncover phonological patterns that could easily be obscured by non-phonological factors and therefore missed in an impressionistic survey. That is, speakers' focus on syntactic coherence and semantic humor may well obscure the purely phonetic and phonological aspects of two corresponding words (Hempelmann 2003). Here we draw on a large amount of data to verify whether discovered phonetic and phonological patterns are convincingly real.

The rest of this paper proceeds as follows. In section 2 we provide an overview of Japanese imperfect puns, and describe how we collected and analyzed the data. In section 3, as a preliminary, we statistically establish the role of similarity in the formation of Japanese imperfect puns. In section 4 we make the case that speakers deploy psychoacoustic similarity, rather than featural similarity, when they compose imperfect puns.

2. Method

2.1 *An overview of Japanese imperfect puns*

Puns are very common in Japanese, and some speakers create new puns on a daily basis (Mukai 1996, Takizawa 1996). In composing puns or *dajare*, speakers deliberately juxtapose a pair of 'similar-sounding' words or phrases – we argue in section 4 that this notion of 'similarity' is based on psychoacoustic information. The two corresponding parts can include identical sound sequences as in *buta-ga butareta* 'a pig was hit', an example of a PERFECT PUN; or the two corresponding words can include similar but non-identical sounds. In (1) we provide an example of an IMPERFECT PUN, which includes corresponding pairs of non-identical sounds (indicated in bold letters: [m–t], [h–b]). In this paper we focus on imperfect puns that include mismatched consonant pairs.[3]

(1) **m**an**h**attan-de su**t**an**b**attan
Manhattan-LOC stand-by
'I stood by in Manhattan.'

In (1), *manhattan* and *sutanbattan* are a pair of similar phrases that form an imperfect pun.[4] This paper uses the following notations. We mark CORRESPONDING DOMAINS in imperfect puns by underlining: we define corresponding domains as sequences of syllables in which corresponding vowels

[3] Other types of imperfect puns involve metathesis (e.g. *shi**mane**-no shi**nema*** 'Cinema in Shimane'), phrase boundary mismatches (e.g. *wakusee waa, kusee* 'Man, that planet stinks'), syllable intrusion (e.g. *bundoki bundo**tto**ki* 'Just take the protractor away from him'), etc. Analyzing these patterns is beyond the scope of this paper, but it is an interesting topic for future research.

[4] Speakers sometimes change the underlying shape of one form to achieve better resemblance to the corresponding form, as in (i) (Takizawa 1996, Shinohara 2004):

(i) Hokkaidoo-wa dekkai **d**oo
Hokkaido-TOP big PARTICLE
'Hokkaido is big.'

In this example, the sentence-final particle [doo] is pronounced as [zo] in canonical speech, but speakers pronounce /zo/ as [doo] here to make the word *dekkaizo* sound similar to *Hokkaidoo*. We also analyzed examples of puns like (i), but found that these types of puns are mainly created by mimicking non-standard speech styles. For example, for the change from /zo/ to [do] in (i), speakers are probably mimicking dialects of Japanese where the particle *zo* is pronounced as [do]. Another common pattern is palatalization, as in (ii):

(ii) Undooʒoo-o kaʃite kudasai. Un dooʒo.
playground-ACC lend please yes please.
'Please let us use your playground. Yes, please.'

The second [dooʒo] is usually pronounced as [doozo]; here /z/ is palatalized to mimic *undooʒoo*. This sort of palatalization is common in Japanese child speech and motherese (Mester & Itô 1989), and composers of puns may be mimicking these styles of speech. Our database thus excludes examples like (i) and (ii) which involve some change from underlying form to surface form.

are identical.[5] We use bold letters to indicate non-identical pairs of corresponding consonants, which is the focus of our analysis.

The pun in (1) is an imperfect pun in that it does not involve perfectly identical sequences of sounds. Though Japanese puns can involve non-identical pairs of consonants, previous studies have noted that imperfect pun pairs tend to involve 'similar' consonants (Otake & Cutler 2001, Cutler & Otake 2002, Shinohara 2004). Deferring detailed discussion of the precise definition of 'similarity', the importance of similarity in imperfect puns can be informally illustrated by examples like: *Ai**z**usan-no ai**s**u* 'ice cream from Aizu' ([z] and [s] are minimally different in voicing) and *okosa**m**a-o okosa-**n**aide* 'please do not wake up the child' ([m] and [n] are minimally different in place). Drawing on this observation, this paper statistically assesses what *kind* of similarity plays a role in imperfect puns, and demonstrates that speakers deploy their knowledge about psychoacoustic similarity based on detailed acoustic information. Our findings show that speakers possess a rich knowledge of psychoacoustic similarity and make use of that knowledge to form verbal art patterns (Steriade 2001b, 2003; Minkova 2003; Fleischhacker 2005; Kawahara 2007).

2.2 *Data collection*

This subsection describes how we compiled the database for this study. In order to investigate how similarity influences the formation of Japanese imperfect puns, we collected examples of imperfect puns from three websites.[6] We also elicited data from native Japanese speakers by asking them to create puns of their own 'out of the blue', from which we selected imperfect puns.[7] In total, we collected 2371 examples of Japanese imperfect puns.

Recall that we define 'corresponding domains' in imperfect puns as sequences of syllables in which corresponding vowels are identical. For

[5] In an example like *Ha**i**degaa-no zense-wa ha**e** dekka?* 'Was Heidegger a fly in his previous life?', the corresponding domain could be interpreted as coinciding with word boundaries; note that this example involves a pair of non-identical vowels ([i–e]). However, corresponding domains and word boundaries do not necessarily coincide with each other, as in *huton-ga huttonda* 'a futon was blown away'. In this study, we therefore define corresponding domains based on sequences of matching vowels rather than on word boundaries. Since examples like *Haidegaa-no zense-wa hae dekka?* were relatively rare in our corpus, our results below should not depend on the particular definition of the corresponding domain we have chosen.

[6] http://www.ipc-tokai.or.jp/~y-kamiya/Dajare/, http://www.geocities.co.jp/Milkyway-Vega/8361/umamoji3.html, http://www.dajarenavi.net/. Data were obtained from these websites in March 2007.

[7] The data collection was primarily done by the second author. She plays with *dajare* in her daily life, and contributed some examples to our database. However, she was unaware of the purpose of this project at the time of data collection.

example, in *okosama-o okosanaide* 'please do not wake up the child', the corresponding domains include the first four syllables. In the domains defined as such, we counted pairs of corresponding consonants between the two words. We ignored identical pairs of consonants, because we are interested in similarity, not identity. Therefore, in the example *okosama-o okosanaide*, for instance, we extracted just the pair [m–n].

A few remarks on our coding convention are in order. We counted only onset consonants, and ignored coda consonants because they assimilate in [place] to the following consonant in Japanese (Itô 1986). We also ignored singleton–geminate distinctions because our interest lies in segmental similarity. We coded the data based on surface forms, following a recent body of work which argues that similarity is based on surface forms rather than phonemic forms (Steriade 2001b; Kang 2003; Kenstowicz 2003; Minkova 2003; Kawahara 2006, 2007; Kawahara, Ono & Sudo 2006; though see Jakobson 1960; Kiparsky 1970, 1972, 1981; Malone 1987 for a different view). For instance, the onset of the syllable [ʃi] is arguably derived from an underlying /s/, but it was counted as [ʃ]. Finally, we excluded pairs which are underlyingly different but are neutralized at the surface because of independent phonological processes (even when such neutralization is optional). For example, the second consonant [dd] in *beddo* 'bed' can optionally devoice in the presence of another voiced obstruent (Nishimura 2003, Kawahara 2006). Thus given an example like *beddo-dai-wa betto itadakimasu* 'please pay for the bed separately', we did not consider the [dd–tt] pair as a pair of non-identical consonants, since the pun-composer could have pronounced /beddo/ as [betto] when s/he composed the pun. Similarly, many puns included examples of [y] corresponding with Ø in the context of [i _ a], as in *kanariØa-o kauno-wa kanari iya* 'I don't like having a canary'. We did not count these cases since /ia/ can surface as [iya] (Katayama 1998, Kawahara 2003).

2.3 *Data analysis: O/E ratios*

This paper investigates what kind of measure of similarity speakers use when they compose imperfect puns. For this purpose, we need a measure of combinability between two segments. We use O(bserved)/E(xpected) ratios as a measure of how well two elements combine. O/E ratios are ratios between how often a pair is actually observed (O-values) and how often it would be expected to occur if two elements were combined at random (E-values). Mathematically, the O-value of a sound [A] is how many times [A] occurs in the corpus, and the E-value of a pair [A–B] is $P(A) \times P(B) \times N$ (where $P(X)$ = the probability of the sound [X] occurring in the corpus; N = the total number of consonants). O/E ratios greater than 1 indicate that the given consonant pairs are combined as pun pairs more frequently than expected (overrepresentation), while O/E ratios less than 1 indicate that the given consonant pairs are combined less frequently than expected (underrepresentation) (see Trubetzkoy 1969, Frisch, Pierrehumbert & Broe 2004, Kawahara 2007 for the notion of relativized frequencies).

In our O/E analysis, we excluded consonants whose O-values are less than 20 because including them resulted in extraordinarily high O/E ratios: combining two rare consonants yields a very low E-value, so that even a single observed pair of that type would yield an artificially high O/E ratio (e.g. the O/E ratio of the [p^j–b^j] pair is 1121.7). For this reason, we excluded [ts], [dʒ], [y], [n^j], [r^j], and all non-coronal palatalized consonants. The exclusions left 535 pairs of consonants for the subsequent analysis.

Before completing the method section, a remark on our statistical analysis is in order. We did not test whether each individual O/E ratio was significantly larger than 1, for two reasons. First, applying the same test on multiple O/E ratios requires a drastic post-hoc α-level adjustment, and it is unlikely to find statistically significant overrepresentations – i.e. potential inflation of Type 2 errors (Klockars & Sax 1986, Myers & Well 2003: 84–86, 241–245). Second, since each O/E ratio is dependent on other O/E ratios, applying the same test on multiple O/E ratios violates the assumption of independence, which usually results in an increase in Type 1 errors (Seco, Menéndez de la Fuente & Escudero 2001). We thus instead applied a statistical test for general patterning of O/E ratios to test specific null hypotheses.

3. The correlation between similarity and combinability: a statistical preliminary

Table 1 shows the result of the O/E analysis, where shading indicates underrepresentation.

To statistically verify the correlation between combinability in puns on the one hand and similarity on the other, as a first approximation, we estimated the similarity of consonant pairs in terms of how many feature specifications they share (Klatt 1968, Fay & Cutler 1977, Shattuck-Hufnagel & Klatt 1977, Stemberger 1991, Bailey & Hahn 2005). We will later discuss the importance of psychoacoustic similarity in section 4 and argue that featural similarity ultimately fails to account for the detailed aspects of the pun pairing patterns. We nevertheless start with this simple version involving featural similarity in order to first statistically establish the correlation between similarity and combinability.

To estimate the numbers of shared feature specifications among the set of distinctive consonants in Japanese, we used eight features: [sonorant], [consonantal], [continuant], [nasal], [strident], [voice], [palatalized], and [place]. [Palatalized] is a secondary place feature which distinguishes plain consonants from palatalized consonants (*yoo-on*, in the traditional Japanese terminology). According to this featural similarity system, a [p–b] pair, for

	Ø	p	b	ϕ	m	w	t	tʃ	s	ʃ	z	d	j	n	r	k	g	h
Ø	0	.7	.24	.77	2.5	6.3	.64	.45	.36	.37	0	0	0	1.5	4.6	1.4	.46	3.7
p		0	8.5	5.6	0	0	1.1	0	1.1	0	0	0	0	0	.37	1.1	.24	.77
b			0	1.6	4.7	5.2	.30	0	0	0	.42	1.1	0	.35	0	.25	.65	2.9
ϕ				0	2.3	0	0	0	.81	.84	0	0	0	0	0	2.7	0	4.2
m					0	0	.64	0	.53	0	0	0	0	8.9	2.1	0	.34	1.7
w						0	0	0	0	0	0	1.2	2.1	1.5	0	.53	0	4.5
t							0	.57	.90	0	0	7.6	0	.62	.44	.87	.14	.94
tʃ								0	.95	7.9	0	0	13	0	.47	.23	0	.49
s									0	4.7	11	.2	0	1.6	.37	.73	0	.78
ʃ										0	0	0	6.8	.54	0	.38	0	.41
z											0	1.7	1.2	0	.62	0	0	0
d												0	.39	1.1	4.0	0	.39	.42
j													0	2.0	.72	0	.93	0
n														0	4.1	.25	1.7	1.1
r															0	.36	1.2	.38
k																0	8.0	1.7
g																	0	0
h																		0

Table 1
O/E ratio (shading indicates underrepresentation).

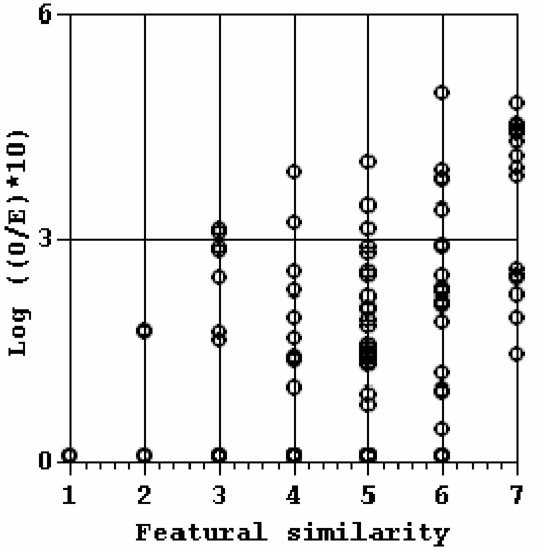

Figure 1
Correlation between the number of shared feature specifications and the likelihood of two consonants being paired in imperfect puns.

example, shares seven feature specifications, and hence is treated as highly similar. On the other hand, a pair like [ʃ–m] agrees only in [cons], and is thus considered as a highly non-similar pair.

We made the following assumptions when assigning feature specifications to Japanese sounds. Affricates are specified both as [+cont] and [−cont] (Sagey 1986, Lombardi 1990). We treated [h] as a voiceless fricative, not as an approximant (Lass 1976: 64–68, Jaeger & Ohala 1984, Sagey 1986, Parker 2002). We also assumed that sonorants are specified as [+voice] and thus agree with voiced obstruents in terms of [voice], although phonologically, voicing in sonorants behaves differently from voicing in obstruents in Japanese phonology (see section 4.3 for further discussion). The appendix gives the complete feature matrix used in this analysis.

The scatter plot in figure 1 illustrates the general correlation between similarity and combinability of consonant pairs. For each consonant pair, it plots the number of shared feature specifications on the x-axis and plots the natural log-transformed O/E ratios multiplied by 10 ($Log_e((O/E)*10)$) on the y-axis. We use log-transformed values on the y-axis because it allows us to fit all the data points in a reasonably small-sized graph. Log_e(zero) is undefined, so we replaced Log_e(zero) with zero. However, the log of O/E ratios smaller than 1 is negative, and hence would be incorrectly treated as a value smaller than zero (e.g. $0.5 > 0$, but $Log_e(0.5) < 0$). Therefore, we multiplied O/E ratios by 10 before log-transformation. The plot excludes all pairs in which one member is Ø (null), since it is difficult if not impossible to define Ø in terms of distinctive features; we discuss such pairs in section 4.4.

We observe in figure 1 an upward trend in which the combinability correlates with featural similarity: the more feature specifications a pair shares, the higher the corresponding O/E ratio is. To statistically assess the linear relationship in figure 1, we calculated a Spearman's rank correlation coefficient ρ, a non-parametric measure of linear correlation. We used a non-parametric measure because we cannot assume a bivariate normal distribution. The result is that ρ is .497 ($t(134)=6.63$; $p<.001$), a statistically significant correlation.

Another measure of featural similarity can be calculated based on the number of shared natural classes divided by the number of shared and unshared natural classes $\left(\frac{\textit{shared natural classes}}{\textit{shared natural classses} + \textit{unshared natural classes}}\right.$; Frisch et al. 2004). Based on this formula, we calculated the similarity matrix among the Japanese consonants using a similarity calculator developed by Adam Albright (Albright n.d.). We then calculated the correlation between combinability and this similarity matrix, yielding a ρ value of .488 ($t(134)=6.47$, $p<.001$), a value similar to the ρ value in the analysis above.

These two analyses establish statistically that a general correlation exists between combinability and some kind of similarity. Having established the statistical correlation, we now turn to some inadequacies of featural similarity in capturing imperfect pun patterns and argue for the importance of psychoacoustic similarity.

4. Beyond featural similarity: the role of psychoacoustic similarity

In section 3, we used distinctive features to estimate similarity. We argue in this section that, although featural similarity accounts for some broad patterns, it ultimately fails to capture some of the detailed aspects of pun pairings. We propose that Japanese speakers instead use psychoacoustic similarity. By psychoacoustic similarity, we mean perceived similarity between sounds based on acoustic details.

One example that supports the role of psychoacoustic similarity is the [r–d] pair, whose O/E ratio is 3.99. According to the distinctive feature system, this pair agrees in six features (all but [son] and [cont]); but the average O/E ratio of other consonant pairs agreeing in six distinctive feature is 1.49. The 95% confidence interval of the O/E ratios of such pairs is 0.39 ~ 2.59, and 3.99 lies outside this interval. Thus the high O/E ratio of the [r–d] pair is unexpected from the standpoint of featural similarity, but makes sense from a psychoacoustic perspective. Japanese [r] is a flap which involves a ballistic – or instantaneous – constriction (Nakamura 2002): [r] and [d] are similar to each other in that they are both voiced consonants with relatively short closures (Price 1981, Steriade 2000).[8]

Below we present four more kinds of arguments that pun composers exploit psychoacoustic similarity rather than featural similarity. First, speakers are sensitive to the different perceptual salience of the same feature in different contexts (section 4.1.). Second, composers take into consideration the different perceptual salience of different features (section 4.2.). Third, composers are sensitive to similarity contributed by a phonologically inert feature (section 4.3.). Finally, pun-makers are willing to pair Ø with consonants that are psychoacoustically similar to Ø (section 4.4.).

4.1 *Sensitivity to context-dependent salience of the same feature*

The first piece of evidence that pun composers make use of psychoacoustic similarity is the fact that composers are sensitive to different degrees of salience of the same feature in different contexts. Here, we focus on the perceptual salience of [place]. In (2), we list the O/E ratios of minimal pairs of consonants differing in place (the Japanese moraic nasal [N] appears only in codas, and was therefore excluded from this study).

(2) *O/E ratios of minimal pairs differing in place*

[m–n]: 8.85	[b–d]: 1.09	[p–t]: 1.11
	[b–g]: 0.65	[p–k]: 1.08
	[d–g]: 0.39	[t–k]: 0.87

We observe that the O/E ratio of the [m–n] pair is higher than that of any other minimal pair in (2) (by a binomial test, $p < .05$).[9] Thus the data in (2) indicate that composers treat the [m–n] pair as more similar than any minimal pairs of oral consonants: they treat the place distinction in nasal consonants as less salient than in oral consonants.

Evidence from previous phonetic and psycholinguistic studies supports the lower perceptibility of [place] in nasals than in oral consonants. First, a similarity judgment experiment by Mohr and Wang (1968) shows that nasal

[8] Other examples of this kind are the pairs [tʃ–kʲ] and [dʒ–gʲ], which are psychoacoustically – but not featurally – similar (Blumstein 1986, Ohala 1989, Guion 1998, Wilson 2006). A long burst of velar consonants resembles affrication, and raising of F2 in [kʲ] and [gʲ] due to palatalization makes them sound similar to coronals. Recall, however, that we excluded [kʲ] and [gʲ] from our O/E analysis (section 2.3), because they are rare consonants. Yet our original O/E analysis, which did include rare consonants, showed that both the [tʃ–kʲ] pair and [dʒ–gʲ] are highly overrepresented (O/E = 18.3, 21.6, respectively). To the extent that these high O/E ratios are statistically reliable, these examples provide further support for the importance of psychoacoustic similarity.

[9] Pairs of nasal consonants with different [place] specifications also seem to be common in English rock lyrics (Zwicky 1976), English and German poetry (Maher 1969, 1972), English imperfect puns (Zwicky & Zwicky 1986), and Japanese rap lyrics (Kawahara 2007).

minimal pairs are considered perceptually more similar to each other than oral consonant minimal pairs. Second, Pols (1983) demonstrated with Dutch speakers that [place] in nasals is less accurately perceived than [place] in oral consonants in noisy environments.[10] Place cues are less salient in nasal consonants than in oral consonants for two reasons: (i) formant transitions into and out of the neighboring vowels are obscured by coarticulatory nasalization, and (ii) burst spectra play an important role in distinguishing different places of articulation (Jakobson, Fant & Halle 1952, Stevens & Blumstein 1978, Blumstein & Stevens 1979), but nasals have weak or no bursts (see Malécot 1956, Pols 1983, Hura et al. 1992, Kurowski & Blumstein 1993, Ohala & Ohala 1993, Steriade 1994, Boersma 1998, Davis & MacNeilage 2000, Steriade 2001a, Jun 2004 for further discussion of the low perceptibility of nasals' [place] and its phonological consequences). To summarize, in measuring similarity due to [place], Japanese speakers take into account the lower perceptibility of [place] in nasal consonants. The lower perceptibility of [place] in nasals has a psychoacoustic basis: the blurring of formant transitions caused by coarticulatory nasalization and weak bursts. The data in (2) therefore show that speakers use psychoacoustic similarity in composing puns.

By contrast, if speakers were using featural similarity, then the higher combinability of the nasal pair would remain unexplained. One could postulate that featural similarity is affected less when two nasal segments disagree in [place] than when two oral segments disagree in [place]. However, introducing such weighting is post-hoc: it simply restates the observation and has no explanatory power.

An anonymous reviewer points out that feature weighting would not be ad-hoc 'if it is derived from an independently motivated model of feature relationship'. However, no previously established models of distinctive features distinguish [place] in oral consonants and [place] in nasal consonants. Feature geometry is one well-elaborated theory of feature relationships, which postulates that features are organized in a hierarchical structure (Clements 1985, Sagey 1986, McCarthy 1988, Halle 1992, Clements & Hume 1995). However, no versions of feature geometry postulate a structural difference between [place] in nasal consonants and [place] in oral consonants.

Another elaboration of feature relationships is the theory of underspecification. This theory postulates that non-contrastive or unmarked specifications are underlyingly unspecified (Kiparsky 1982, Itô & Mester 1986, Archangeli 1988, Mester & Itô 1989, Paradis & Prunet 1991, Itô, Mester & Padgett 1995; see also Lahiri & Reetz 2002 for a model of word recognition using underspecification). Underspecification theory cannot explain the

[10] Some studies found that voiced obstruent pairs are more similar to each other than voiceless obstruent pairs are (Greenberg & Jenkins 1964, Mohr & Wang 1968), but the difference is not observed in (2) – possibly because the difference is too small to be reflected in our database.

difference between [place] in nasal consonant and [place] in oral consonants either, because [place] is distinctive both in nasal and in oral consonants, and therefore [place] must be specified in both of these environments.

As a last alternative analysis, it has been claimed that when two sounds alternate with each other in phonology, their phonological relation may enhance similarity between them (Zwicky 1976; Malone 1987, 1988a, b; Hume & Johnson 2003; Shinohara 2004; Kawahara 2007; cf. Steriade 2003, who doubts this claim). However, it is impossible to derive the high O/E ratio of the [m–n] pair from their phonological alternating status. In Japanese phonology, [place] in nasals and [place] in oral consonants do not behave differently. In onset position, neither nasal nor oral consonants change their place specifications. In codas, both nasal and oral consonants place-assimilate to the following consonant (Itô 1986), as exemplified in (3).

(3) (a) *Nasal place assimilation*

ho**N**	'book'
ho**m-b**ako	'book-box'
ho**n-d**ana	'book-shelf'
ho**ŋ-g**a	'book-NOM'

(b) *Place assimilation of oral consonants*

bu**t**(u)	'to hit'	hi**k**(u)	'to pull'
bu**p-p**anasu	'to fire a bullet'	hi**p-p**aru	'to pull off'
bu**t-t**oosu	'to permeate'	hi**t-t**oraeru	'to arrest'
bu**k-k**orosu	'to kill'	hi**k-k**aku	'to scratch'

To summarize, no independent mechanisms of distinctive features derive the difference between [place] in nasals and [place] in oral consonants.

4.2 *Sensitivity to different salience of different features*

The second argument that pun composers deploy psychoacoustic rather than featural similarity rests on the fact that they treat some features as perceptually more salient than other features. A large body of psycholinguistic work has demonstrated the non-equivalence of perceptibility of different features (Miller & Nicely 1955, Peters 1963, Singh & Black 1966, Mohr & Wang 1968, Ahmed & Agrawal 1969, Singh, Woods & Becker 1972, Wang & Bilger 1973, Walden & Montgomery 1975, van den Broecke 1976, Benkí 2003, Bailey & Hahn 2005). Therefore, if speakers are using psychoacoustic similarity, it is predicted that we would observe that different features contribute to pun pairing frequencies to different degrees. We now show that this prediction is correct.

Here we discuss the perceptibility of [voice]. Some scholars have proposed that among manner features that determine spectral continuity ([cont, nasal, voice]), the [voice] feature contributes only weakly to consonant distinctions. The weaker perceptibility of [voice] obtains support from previous psycholinguistic findings, such as Multi-Dimensional Scaling (Peters 1963, Walden & Montgomery 1975; see also Singh et al. 1972), a similarity judgment task (Greenberg & Jenkins 1964, van den Broecke 1976, Bailey & Hahn 2005: 352), and an identification experiment in noise (Wang & Bilger 1973; see also Singh & Black 1966).[11] See also Kenstowicz (2003) and Steriade (2001b), who argue that [voice] is phonologically more likely to neutralize than other manner features because [voice] is less perceptible.

Given the relatively low perceptibility of [voice], if pun composers are sensitive to psychoacoustic similarity, we expect that pairs of consonants that are minimally different in [voice] should combine more frequently than minimal pairs that disagree in other manner features. To compare the effect of [cont], [nasal], and [voice], we compare the O/E ratios for the minimal pairs defined by these features, as in (4).

(4)	[cont]	[nasal]	[voice]
	[p–Φ]: 5.58	[b–m]: 4.68	[p–b]: 8.51
	[t–s]: 0.90	[d–n]: 1.12	[t–d]: 7.64
	[d–z]: 1.68		[k–g]: 8.03
			[s–z]: 11.3
			[ʃ–ʒ]: 6.81

A non-parametric Mann-Whitney test reveals that the O/E ratios of the minimal pairs that differ in [voice] are significantly larger than those of the minimal pairs that differ in [cont] or [nasal] (*Wilcoxon* $W = 15$, $z = 2.61$, $p < .01$).[12]

Japanese pun-makers treat minimal pairs that differ only in [voice] as very similar, which indicates that composers know that a disagreement in [voice] does not disrupt perceptual similarity as much as other manner features would. In other words, Japanese speakers have an awareness of the varying perceptual salience of different features.[13]

[11] A voicing contrast is more robust than other features in white noise (Miller & Nicely 1955). However, its robustness is probably due to the fact that voicing contrasts involve durational cues (such as VOT, preceding vowel duration, closure duration, and closure voicing duration; Lisker 1986), and white noise does not cover such durational cues. In fact, in signal-dependent noise, the difference in perceptibility between voicing and other manner features is highly attenuated, if not completely obliterated (Benkí 2003).

[12] Previous studies have observed that voicing disagreement is common in Japanese imperfect puns (Otake & Cutler 2001, Shinohara 2004); our study provides statistical support for this observation.

[13] The perceived similarity between minimal pairs of consonants differing in [voice] may be enhanced in Japanese because they are distinguished in Japanese orthography only by a diacritic, i.e. they are orthographically similar (see Itô, Kitagawa & Mester 1996 for the influence of orthography on verbal art). However, [voice]'s weak contribution to perceptual similarity has been demonstrated in languages other than Japanese (see the references cited above), showing that orthographic similarity is not the only factor. Moreover, voicing disagreement is also more common than nasality disagreement in other languages, including English imperfect puns (Lagerquist 1980, Zwicky & Zwicky 1986), slant rhyme in the

A theory that relies on featural similarity cannot explain why a [voice] difference does not reduce pun pairing likelihood as much as differences in other manner features do. To salvage the featural similarity view, one could augment the theory by postulating that [voice] contributes less to featural similarity; but this analysis is ad-hoc. This augmentation simply restates the observation and misses the insight that [voice]'s lesser contribution to the combinability of pun pairs has its basis in its lower perceptibility. Relatedly, Frisch et al. (2004: 204) consider augmenting their model of similarity by introducing weighting. However, weighting does not automatically follow from their model, but is rather 'perceptually and cognitively plausible' (p. 204) – Frisch et al. too, assume that weighting has a psychoacoustic basis.[14]

There are no independent mechanisms in feature theories that derive the observation that [voice] contributes to featural similarity less than other manner features. One might attempt to derive the low contribution of [voice] from its hierarchical position in feature geometry. McCarthy (1988) places [voice] under a Laryngeal node and places [nasal] and [cont] directly under a root node. One could argue that the fact that [voice] is dominated by a higher node – other than a root node – may make the contribution of [voice] smaller than that of [nasal] and [cont] (Golston 1994, cited in Holtman 1996: 252). However, it is not clear why being dominated by some higher node should result in a lower contribution to featural similarity. Nor, to our knowledge, have the proponents of feature geometry proposed to derive different degrees of contribution to featural similarity from hierarchical organization. Moreover, in other models of feature geometry, [nasal] and [cont] are dominated by some higher node as well; both of these features are dominated by a Manner node in Clements (1985); [nasal] is dominated by a Soft Palate and Supralaryngeal node in Sagey (1986) and Halle (1992) (although Sagey notes that the evidence for the Supralaryngeal node is not so strong); [cont] is dominated by an Oral Cavity node in Clements and Hume (1995) and by a Place node in Padgett (1991). Underspecification is of no help either, because all the features in (4) are contrastive in Japanese phonology.

Nor can [voice]'s weak contribution to featural similarity be derived from an alternating status of minimal pairs that differ in voicing. We do find a change in [voice] in Japanese phonology, namely Rendaku, which voices the initial consonant of the second member of compounds (Itô & Mester

poetry of Robert Pinsky (Hanson 1999, cited in Steriade 2001b), and many half rhyme patterns (English: Zwicky 1976, Romanian: Steriade 2003, and Slavic: Eekman 1974). This observation therefore shows that even without orthographic similarity, speakers know that a [voice] difference contributes to perceptual similarity less than other features.

[14] Frisch et al.'s model alone does not predict that a minimal pair of sounds differing in voicing would be most similar to each other. For example, the [p–b] pair has a similarity value of 0.56 while the [p–Φ] pair has a value of 0.67; similarly, the [t–d] pair has a value of 0.54 while the [z–d] pair has a value of 0.57.

	Voiced obstruent	Voiceless obstruent
Paired with sonorants	63 (18.2%)	30 (6.0%)
Total occurrences	346	497

Table 2
Probabilities of sonorants being paired with voiced obstruents and voiceless obstruents.

1986), as in (5a) (see also Nishimura 2003, Kawahara 2006 for another phonological alternation that changes consonantal voicing). However, [Φ] and [p] also alternate allophonically with each other in Japanese phonology, where [Φ] appears before [u] but [p] appears when geminated (Itô & Mester 1999), as in (5b). Moreover, voiced obstruents, when geminated, are realized as nasal–obstruent clusters, as in (5c) (Kuroda 1965; Itô & Mester 1986, 1999; Kawahara 2006).

(5) (a) *Rendaku*
tako 'octopus' oo-**d**ako 'big octopus'
kai 'shell' hotate-**g**ai 'scallop'
(b) *[ϕ]~[p] alternation*
Φuro 'bath' hito**p-p**uro 'one bath'
Φun 'minute' ro**p-p**un 'six minutes'
(c) *Coda nasalization*
/za**b**u+μ+ri/ → [za**mb**uri] 'splashing'
(cf. /ta**p**u+μ+ri/→ [ta**pp**uri] 'a lot')
/o+μ+**d**eru/→ [o**nd**eru] 'get out'
(cf. /o+μ+**k**akeru/→o**kk**akeru 'chase')
('μ' represents a floating mora that causes gemination)

Since Japanese phonology shows evidence for changes in all of [voice], [cont] and [nasal], it is difficult to derive the asymmetry in (4) from alternating statuses in phonology.

4.3 *Sensitivity to similarity contributed by voicing in sonorants*

The third piece of evidence that speakers resort to psychoacoustic similarity is that speakers are sensitive to similarity contributed by a phonologically inert feature – voicing in sonorants. Table 2 shows how often Japanese speakers combine sonorants with voiced obstruents and voiceless obstruents in imperfect puns in our database.

Out of 346 tokens of voiced obstruents, 63 of them are paired with sonorants, whereas out of 497 tokens of voiceless obstruents, only 30 are paired with sonorants. In other words, the probability of voiced obstruents corresponding with sonorants (0.18; s.e. (standard error) = 0.02) is higher

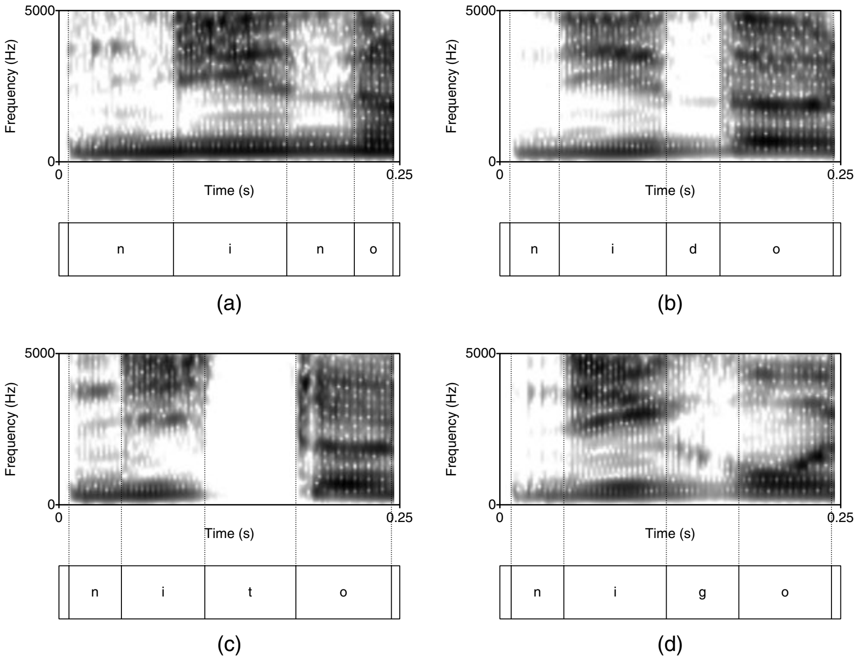

Figure 2
Illustrative spectrograms of Japanese [n, d, t, g] in the environment of [ni _ o]. Read by a female native speaker of Japanese. Pictures generated by Praat (Boersma & Weenink 2007).

than the probability of voiceless obstruents corresponding with sonorants (0.06; s.e. = 0.01). The difference between these ratios is statistically significant (by approximation to a Gaussian distribution, $z = 5.22$, $p < .001$). The pattern in table 2 suggests that pun composers treat sonorants as being more similar to voiced obstruents than to voiceless obstruents (see Stemberger 1991, Walker 2003, Côté 2004, Frisch et al. 2004, Rose & Walker 2004, Coetzee & Pater 2005 for evidence from other languages that obstruent voicing promotes similarity with sonorants; see Kawahara 2007 for the same pattern in Japanese rhymes).

The thesis that speakers use psychoacoustic similarity when forming puns correctly predicts the effect of sonorant voicing on their similarity with voiced obstruents. First, low frequency energy is present during the consonantal constriction in both sonorants and voiced obstruents (figure 2a, b), but not in voiceless obstruents (figure 2c). Second, Japanese voiced stops, especially [g], are often lenited intervocalically, resulting in formant continuity (figure 2d), as with sonorants. Voiceless stops do not spirantize, resulting in complete formant discontinuity, and thus differ from sonorants. For these reasons, voiced obstruents are acoustically more similar to sonorants than voiceless obstruents are.

By contrast, if speakers were deploying featural rather than psychoacoustic similarity, the pattern in table 2 would not be predicted, given the behavior of [+voice] in Japanese sonorants. Phonologically, voicing in Japanese sonorants behaves differently from voicing in obstruents. A well-known phonological restriction in Japanese requires that there be no more than one 'voiced segment' within a stem; but only voiced obstruents, not voiced sonorants, count as 'voiced segments'. Consequently, previous studies have proposed that [+voice] in sonorants is underspecified (Itô & Mester 1986, Itô et al. 1995), or that sonorants do not bear the [voice] feature at all (Mester & Itô 1989), or that sonorants and obstruents bear different [voice] features (Rice 1993). Regardless of how we featurally differentiate voicing in sonorants and voicing in obstruents, sonorants and voiced obstruents do not share the same phonological feature for voicing. Therefore, if speakers were deploying phonological featural similarity, the pattern in table 2 would not be predicted. The featural similarity view – augmented with underspecification or structural differences between sonorant voicing and obstruent voicing – in fact makes the incorrect prediction that voicing in sonorants and voicing in obstruents should not interact with each other.

4.4 *Sensitivity to consonants' similarity to Ø*

As a final piece of evidence for the importance of psychoacoustic similarity, we discuss consonants' similarity to Ø: consonants that correspond with Ø are those that are psychoacoustically similar to Ø. In some imperfect puns, consonants in one phrase do not have a corresponding consonant in the other phrase, as in ***h****ayamatte* ***Ø****ayamatta* 'I apologized without thinking too much' and ***Ø****akagaeru-ga* ***w****akagaeru* 'A brown frog will become young'. In (6) we list the set of consonants whose O/E ratios with Ø are larger than 1 – these are consonants which are paired with Ø more frequently than expected.[15]

(6) [w]: 6.25, [r]: 4.59, [h]: 3.72, [m]: 2.54, [n]: 1.49, [k]: 1.39

The consonants listed in (6) are in fact those that are likely to be psychoacoustically similar to Ø, which supports the claim that speakers use psychoacoustic information when they create imperfect puns.

First, since [w] is a glide, the transition between [w] and the following vowel is blurry, which can make the presence of [w] hard to detect.[16] Myers

[15] Shinohara (2004) also found that [r] and [h] are most likely to correspond with Ø in Japanese imperfect puns, although this observation was based on O-values rather than O/E ratios. See also Lagerquist (1980) for a similar pattern in English (cf. Zwicky & Zwicky 1986).

[16] Since the other palatal glide [y] was very rare in our data (O = 18), it did not enter the O/E analysis.

and Hanssen (2005) demonstrate that given a sequence of two vocoids, listeners misattribute the transitional portion to the second vocoid, effectively lengthening the percept of the second vocoid and shortening the percept of the first vocoid.[17] Thus, due to [w]'s blurry boundaries and consequent misparsing, the presence of [w] is perceptually hard to detect.

Next, Japanese [r] is a flap, which involves a brief and ballistic constriction in Japanese (Nakamura 2002). The shortness of the constriction makes [r] sound similar to Ø. Third, the propensity of [h] to correspond with Ø accords with the observation that [h] lacks a superlaryngeal constriction and hence its spectral property assimilates to that of neighboring vowels (Keating 1988, Pierrehumbert & Talkin 1992), making the presence of [h] difficult to detect; [h] is a perceptually weak consonant (see Mielke 2003 for the perceptibility of [h] in various phonetic contexts).

Fourth, for [m] and [n], we speculate that the edges of these consonants with flanking vowels are blurry due to coarticulatory nasalization, causing them to be interpreted as belonging to the neighboring vowels. Downing (2005) argues that the transitions between vowels and nasals can be misparsed due to this blurriness, and that the misparsing effectively lengthens the percept of the vowel. As a result of misparsing, the perceived duration of nasals may become shortened.

Finally, [k] extensively coarticulates with adjacent vowels in terms of tongue backness (Sussman, McCaffrey & Matthews 1991, Keating & Lahiri 1993, Keating 1996). As a result of this extensive coarticulation, [k] can fade into its environment and become perceptually similar to Ø (de Lacy & Kingston 2006).[18]

To summarize, speakers pair Ø with segments that are psychoacoustically similar to Ø – especially those that fade into their environments – but not with consonants whose presence is highly perceptible. Stridents, for example, are not likely to correspond with Ø because of their salient long duration and great intensity of the noise spectra (Wright 1996, Steriade 2001a). Additionally, coronal stops coarticulate least with surrounding vowels (Sussman et al. 1991, de Lacy & Kingston 2006) and hence they stand out perceptually from their environments. As a result, they are unlikely to correspond with Ø.

By contrast, phonological similarity does not offer a straightforward explanation for the set of consonants in (6). The list of consonants in (6) does not alternate systematically with Ø in Japanese phonology. We observe that

[17] The perceived shortening of the first vocoid may be exaggerated by durational contrast, whereby the percept of one interval gets shortened next to long intervals (Kluender, Diehl & Wright 1988, Diehl & Walsh 1989). However, we also acknowledge that durational contrast has not been replicated by some later studies (Fowler 1992, van Dommelen 1999).

[18] Given this explanation, one may wonder why [g] does not behave like [k]. We conjecture that since Japanese [g] is spirantized intervocalically (Kawahara 2006; see figure 2d above), it stands out from its environment with its frication noise.

(6) includes sonorant consonants, with the exception of [h] and [k];[19] but as Kirchner (1998: 17–18) notes, there is no sense in which sonorous consonants are similar to Ø phonologically. We should in fact not consider Ø as being sonorous: while languages prefer to have sonorous segments in syllable nuclei (Dell & Elmedlaoui 1985, Prince & Smolensky 1993/2004), no languages prefer to have Ø nuclei. Sonority thus fails to explain why speakers prefer to pair Ø with the set of consonants in (6).

Rather, the list of consonants in (6) includes sonorants because their phonetic properties make them akin to Ø: their edges with flanking vowels are blurry, fading into their environments. This view is supported by the fact that [k] and [h], which like sonorants blend into their environment, are treated as similar to Ø.

As another alternative, one could try to characterize Ø in terms of distinctive features, making use of the theory of underspecification. Recall that this theory postulates that non-contrastive or unmarked feature specifications are underlyingly unspecified (Kiparsky 1982, Itô & Mester 1986, Archangeli 1988, Mester & Itô 1989, Paradis & Prunet 1991, Itô et al. 1995). Based on this view, we could postulate that the segment which has the sparsest underlying specification is closest to Ø. This analysis, however, does not predict that the list of consonants in (6) would be close to Ø. For example, nasal consonants are marked while oral consonants are not, and sonorants are marked while obstruents are not (Chomsky & Halle 1968); therefore, [−nasal] and [−son] should be underspecified. As a result, oral obstruents are predicted to be closest to Ø because they lack underlying specifications for [±nasal] and [±son]. However, the list in (6) includes nasal and sonorant consonants. Moreover, no version of underspecification postulates that [k] is more sparsely specified than [t]: in fact, the proponents of Radical Underspecification argue for the opposite (Archangeli 1988, Paradis & Prunet 1991, Lahiri & Reetz 2002).[20] For these reasons, the theory of underspecification does not account for why the consonants in (6) are treated as close to Ø in Japanese imperfect puns.

4.5 *Summary: psychoacoustic similarity vs. featural similarity*

Japanese speakers perceive similarity between sounds based on acoustic information, and use that knowledge of psychoacoustic similarity to compose imperfect puns. Featural similarity fails to capture the bases of similarity

[19] Some proposals treat [h] as a (voiceless) sonorant (Chomsky & Halle 1968), but the [−son] status of [h] is supported by debuccalization processes (/s/ → [h]: Lass 1976, Sagey 1986) as well as by a psycholinguistic experiment (Jaeger & Ohala 1984) and an acoustic experiment (Parker 2002).

[20] Spring (1994) argues that in Axininca Campa, a velar glide is underlyingly underspecified, while [t] is not. However, Spring does not argue for general velar underspecification, and hence even under this theory [k] is no more underspecified than [t].

judgments that speakers make in composing pun pairs. We have presented four kinds of arguments: (i) context-dependent salience of the same feature, (ii) relative salience of different features, (iii) similarity contributed by a phonologically inert feature, and (iv) similarity between consonants and Ø. We have also argued throughout that these shortcomings of the featural similarity model cannot be salvaged by existing elaboration of feature theories, such as feature geometry or underspecification theory. Recall also that there are pairs like the [r–d] pair, whose high degree of over-representation cannot be explained in terms of featural similarity (see also footnote 8 for other potential examples of this kind).

Rather, Japanese speakers use psychoacoustic similarity when they make imperfect puns. A future investigation should construct a complete psycho-acoustic similarity matrix for Japanese sounds through psycholinguistic experiments (e.g. identification experiments in noise and similarity judgment tasks) and apply it to a further analysis of the imperfect pun patterns. Since this task will require a completely new set of experiments, we leave it for future research.

5. Conclusion

5.1 *Parallels between verbal art and phonological patterns*

In closing this paper, we discuss non-trivial parallels which we found between the imperfect pun patterns and phonological patterns. For example, in Japanese imperfect puns, [m] and [n] are more likely to correspond with each other than [p] and [t] are, just as in other languages' phonology in which [m] and [n] alternate with each other (i.e. nasal place assimilation) while [p] and [t] do not (Cho 1990, Mohanan 1993, Jun 2004). In both verbal art and phonology, nasal pairs are more likely to correspond with each other than oral consonant pairs.

Similarly, in the pun pairing patterns, minimal pairs that differ in voicing are more likely to correspond with each other than other types of minimal pairs that differ in nasality and continuency. This patterning parallels the cross-linguistic observation that a voicing contrast is more easily neutralized than other manner features (Steriade 2001b, Kenstowicz 2003). In both cases, a voicing mismatch between two corresponding elements is tolerated, more so than a mismatch in other features.

Third, in section 4.3 we observed that voicing in sonorants enhances similarity with voiced obstruents, and this finds its parallel in linguistic similarity avoidance patterns. In some languages, pairs of sonorants and voiced obstruents are disfavored, as in Arabic (Frisch et al. 2004: 195) and Muna (Coetzee & Pater 2005). Finally, we demonstrated in section 4.4 that when consonants correspond with Ø, speakers tend to choose those consonants that fade into their environment. Again, this pattern finds a parallel in phonology: when speakers epenthesize a segment, they typically choose segments that are the least intrusive (usually glottal consonants for consonants and schwa or high vowels for vowels: Dupoux et al. 1999, Steriade 2001b, Kenstowicz 2003, Howe & Pulleyblank 2004, Walter 2004, Kwon 2005, though see de Lacy & Kingston 2006 for the observation that [k] is never epenthetic in natural languages, presumably for a markedness reason).

With these results we conclude that there exist non-trivial parallels between verbal art patterns and phonological patterns.[21] As we have argued throughout this paper, the parallels exist because speakers use psychoacoustic similarity in shaping both phonological and verbal art patterns. Given the parallels, moreover, we further conclude that it can be fruitful to investigate speakers' grammatical knowledge by examining para-linguistic patterns (Ohala 1986, Itô et al. 1996, Fabb 1997, Yip 1999, Minkova 2003, Steriade 2003, Kawahara 2007 as well as references cited above). Our strategy has been inspired by a growing interest in using verbal art as a source of information about our linguistic knowledge. We hope our studies have shown the fruitifulness of investigating phonological hypotheses using external evidence (Churma 1979).

5.2 *Overall summary*

The combinability of two consonants in imperfect puns in Japanese correlates positively with their perceptual similarity. The analysis of imperfect puns has shown that speakers possess a rich knowledge of psychoacoustic similarity and deploy it in composing imperfect puns, supporting the tenets of the P-map hypothesis (Steriade 2001a, b, 2003; Fleischhacker 2005; Kawahara 2006, 2007; Zuraw 2007). Featural similarity, on the other hand, fails to capture the bases of the similarity decisions that speakers evidently make.

[21] Recall from section 4 that the effects of psychoacoustic similarity cannot be derived from patterns in Japanese phonology. In other words, the above parallels exist between Japanese imperfect pun patterns and phonological patterns in other languages. Therefore, the parallels arise not because speakers compose imperfect puns based on their knowledge of Japanese phonology, but rather because both Japanese imperfect pun patterns and phonological patterns in other languages are grounded in the same principle – the principle of the minimization of perceptual disparities.

APPENDIX

Feature matrix used for the analysis in section 3

	son	cons	cont	nasal	voice	strid	place	palatalized
p	−	+	−	−	−	−	lab	−
b	−	+	−	−	+	−	lab	−
Φ	−	+	+	−	−	−	lab	−
m	+	+	−	+	+	−	lab	−
w	+	−	+	−	+	−	lab	−
t	−	+	−	−	−	−	cor	−
d	−	+	−	−	+	−	cor	−
s	−	+	+	−	−	+	cor	−
z	−	+	+	−	+	+	cor	−
ʃ	−	+	+	−	−	+	cor	+
ʒ	−	+	+	−	+	+	cor	+
ʧ	−	+	±	−	−	+	cor	+
n	+	+	−	+	+	−	cor	−
r	+	+	+	−	+	−	cor	−
k	−	+	−	−	−	−	dors	−
g	−	+	−	−	+	−	dors	−
h	−	+	+	−	−	−	phary	−

REFERENCES

Ahmed, Rais & S. S. Agrawal. 1969. Significant features in the perception of (Hindi) consonants. *Journal of the Acoustical Society of America* 45, 758–773.

Albright, Adam (n.d.). Segmental similarity calculator [software]. http://web.mit.edu/albright/www/.

Archangeli, Diana. 1988. Aspects of underspecification theory. *Phonology* 5, 183–208.

Bailey, Todd & Ulrike Hahn. 2005. Phoneme similarity and confusability. *Journal of Memory and Language* 52, 339–362.

Benkí, José. 2003. Analysis of English nonsense syllable recognition in noise. *Phonetica* 60, 129–157.

Blumstein, Sheila. 1986. On acoustic invariance in speech. In Joseph Perkell & Dennis Klatt (eds.), *Invariance and variability in speech processes*, 178–193. Hillsdale, NJ: Lawrence Erlbaum.

Blumstein, Sheila & Kenneth Stevens. 1979. Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America* 66, 1001–1017.

Boersma, Paul. 1998. *Functional phonology: Formalizing the interaction between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.

Boersma, Paul & David Weenink. 2007. Praat: Doing phonetics by computer (version 5.0.34). http://www.praat.org.

Cho, Young-mee Yu. 1990. *Parameters of consonantal assimilation*. Ph.D. dissertation, Stanford University.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Churma, Don. 1979. *Arguments from external evidence in phonology*. Ph.D. dissertation, Ohio State University.

Clements, G. N. 1985. The geometry of phonological features. *Phonology Yearbook* 2, 225–252.
Clements, G. N. & Elizabeth Hume. 1995. The internal organization of speech sounds. In John A. Goldsmith (ed.), *The handbook of phonological theory*, 245–306. Malden, MA & Oxford: Blackwell.
Coetzee, Andries & Joe Pater. 2005. Lexically gradient phonotactics in Muna and Optimality Theory. Ms., University of Massachusetts, Amherst.
Côté, Marie-Hélène. 2004. Syntagmatic distinctness in consonant deletion. *Phonology* 21, 1–41.
Cutler, Anne & Takashi Otake. 2002. Rhythmic categories in spoken-word recognition. *Journal of Memory and Language* 46, 296–322.
Davis, Barbara & Peter MacNeilage. 2000. An embodiment perspective on the acquisition of speech perception. *Phonetica* 57, 229–241.
de Lacy, Paul & John Kingston. 2006. Synchronic explanation. Ms., Rutgers University & University of Massachusetts, Amherst.
Dell, François & Mohamed Elmedlaoui. 1985. Syllabic consonants and syllabification in Imdlawn Tashlhiyt Berber. *Journal of African Languages and Linguistics* 7, 105–130.
Diehl, Randy & Margaret Walsh. 1989. An auditory basis for the stimulus-length effect in the perception of stops and glides. *Journal of the Acoustical Society of America* 85, 2154–2164.
Downing, Laura. 2005. On the ambiguous segmental status of nasal in homorganic NC sequences. In Marc van Oostendorp & Jeroen M. van der Weijer (eds.), *The internal organization of phonological segments*, 183–216. Berlin: Mouton de Gruyter.
Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier & Jacques Mehler. 1999. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25, 1568–1578.
Eekman, Thomas. 1974. *The realm of rime: A study of rime in the poetry of the Slavs*. Amsterdam: Adolf M. Hakkert.
Fabb, Nigel. 1997. *Linguistics and literature: Language in the verbal arts of the world*. Oxford: Blackwell.
Fay, David & Anne Cutler. 1977. Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry* 8, 505–520.
Fleischhacker, Heidi. 2005. *Similarity in phonology: Evidence from reduplication and loan adaptation*. Ph.D. dissertation, UCLA.
Fowler, Carol. 1992. Vowel duration and closure duration in voiced and unvoiced stops: There are no contrast effects here. *Journal of Phonetics* 20, 143–165.
Frisch, Stephan, Janet B. Pierrehumbert & Michael Broe. 2004. Similarity avoidance and the OCP. *Natural Language & Linguistic Theory* 22, 179–228.
Golston, Chris. 1994. The phonology of rhyme. Presented at 1994 LSA Meeting.
Greenberg, Joseph & James Jenkins. 1964. Studies in the psychological correlates of the sound system of American English. *Word* 20, 157–177.
Guion, Susan. 1998. The role of perception in the sound change of velar palatalization. *Phonetica* 55, 18–52.
Halle, Morris. 1992. Phonological features. In William Bright (ed.), *International encyclopedia of linguistics*, vol. 3, 207–212. Oxford: Oxford University Press.
Hanson, Kristin. 1999. Formal variation in the rhymes of Robert Pinsky's *The Inferno* of Dante. Ms., University of California, Berkeley.
Hempelmann, Christian. 2003. *Paronomasic puns: Target recoverability towards automatic generation*. Ph.D. dissertation, Purdue University.
Holtman, Astrid. 1996. *A generative theory of rhyme: An optimality approach*. Ph.D. dissertation, Utrecht Institute of Linguistics.
Howe, Darin & Douglas Pulleyblank. 2004. Harmonic scales as faithfulness. *Canadian Journal of Linguistics* 49, 1–49.
Huang, Tsan. 2001. The interplay of perception and phonology in tone 3 sandhi in Chinese Putonghua. In Elizabeth Hume & Keith Johnson (eds.), *Studies on the interplay of speech perception and phonology* (Ohio State University Working Papers in Linguistics 55), 23–42. Columbus, OH: OSU Working Papers in Linguistics.
Huffman, Marie K. & Rena A. Krakow (eds.). 1993. *Nasals, nasalization, and the velum*. New York: Academic Press.

Hume, Elizabeth & Keith Johnson. 2003. The impact of partial phonological contrast on speech perception. *15th International Congress of Phonetic Sciences* (*ICPhS*) *2003, Barcelona*, 2385–2388.

Hura, Susan, Björn Lindblom & Randy Diehl. 1992. On the role of perception in shaping phonological assimilation rules. *Language and Speech* 35, 59–72.

Itô, Junko. 1986. *Syllable theory in prosodic phonology*. Ph.D. dissertation, University of Massachusetts, Amherst.

Itô, Junko, Yoshihisa Kitagawa & Armin Mester. 1996. Prosodic faithfulness and correspondence: Evidence from a Japanese argot. *Journal of East Asian Linguistics* 5, 217–294.

Itô, Junko & Armin Mester. 1986. The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry* 17, 49–73.

Itô, Junko & Armin Mester. 1999. The phonological lexicon. In Natsuko Tsujimura (ed.), *The handbook of Japanese linguistics*, 62–100. Oxford: Blackwell.

Itô, Junko, Armin Mester & Jaye Padgett. 1995. Licensing and underspecification in Optimality Theory. *Linguistic Inquiry* 26, 571–614.

Jaeger, Jeri & John Ohala. 1984. On the structure of phonetic categories. *Tenth Annual Meeting of the Berkeley Linguistics Society* (*BLS* 10), 15–26.

Jakobson, Roman. 1960. *Language in literature*. Cambridge, MA: Harvard University Press.

Jakobson, Roman, Gunnar Fant & Morris Halle. 1952. *Preliminaries to speech analysis* (Technical Report 13). Cambridge, MA: MIT.

Johnson, Keith. 2003. *Acoustic and auditory phonetics*, 2nd edn. Malden, MA & Oxford: Blackwell.

Jun, Jongho. 2004. Place assimilation. In Bruce Hayes, Robert Kirchner & Donca Steriade (eds.), *Phonetically based phonology*, 58–86. Cambridge: Cambridge University Press.

Kang, Yoonjung. 2003. Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean. *Phonology* 20, 219–273.

Katayama, Motoko. 1998. *Optimality Theory and Japanese loanword phonology*. Ph.D. dissertation, University of California, Santa Cruz.

Kawahara, Shigeto. 2003. On a certain kind of hiatus resolution in Japanese. *Phonological Studies* 6, 11–20.

Kawahara, Shigeto. 2006. A faithfulness ranking projected from a perceptibility scale: The case of voicing in Japanese. *Language* 82, 536–574.

Kawahara, Shigeto. 2007. Half-rhymes in Japanese rap lyrics and knowledge of similarity. *Journal of East Asian Linguistics* 16, 113–144.

Kawahara, Shigeto, Hajime Ono & Kiyoshi Sudo. 2006. Consonant co-occurrence restrictions in Yamato Japanese. In Timothy Vance & Kimberly Jones (eds.), *Japanese/Korean Linguistics* 14, 27–38. Stanford, CA: CSLI Publications.

Kawahara, Shigeto & Kazuko Shinohara. 2007. The role of psychoacoustic similarity in Japanese imperfect puns. In Michael Becker (ed.), *University of Massachusetts Occasional Papers in Linguistics* 36, 111–133. Amherst, MA: Graduate Linguistic Student Association (GLSA).

Keating, Patricia. 1988. Underspecification in phonetics. *Phonology* 5, 275–292.

Keating, Patricia. 1996. The phonology–phonetics interface. *UCLA Working Papers in Phonetics* 92, 45–60.

Keating, Patricia & Aditi Lahiri. 1993. Fronted velars, palatalized velars, and palatals. *Phonetica* 50, 73–101.

Kenstowicz, Michael. 2003. Salience and similarity in loanword adaptation: A case study from Fijian. Ms., MIT.

Kiparsky, Paul. 1970. Metrics and morphophonemics in the *Kalevala*. In Donald Freeman (ed.), *Linguistics and literary style*, 165–181. Orlando, FL: Holt, Rinehart and Winston.

Kiparsky, Paul. 1972. Metrics and morphophonemics in the Rigveda. In Michael Brame (ed.), *Contributions to generative phonology*, 171–200. Austin, TX: University of Texas Press.

Kiparsky, Paul. 1981. The role of linguistics in a theory of poetry. In Donald Freeman (ed.), *Essays in modern stylistics*, 9–23. London: Methuen.

Kiparsky, Paul. 1982. Lexical phonology and morphology. In In-Seok Yang (ed.), *Linguistics in the morning calm*, 3–91. Seoul: Hanshin.

Kirchner, Robert. 1998. *An effort-based approach to consonant lenition*. Ph.D. dissertation, UCLA.

Klatt, Dennis. 1968. Structure of confusions in short-term memory between English consonants. *Journal of the Acoustical Society of America* 59, 1208–1221.

Klockars, Alan J. & Gilbert Sax. 1986. *Multiple comparisons*. London: Sage.

Kluender, Keith, Randy Diehl & Beverly Wright. 1988. Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics* 16, 153–169.

Kohler, Klaus. 1990. Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In William J. Hardcastle & Alain Marchal (eds.), *Speech production and speech modelling*, 69–92. Dordrecht: Kluwer.

Kuroda, S.-Y. 1965. *Generative grammatical studies in the Japanese language*. Ph.D. dissertation, MIT.

Kurowski, Kathleen & Sheila Blumstein. 1993. Acoustic properties for the perception of nasal consonants. In Huffman & Krakow (eds.), 197–224.

Kwon, Bo-Young. 2005. The patterns of vowel insertion in IL phonology: The P-map account. *Studies in Phonetics, Phonology and Morphology* 11, 21–49.

LaCharité, Darlene & Carole Paradis. 2005. Category preservation and proximity versus phonetic approximation in loanword adaptation. *Linguistic Inquiry* 36, 223–258.

Lagerquist, Linnes. 1980. Linguistic evidence from paronomasia. *Chicago Linguistic Society* (*CLS*) 16, 185–191.

Lahiri, Aditi & Henning Reetz. 2002. Underspecified recognition. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory Phonology* 7, 637–675. Berlin: Mouton de Gruyter.

Lass, Roger. 1976. *English phonology and phonological theory*. Cambridge: Cambridge University Press.

Lisker, Leigh. 1986. 'Voicing' in English: A catalog of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29, 3–11.

Lombardi, Linda. 1990. The nonlinear organization of the affricate. *Natural Language & Linguistic Theory* 8, 375–425.

Maher, Peter John. 1969. English-speakers' awareness of distinctive features. *Language Sciences* 5, 14.

Maher, Peter John. 1972. Distinctive feature rhyme in German folk versification. *Language Sciences* 19, 19–20.

Malécot, André. 1956. Acoustic cues for nasal consonants: An experimental study involving a tape-splicing technique. *Language* 32, 274–284.

Malone, Joseph. 1987. Muted euphony and consonant matching in Irish. *Germanic Linguistics* 27, 133–144.

Malone, Joseph. 1988a. On the global-phonologic nature of classical Irish alliteration. *Germanic Linguistics* 28, 93–103.

Malone, Joseph. 1988b. Underspecification theory and Turkish rhyme. *Phonology* 5, 293–297.

McCarthy, John. 1988. Feature geometry and dependency: A review. *Phonetica* 43, 84–108.

McCarthy, John & Alan Prince. 1995. Faithfulness and reduplicative identity. In Jill Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.), *University of Massachusetts Occasional Papers in Linguistics* 18, 249–384. Amherst, MA: Graduate Linguistic Student Association (GLSA).

Mester, Armin & Junko Itô. 1989. Feature predictability and underspecification: Palatal prosody in Japanese mimetics. *Language* 65, 258–293.

Mielke, Jeff. 2003. The interplay of speech perception and phonology: Experimental evidence from Turkish. *Phonetica* 60, 208–229.

Miller, George & Patricia Nicely. 1955. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America* 27, 338–352.

Minkova, Donca. 2003. *Alliteration and sound change in early English*. Cambridge: Cambridge University Press.

Mohanan, K. P. 1993. Fields of attraction in phonology. In John A. Goldsmith (ed.), *The last phonological rule: Reflections on constraints and derivations*, 61–116. Chicago: University of Chicago Press.

Mohr, B. & W. S. Wang. 1968. Perceptual distance and the specification of phonological features. *Phonetica* 18, 31–45.

Mukai, Yoshihito. 1996. *Kotoba asobi no jyugyoo zukuri* [Teaching with word games]. Tokyo: Meiji Tosho.

Myers, Jerome & Arnold Well. 2003. *Research design and statistical analysis*. Mahwah, NJ: Lawrence Erlbaum.

Myers, Scott & Benjamin Hanssen. 2005. The origin of vowel-length neutralization in vocoid sequences. *Phonology* 22, 317–344.

Nakamura, Mitsuhiro. 2002. The articulation of the Japanese /r/ and some implications for phonological acquisition. *Phonological Studies* 5, 55–62.

Nishimura, Kohei. 2003. Lyman's law in loanwords. Ms., University of Tokyo.

Ohala, John. 1986. Consumer's guide to evidence in phonology. *Phonology* 3, 3–26.

Ohala, John. 1989. Sound change is drawn from a pool of synchronic variation. In Leiv Egil Breivik & Ernst Håkon Jahr (eds.), *Language change: Contributions to the study of its causes*, 173–198. Berlin: Mouton de Gruyter.

Ohala, John & Manjari Ohala. 1993. The phonetics of nasal phonology: Theorems and data. In Huffman & Krakow (eds.), 225–249.

Otake, Takashi & Anne Cutler. 2001. Recognition of (almost) spoken words: Evidence from word play in Japanese. *European Conference on Speech Communication and Technology* 7 (EUROSPEECH 2001 Scandinavia), 464–468.

Padgett, Jaye. 1991. *Stricture in feature geometry*. Ph.D. dissertation, University of Massachusetts, Amherst.

Paradis, Carol & Jean-François, Prunet (eds.). 1991. *The special status of coronals: Internal and external evidence*. San Diego, CA: Academic Press.

Parker, Steve. 2002. *Quantifying the sonority hierarchy*. Ph.D. dissertation, University of Massachusetts, Amherst.

Peters, Robert. 1963. Dimensions of perception for consonants. *Journal of the Acoustical Society of America* 35, 1985–1989.

Pierrehumbert, Janet B. & David Talkin. 1992. Lenition of /h/ and glottal stop. In Gerard Docherty & Robert Ladd (eds.), *Papers in Laboratory Phonology II*, 90–117. Cambridge: Cambridge University Press.

Pols, Louis. 1983. Three mode principle component analysis of confusion matrices, based on the identification of Dutch consonants, under various conditions of noise and reverberation. *Speech Communication* 2, 275–293.

Price, Patti. 1981. *A cross-linguistic study of flaps in Japanese and in American English*. Ph.D. dissertation, University of Pennsylvania.

Prince, Alan & Paul Smolensky. 1993/2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA & Oxford: Blackwell.

Rice, Keren. 1993. A reexamination of the feature [sonorant]: The status of sonorant obstruents. *Language* 69, 308–344.

Rose, Sharon & Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language* 80, 475–532.

Sagey, Elizabeth. 1986. *The representation of features and relations in nonlinear phonology*. Ph.D. dissertation, MIT.

Seco, G. V., I. A. Menéndez de la Fuente & J. R. Escudero. 2001. Pairwise multiple comparisons under violation of the independence assumption. *Quality and Quantity* 35, 61–76.

Shattuck-Hufnagel, Stefanie & Dennis Klatt. 1977. The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Behavior* 18, 41–55.

Shinohara, Shigeko. 2004. A note on the Japanese pun, *dajare*: Two sources of phonological similarity. Ms., Laboratoire de Psychologie Experimentale.

Singh, Rajendra, David Woods & Gordon Becker. 1972. Perceptual structure of 22 prevocalic English consonants. *Journal of the Acoustical Society of America* 52, 1698–1713.

Singh, Sadanand & John Black. 1966. Study of twenty-six intervocalic consonants as spoken and recognized by four language groups. *Journal of the Acoustical Society of America* 39, 372–387.

Sobkowiak, Włodzimierz. 1991. *Metaphonology of English paronomasic puns*. Frankfurt am Main: Lang.

Spring, Cari. 1994. Combinatorial specification of a consonantal system. In Phillip Spaelti (ed.), *Twelfth West Coast Conference on Formal Linguistics* (*WCCFL* 12), 133–152. Stanford, CA: CSLI Publications.

Stemberger, Joseph. 1991. Radical underspecification in language production. *Phonology* 8, 73–112.

Steriade, Donca. 1994. Positional neutralization and the expression of contrast. Ms., UCLA. [A revised, written-up version of 1993 NELS handout.]
Steriade, Donca. 2000. Paradigm uniformity and the phonetics-phonology boundary. In Michael B. Broe & Janet B. Pierrehumbert (eds.), *Acquisition and the lexicon* (Papers in Laboratory Phonology V), 313–334. Cambridge: Cambridge University Press.
Steriade, Donca. 2001a. Directional asymmetries in place assimilation: A perceptual account. In Elizabeth Hume & Keith Johnson (eds.), *The role of speech perception in phonology*, 219–250. New York: Academic Press.
Steriade, Donca. 2001b. The phonology of perceptibility effect: The P-map and its consequences for constraint organization. Ms., UCLA.
Steriade, Donca. 2003. Knowledge of similarity and narrow lexical override. *29th Annual Meeting of Berkeley Linguistics Society* (*BLS* 29), 583–598.
Stevens, Kenneth & Sheila Blumstein. 1978. Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America* 64, 1358–1368.
Sussman, Harvey, Helen McCaffrey & Sandra Matthews. 1991. An investigation of locus equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America* 90, 1309–1325.
Takizawa, Osamu. 1996. *Nihongo syuuzihyougen no kougakuteki kenkyuu* [Technological study of Japanese rhetorical expressions]. Tokyo: Tsuushin Sogo Kenkyuusho.
Trubetzkoy, Nikolai. 1969. *Principles of phonology*. Berkeley, CA: University of California Press.
van den Broecke, M. P. R. 1976. *Hierarchies and rank orders in distinctive features*. Amsterdam: Van Gorcum Assen.
van Dommelen, Wim. 1999. Auditory accounts of temporal factors in the perception of Norwegian disyllables and speech analogs. *Journal of Phonetics* 27, 107–123.
Walden, Brian & Allen Montgomery. 1975. Dimensions of consonant perception in normal and hearing-impaired listeners. *Journal of Speech and Hearing Research* 18, 445–455.
Walker, Rachel. 2003. Nasal and oral consonant similarity in speech errors: Exploring parallels with long-distance nasal agreement. Ms., University of Southern California.
Walter, Mary-Ann. 2004. Loan adaptation in Zazaki Kurdish. In Michael Kenstowicz (ed.), *Working Papers in Endangered and Less Familiar Languages* 6 (MIT Working Papers in Linguistics), 97–106.
Wang, Marilyn D. & Robert C. Bilger. 1973. Consonant confusions in noise: A study of perceptual features. *Journal of the Acoustical Society of America* 54, 1248–1266.
Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30, 945–980.
Wright, Richard. 1996. *Consonant clusters and cue preservation in Tsou*. Ph.D. dissertation, UCLA.
Yip, Moira. 1999. Reduplication as alliteration and rhyme. *Glot International* 4, 1–7.
Zuraw, Kie. 2007. The role of phonetic knowledge in phonological patterning: Corpus and survey evidence from Tagalog reduplication. *Language* 83, 277–316.
Zwicky, Arnold. 1976. This rock-and-roll has got to stop: Junior's head is hard as a rock. *Chicago Linguistic Society* (*CLS*) 12, 676–697.
Zwicky, Arnold & Elizabeth Zwicky. 1986. Imperfect puns, markedness, and phonological similarity: With fronds like these, who needs anemones? *Folia Linguistica* 20, 493–503.

Authors' addresses : (*Kawahara*)
Linguistics Department, Rutgers University, 18 Seminary Place,
New Brunswick, NJ 08901-1108, USA
kawahara@rci.rutgers.edu

(*Shinohara*)
Institute of Symbiotic Science and Technology,
Tokyo University of Agriculture and Technology, 2-24-16 Nakacho,
Koganei, Tokyo 184-8588, Japan
k-shino@cc.tuat.ac.jp

4. Kawahara, Shigeto and Kazuko Shinohara (2010) Calculating vocalic similarity through puns. The Journal of Phonetic Society of Japan 14.

Calculating vocalic similarity through puns*

Shigeto Kawahara & Kazuko Shinohara
Rutgers University &
Tokyo University of Agriculture and Technology

kawahara@rci.rutgers.edu & k-shino@cc.tuat.ac.jp

Keywords: (perceptual) similarity, vowel, pun, verbal art, corpus analysis

1 Introduction and background

In the verbal art patterns of rhyming and punning, speakers pair two words that involve identical sound sequences. In English rhyming for example, speakers combine two words that have the same syllable rhymes (e.g. *hop* and *stop*). In so-called half rhymes, however, speakers can pair two words with similar—but not identical—sounds; for example, in English, *rock* and *stop* can rhyme,[1] despite the different coda consonants. Though the sounds that stand in correspondence in verbal art patterns need not be identical, recent studies have argued that speakers attempt to maximize the similarity between corresponding segments (we clarify several notions of "similarity" below, particularly in section 4). Recent studies have also argued that speakers measure similarity on *psychoacoustic* or *perceptual* grounds (Fleischhacker, 2000, 2005; Kawahara, 2007; Minkova, 2003; Steriade and Zhang, 2001; Steriade, 2003); i.e., when speakers pair non-identical sounds, they are more likely to pair two sounds if the two sounds are perceptually similar. This body of work argues more particularly that it is perceptual similarity, rather than abstract phonological similarity, that shapes verbal art patterns.

This principle of maximization of perceptual similarity has been argued to hold in Japanese puns (*dajare*) as well (Kawahara, 2009b; Kawahara and Shinohara, 2009, to appear). In composing puns, speakers pair two identical or similar words to create a meaningful expression. They can create puns using identical sound sequences, as in *buta-ga buta-reta* 'A pig was hit' or can use similar but non-identical sound sequences, as in *okosa***m***a-o okosa***n***ai-de* 'Please don't wake up the kid'. We refer to the latter type as "imperfect puns". A previous corpus-based study of Japanese imperfect puns (Kawahara and Shinohara, 2009) has argued that Japanese speakers deploy psychoacoustic or perceptual similarity in composing puns, and as a result puns involving

*We are grateful to Aaron Braver, Edward Flemming, Seunghun Lee, Dan Mash, Toshio Matsuura, Jeff Mielke, Maki Shimotani, Donca Steriade, Kyoko Yamaguchi, two anonymous reviewers, and the audiences at the University of Chicago, the University of Delaware and MIT for their feedback and assistance on this project as well as comments on earlier versions of this paper. A version of this paper appears in A.Altshuler et al (eds.) Rutgers Linguistics Working Papers (RuLing) 3. This project is partly supported by a Research Council Grant from Rutgers University to the first author. Remaining errors are ours.

[1]These examples (*hop* vs. *stop* and *rock* vs. *stop*) are from *Rapper's Delight* by the Sugerhill Gang.

more perceptually similar consonants are found more often than those involving less perceptually similar consonants. For example, Japanese speakers are more likely to pair /m/-/n/ than /p/-/t/ in puns, even though both pairs are distinguished by the same feature—[place]. The high likelihood of /m/-/n/ being paired in puns is arguably grounded in the perceptual similarity between their phonetic realizations, which are perceptually similar for two reasons: (i) place information in formant transition next to nasal consonants is blurred by coarticulatory nasalization; (ii) bursts, which provide important cues for place contrasts (Smits et al., 1996; Stevens and Blumstein, 1978; Takieli and Cullinan, 1979; Winitz et al., 1972), are rarely if ever present in nasals because intraoral airpressure does not rise in sonorants. For these reasons, the members of [m]-[n] pair are perceptually more similar to each other than the members of [p]-[t] pair (see Jun 2004 following Malécot 1956). (Here and throughout, we use / / when we discuss correspondence in pun pairing, and [ ] when we discuss phonetic realizations of particular sounds.)

Evidence from some psycholinguistic studies indeed supports the weaker perceptibility of a nasal place contrast. A similarity judgement experiment by Mohr & Wang (1968) showed that speakers judge nasal minimal pairs as more similar than oral consonant minimal pairs. Pols (1983) found that Dutch speakers perceive place contrast in oral consonants more accurately than in nasal consonants (though see Kawahara 2009a and Winter 2003 for complications). Kawahara & Shinohara (2009) further found in their corpus that other perceptually similar consonants are more often paired than non-similar pairs. They thus conclude that Japanese speakers compose puns using the principle of maximization of perceptual similarity.[2]

Building on the previous analyses of Japanese puns, this paper calculates the distances between the five vowels in Japanese (/a/, /e/, /i/, /o/, /u/) through the analysis of imperfect puns, and shows that the obtained matrix closely corresponds to a matrix of psychoacoustic distance. We thus argue that perceptual similarity governs not only consonantal pairing but also vowel pairing. A model of featural similarity, which measures similarity in terms of the numbers of shared feature specifications (Bailey and Hahn, 2005; Shattuck-Hufnagel, 1986), fails to capture the similarity judgments that Japanese speakers deploy in composing puns.

We will first briefly illustrate how Japanese speakers create puns and clarify our working definition of puns (see also Cutler and Otake, 2002; Kawahara, 2009b; Kawahara and Shinohara, 2009; Otake and Cutler, 2001; Shinohara, 2004). There are two general types of Japanese puns. In the first type, speakers put together two identical or similar sounding words or phrases to make one meaningful expression. In the second type, speakers replace a part of a proper name, a cliché, or a well-known phrase with a similar sounding word; e.g. *maccho-ga uri-no shoojo* 'A girl who's proud to be a macho', which is based on *macchi uri-no shoojo* 'The Little Match Girl'—in these sort of examples, speakers utter one phrase to imply some other phrase (which is to be recovered by the listener). Since the two types of puns may involve different systems, this paper analyzes only

[2]Recent studies also argue that the use of psychoacoustic similarity in verbal art patterns has a parallel in phonological patterns. In phonology too, speakers maximize the perceptual similarity between corresponding segments, say, input strings and output strings (in addition to those cited above, see Hura et al., 1992; Kohler, 1990; Steriade, 2001/2008; Zuraw, 2007). For example, Jun (2004) observes that cross-linguistically, nasals are more likely to assimilate in place than oral consonants—there are languages in which only nasal consonants assimilate (e.g. Malayalam) but there are no languages in which only oral consonants assimilate. Jun argues that this asymmetry derives from the perceptibility difference of the place contrast in nasals and oral consonants—since the nasal pairs are more similar to each other, they are more likely to be exchanged with each other in phonology. Given such parallels between verbal art and phonology, we can use verbal art to probe our linguistic knowledge (Kawahara, 2009a,b). This research strategy is a new and developing area of study, and there have been some extensive studies on Japanese puns in this framework (Kawahara, 2009b; Kawahara and Shinohara, 2009, to appear; Shinohara, 2004).

the first type; i.e. those puns with two similar words or phrases, as in the example like *okosama-o okosanaide*, where two corresponding words/phrases are overtly expressed. In particular this paper analyzes puns using two words that differ in one (or more) vowel(s), as in (1), (2) and (3) (mismatched vowels are shown in bold).

(1) Ha**i**deggaa-no zense-wa ha**e** dek-ka?
Heideggar-GEN previous life-TOP fly copula-Q particle
'Is Heidegger a fly in his previous life?'

(2) Shib**u**ya-wa shib**i**ya.
Shibuya-TOP severe
'Shibuya is severe.'

(3) Sand**a**ru-ga san-d**o**ru-da.
sandal-NOM three-dollars-copula
'A sandal is three dollars.'

2 Method

An introspection-based approach, in which the authors decide which puns are well-formed and which puns are not, would not suffice to illuminate the measure of similarity deployed in puns, because such an approach would be subject to the authors' biases. Besides the problem of subjectivity, wellformedness of puns may be affected by paralinguistic factors such as funniness and skillfulness. In order to uncover phonetic/phonological patterns which may be obscured by these paralinguistic factors, we took a corpus-based statistical approach (Fleischhacker, 2005; Kawahara, 2007; Kawahara and Shinohara, 2009; Steriade, 2003; Zwicky, 1976; Zwicky and Zwicky, 1986). By collecting a large number of examples which were created as puns by many Japanese speakers, we attempted to observe general patterns beyond each speaker's subjective bias and other extra-linguistic factors (see Shattuck-Hufnagel, 1986, p.149 for relevant discussion).

In order to gather a large number of data, we first collected puns like those in (1)-(3) from 17 summary websites providing many written forms of Japanese puns (see Appendix for the list of URLs).[3] These were summary websites that appeared first in a Google-based search[4] using key words like *dajare* 'puns' and *gyagu* 'joke'. We chose these websites because they were the ones that were judged to be most relevant by Google's search algorithm.[5] From these sites, we compiled examples of puns that are sentences with two words with mismatched vowels. In some cases there were two similar examples of puns that used the same words (e.g. *sandaru-ga sandoru-da* 'A sandal is three dollars' and *sandaru-ga sandoru-shita* 'This sandal cost three dollars'). In such cases we counted only one example to avoid a particular pairing of two words influencing our results. To this web-based data set, we added more examples which we elicited from native speakers by asking them to freely come up with puns out of the blue. We first transcribed these puns, coded corresponding parts in each pun (e.g. /sandaru/-/sandoru/), and finally extracted the mismatched vowel pair(s) (/a/-/o/). Finally, for two reasons we excluded cases in which one or

[3] See the conclusion section for the discussion of some issues using orthography-based data.
[4] http://www.google.co.jp
[5] See http://en.wikipedia.org/wiki/Google_search.

both vowels were long[6]: (i) the degree of (dis-)similarity may be affected by pairs in which only one vowel is long, and (ii) Japanese speakers generally judge long vowel pairs to be less similar to each other than short vowel pairs (Kawahara and Shinohara, to appear). In the end, this process resulted in the total of 547 pairs of mismatched vowels.

Given the corpus data, the next step was to calculate a measure of the combinability of the five vowels. For this purpose, we calculated the O/E ratios of each vowel pair, instead of raw observed frequencies (see below for how to calculate O/E ratios). The raw frequencies are a poor indication of combinability of two segments, since, for example, the /a/-/o/ pair may be frequently observed just because /a/ and/or /o/ are frequent sounds in the corpus in the first place. Instead what we need is a measure of frequency of a pair relative to the frequencies of the individual elements (see Trubetzkoy 1939/1969, p.264).

O/E ratios provide a useful measure for our purpose. The measure is a general statistical notion, and has recently been deployed in linguistic work that analyzes combinability of two linguistic elements in corpora (e.g. Berkley, 1994; Coetzee and Pater, 2008; Frisch, 2000; Frisch et al., 2004; Kawahara, 2007; Kawahara and Shinohara, 2009; Pierrehumbert, 1993; Shatzman and Kager, 2007). O/E ratios are ratios between O-values (the actual occurrences of the pairs observed in the corpus) and E-values (how often the pairs are expected to occur if their two individual elements are combined randomly). The O-value of a sound pair /A/-/B/ is how many times the /A/-/B/ pair occurs in the data set, and the E-value of a pair /A/-/B/ is $P(A) \times P(B) \times N$ (where $P(X)$ = the probability of /X/; N = the total number of elements). For example, suppose that /A/ and /B/ occur 50 and 60 times in the corpus with 300 elements, respectively, and that the /A/-/B/ pair occurs 20 times in the corpus. The O-value is then 20 and the E-value is $P(A) \times P(B) \times N$ = $50/300 \times 60/300 \times 300 = 10$. The O/E ratio is thus $20/10 = 2$, which means that the pair occurs twice as much as expected. For further discussion and illustration of this notation, see Frisch et al. (2004) and Kawahara (2007).

Further, we used the reciprocals of O/E ratios as a measure of distance between the five vowels in puns. For example, the O/E ratio of the /a/-/o/ pair is 2.13, and therefore, its distance is 1/2.13=0.47. The rationale of this final step of the analysis is as follows: since O/E ratios represent the combinability—or proximity—of two elements in puns, taking the reciprocal of O/E amounts to obtaining "distance measures in puns". This measure allows us to make a direct comparison with other measures of distance such as phonological and phonetic distance. On the other hand, if we were to use O/E ratios directly, a high O/E value means that the two vowels are "close" to each other in puns, which makes it difficult to make a direct comparison with phonological and phonetic distances.

3 Results

Table 1 shows O-values and E-values of each vowel pair together with total numbers of each vowel, and Table 2 shows their O/E ratios. Table 3 shows the distance matrix.

In order to obtain its two-dimensional representation, we applied a Principle Component Analysis (PCA) to the matrix (Figure 1). This analysis is a multivariate analysis, which finds uncorrelated dimensions that account for the variability in the data as much as possible (see Baayen, 2008,

[6]Examples include *kookeeki-ni kuukeeki* 'Cake eaten at the time of good economy' and *uchi-no-tabi to uchuu-no tabi* 'A space trip with my sock.'

Table 1: The O-values of each vowel pair, the total numbers of each vowel, and E-values of each vowel pair.

O-values							E-values					
	a	e	o	i	u	total		a	e	o	i	u
a	0	71	93	37	36	237	a	0	44.4	43.8	51.1	46.4
e		0	28	84	22	205	e		0	37.9	44.2	40.1
o			0	20	61	202	o			0	43.6	39.5
i				0	95	236	i				0	46.2
u					0	214	u					0

Table 2: The O/E ratios of the five vowels.

	a	e	o	i	u
a	0	1.60	2.13	0.72	0.78
e		0	0.74	1.90	0.55
o			0	0.46	1.54
i				0	2.06
u					0

Chapter 5).[7] The two components in Figure 1 account for 96% of the variability. We used R (R Development Core Team, 1993-2010) to perform the analysis and generate the figure.

In Figure 1, the five vowels are configured in a way that resembles a vowel space (the parallel becomes clearer if we rotate Figure 1 clockwise by 45 degrees). In other words, the relative configurations between the elements are comparable between Figure 1 and a vowel space.

[7]PCA is similar to a Multi-Dimensional Scaling (MDS) analysis, which has been used to create a vowel distance diagram (Lindblom, 1962).

Table 3: The distance matrix of the five vowels. The distances were calculated as reciprocals of O/E ratios.

	a	e	o	i	u
a	0	0.63	0.47	1.38	1.29
e		0	1.35	0.53	1.82
o			0	2.18	0.65
i				0	0.49
u					0

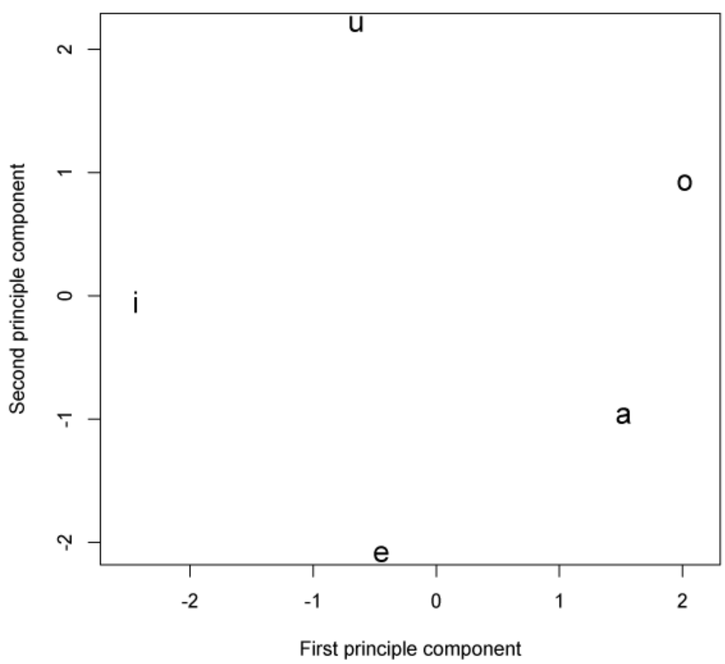

Figure 1: A principle component analysis of the vowel distance matrix which is computed based on combinability in puns (Table 3).

4 Discussion

Figure 1 resembles a standard vowel space, and in this section we further argue that psychoacoustic similarity captures the obtained distance matrix better than phonological, featural similarity. The first model defines similarity in terms of psychoacoustic properties of phonetic realizations between two elements (Fleischhacker, 2000, 2005; Kawahara, 2007; Minkova, 2003; Steriade and Zhang, 2001; Steriade, 2003).[8] The latter model calculates similarity in terms of how many feature specifications two sounds have in common (Bailey and Hahn, 2005; Shattuck-Hufnagel, 1986) (see also Frisch et al., 2004; LaCharité and Paradis, 2005, for other models of phonological similarity using distinctive features). In this paper, for the sake of illustration, we assume the standard feature specifications in Table 4 following Kubozono (1999, pp.100-101) (we assume that non-low back vowels are [+round]—we return to this issue below).

Table 4: The feature specifications of the five vowels.

	a	e	i	o	u
high	-	-	+	-	+
low	+	-	-	-	-
back	+	-	-	+	+
rounded	-	-	-	+	+

There are three reasons why a phonetic model captures the similarity matrix better than a phonological model. First, /a/ is more combinable with /o/ than with /e/ in puns (the distances in Table 3; /a/-/o/: 0.47 vs. /a/-/e/: 0.63). The proximity of /a/ to /o/ in puns has a straightforward explanation from a psychoacoustic point of view; when pronounced by Japanese speakers, both /a/ and /o/ are phonetically realized with low F2 while /e/ is phonetically realized with high F2 (and additionally, the F1 distance between /a/ and /o/ on the one hand and that between /a/ and /e/ on the other are comparable). Relevant evidence from previous production studies is cited in Table 5. The table shows the F2 and F1 values taken from Nishi et al. (2008) (citation forms, averaged over four speakers), Keating & Huffman (1984) (shown as "K&H", averaged over seven speakers), and Hirahara & Kato (1992) (shown as "H&K", in bark values).

In all these studies, in their phonetic realizations, /a/ is closer to /o/ than to /e/ in terms of F2 and /a/ is almost equidistant from /e/ and /o/ in terms of F1. To calculate psychoacoustic distances between the vowels, we calculated Euclidean distances based on the bark values reported by Hirahara & Kato's, using the following equation (Liljencrants and Lindblom, 1972):[9]

$$D_{ij} = \sqrt{(F1_i - F1_j)^2 + (F2_i - F2_j)^2} \qquad (1)$$

[8]A very similar alternative is a model of similarity based on articulatory distances. We chose the psychoacoustic model primarily because psychoacoustic similarity data are readily available to us from previous production/acoustic studies.

[9]We assume, following in particular Klein, Plomp & Pols (1970), that F1 and F2 (in some log scales) provide primary information about vowel quality, although we acknowledge that some other studies propose more complex measures based on the whole spectrum of the vowels (Bladon and Lindblom, 1981; Diehl et al., 2004; Lindblom, 1986, among others). Another measure of perceptual similarity can be obtained from a similarity judgment task (Terbeek, 1977) or a confusion experiment (Shepard, 1972).

Table 5: F2 and F1 of /a,o,e/ in Japanese.

	F2 (Hz) in Nishi et al.	F2 (Hz) in K & H	F2 (Bark) in H & K
a	1182	1383	9.80
o	805	1136	6.97
e	1785	1720	13.10
|a-o|	377	247	2.83
|a-e|	603	337	3.30

	F1 (Hz) in Nishi et al.	F1 (Hz) in K & H	F1 (Bark) in H & K
a	615	631	6.75
o	430	481	4.68
e	437	475	4.69
|a-o|	185	150	2.07
|a-e|	178	156	2.06

where D_{ij} is a distance between V_i and V_j. We calculated the whole distance matrix for the five vowels in Japanese, which is presented in Table 6. We further provide its two-dimensional representation in Figure 2 to compare it with Figure 1.

Table 6: The distance matrix between the five vowels. The distances were calculated as Euclidean distances using F1 and F2 values (in bark) reported in Hirahara & Kato.

	a	e	o	i	u
a	0	3.89	3.51	5.61	3.64
e		0	6.13	2.01	3.47
o			0	7.08	3.41
i				0	3.81
u					0

The analysis in Table 6 shows that [a] is indeed phonetically closer to [o] than to [e] ([a]-[o]: 3.51 vs. [a]-[e]: 3.89). Moreover, in Hirahara & Kato's perception experiment, where they tested the influence of F0 on vowel perception, they found that artificially raising the F0 of original [a] can result in the percept of [o], but not that of [e].

To summarize, the proximity of /a/ to /o/ in pun pairing has a straightforward explanation from a perceptual point of view. On the other hand, the featural similarity view cannot explain the fact that /a/ is closer to /o/ than it is to /e/, because /a/ differs from both /o/ and /e/ in terms of two distinctive features (the /a/-/o/ pair disagrees in [low] and [round]; the /a/-/e/ pair disagrees in [low] and [back]).

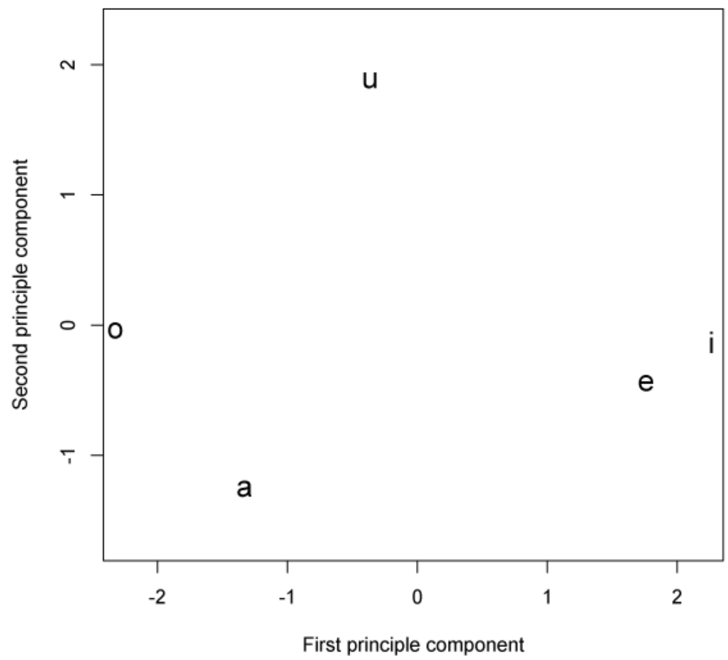

Figure 2: A principle component analysis of the vowel distance matrix in Table 6.

The second piece of evidence for the role of perceptual similarity comes from the behavior of the /i/-/u/ pair. In our pun data, /i/ and /u/ are closer to each other than /e/ and /o/ are (the distances in Table 3; /i/-/u/ : 0.49 vs. /e/-/o/: 1.35). This pattern has a straightforward perceptual explanation because in Japanese, /u/ is phonetically realized with less rounding and backing than /o/ is (Kubozono, 1999; Vance, 1987, and references cited therein), and hence has higher F2, making it closer to front vowels. Table 7 compares F2 and F1 values of /i/-/u/ and /e/-/o/ pairs in the aforementioned three studies.

We observe that /i/ and /u/ are realized with closer F2 than /e/ and /o/, although /e/ and /o/ are realized with slightly closer F1 to each other than /i/ and /u/. Calculating Euclidean distances based on F1 and F2 from Hirahara & Kato's data reveals that [e] and [o] are farther apart from each other than [i] and [u] are ([e]-[o]: 6.13 vs. [i]-[u]: 3.81 in Table 6). Moreover, in Hirahara & Kato's perception experiment, artificially raising F0 in the original [i] could result in the percept of [u], whereas applying the same operation to [e] did not result in the percept of [o]. Thus, the [i]-[u] pair is perceptually closer than the [e]-[o] pair. Hence, perceptual similarity effectively captures the high combinability of /i/ and /u/ in puns.

On the other hand, the featural similarity view fails to explain why /i/-/u/ are closer to each other than /e/-/o/ are in puns, because these pairs both differ in [back] and [round]. One could argue that Japanese /u/ is phonologically [-round]. However, Japanese /u/ is arguably [+round] phonologically for the following reason: when two vowels fuse into one long vowel, the resulting vowel takes the rounding value from the second vowel (e.g. /o/ ([+round]) + /i/ ([-round]) → [e] ([-round]) (Kubozono, 1999); when /a/ ([-round]) and /u/ fuse, the resulting vowel is [+round] [o]. One could also argue that Japanese /u/ is not [+back], but this amendment misses several

Table 7: F2 and F1 of /i,u,e,o/ in Japanese.

	F2 (Hz) in Nishi et al.	F2 (Hz) in K & H	F2 (Bark) in H & K
i	2077	1954	13.80
u	1171	1419	10.00
|i-u|	906	535	3.80

	F2 (Hz) in Nishi et al.	F2 (Hz) in K & H	F2 (Bark) in H & K
e	1785	1720	13.10
o	805	1136	6.97
|e-o|	980	584	6.13

	F1 (Hz) in Nishi et al.	F1 (Hz) in K & H	F1 (Bark) in H & K
i	317	359	2.81
u	349	405	3.12
|i-u|	32	46	0.31

	F1 (Hz) in Nishi et al.	F1 (Hz) in K & H	F1 (Bark) in H & K
e	437	475	4.69
o	430	481	4.68
|e-o|	7	6	0.01

phonological generalizations in Japanese: (i) again, when two vowels fuse into one long vowel, the resulting vowel inherits the backness value from the second vowel (e.g. /a/ ([+back]) + /i/ ([-back]) → [e] ([-back])) (Kubozono, 1999), and when /a/ and /u/ fuse, the resulting vowel is [+back] [o] (ii) Japanese verbal stems cannot end with /a,o,u/, a class of [+back] vowels, and /u/ behaves as [+back]; (iii) only /a, o, u/ can license a palatalization contrast in the preceding (non-coronal) consonants (e.g., /kʲa/-/ka/, /kʲu/-/ku/ vs. */kʲe/-/ke/). Overall, phonologically speaking, /u/ is no less rounded or back than /o/ in Japanese, and therefore featural similarity does not explain why the /i/-/u/ pair is more combinable in puns than the /e/-/o/ pair.

The final piece of evidence for the perceptual similarity hypothesis comes from the comparison between the /e/-/i/ pair and the /o/-/u/ pair. Japanese speakers treat /e/-/i/ as closer than /o/-/u/ in puns (the distances in Table 3; /e/-/i/: 0.53 vs. /o/-/u/: 0.65). As predicted by the perceptual similarity hypothesis, the /e/-/i/ pair has closer phonetic realizations than the /o/-/u/ pair in Japanese (the Euclidean distances in Table 6; [e]-[i] : 2.01 vs. [o]-[u]: 3.41). This phonetic difference in distance between two pairs, /e/-/i/ and /o/-/u/, arises because the latter pair involves a larger F2 difference than the former pair (although the differences in F1 points to the opposite direction), as summarized in Table 8.

Table 8: F2 and F1 of /e,i,o,u/ in Japanese (rearranged from Table 7)

	F2 (Hz) in Nishi et al.	F2 (Hz) in K & H	F2 (Bark) in H & K
e	1785	1720	13.10
i	2077	1954	13.80
|e-i|	292	234	0.70

	F2 (Hz) in Nishi et al.	F2 (Hz) in K & H	F2 (Bark) in H & K
o	805	1136	6.97
u	1171	1419	10.00
|o-u|	366	283	3.03

	F1 (Hz) in Nishi et al.	F1 (Hz) in K & H	F1 (Bark) in H & K
e	437	475	4.69
i	317	359	2.81
|e-i|	120	116	1.88

	F1 (Hz) in Nishi et al.	F1 (Hz) in K & H	F1 (Bark) in H & K
o	430	481	4.68
u	349	405	3.12
|o-u|	81	76	1.56

On the other hand, the featural similarity model offers no explanation for the closer proximity of the /e/-/i/ pair compared to the /o/-/u/ pair, because both of the pairs are distinguished by the same [high] feature.

In summary, the perceptual similarity model accounts for the measure of similarity that Japanese speakers use when composing puns. A phonological similarity model, which relies on the number of shared feature specifications, fails to capture three observations: (i) Japanese speakers treat /a/ as more similar to /o/ than to /e/, (ii) Japanese speakers treat /i/ and /u/ as closer to each other than /e/ and /o/ are, and (iii) Japanese speakers treat /e/ and /i/ as closer to each other than /o/ and /u/.

5 Conclusion

Japanese speakers use knowledge of perceptual similarity in composing puns. On the other hand, there is no clear featural basis for the similarity patterns found in a corpus of Japanese puns. Our result supports a body of recent claims that speakers pay attention to perceptual similarity and deploy that knowledge in creating puns and rhyme patterns (Fleischhacker, 2000, 2005; Kawahara, 2007, 2009b; Kawahara and Shinohara, 2009; Minkova, 2003; Steriade and Zhang, 2001; Steriade, 2003).

Before closing the paper we would like to discuss some remaining issues. First, we obtained our corpus data largely based on written materials. Our use of orthography was not without reason: we followed previous studies that used this approach so as to obtain a large sample of data (Fleischhacker, 2005; Kawahara, 2007; Kawahara and Shinohara, 2009; Steriade, 2003; Zwicky, 1976; Zwicky and Zwicky, 1986). Nevertheless, orthography involves abstraction in the sense that it abstracts away from phonetic details such as coarticulation with surrounding segments and variations due to speech style. To the extent that our claim—that pun formation is based on phonetic details—is on the right track, a future study with speech-based data is warranted to investigate how much phonetic details matter for the formation of puns.

Conversely, we can ask whether standing in correspondence in puns affects phonetic implementation of sounds in puns. An anonymous reviewer pointed out that, for example, when /a/ and /o/ correspond in puns, the [a] can be pronounced as more similar to [o] than otherwise. This pun-specific pronunciation may be an instance of phonetic analogy in which corresponding segments (e.g. morphologically related words, or words that rhyme or are paired in puns) are pronounced as phonetically more similar to each other than otherwise (Steriade, 2000; Yu, 2007). Testing out if phonetic analogy holds in general in the pronunciations of puns is a topic worthy of future investigations. In summary, since we primarily used written sources to obtain a large number of data, we could not analyze the acoustic properties of the vowels as pronounced in puns, but the interplay between the actual pronunciations of puns and the formation of puns provides a promising line for future experimentation.

Finally, another task that is left for future research is to support the perceptual grounding of verbal art patterns with perception experiments. The hypothesis that it is perceptual properties of sounds that are crucial in formation of puns can and should be tested in perception experiments.

6 Appendix

We provide the list of the websites consulted for data collection (the final time for data collection is Sept. 2007):

```
http://www.dajarenavi.net/
http://dajare.jp/
http://www1.tcn-catv.ne.jp/h.fukuda/
http://dajare.noyokan.com/museum/coin.html
http://dajare.noyokan.com/music/index.html
http://www.ipc-tokai.or.jp/˜y-kamiya/Dajare/
http://www.geocities.co.jp/Milkyway-Vega/8361/umamoji3.html.
http://www.webkadoya.com/noumiso/aho/dajare1.htm
http://www.bekkoame.ne.jp/˜novhiko/joke.htm#1
http://planetransfer.com/natsumi/oyajigag/
http://planetransfer.com/natsumi/keijiban/
http://planetransfer.com/natsumi/touko/
http://karufu.net/joke/joke.html
http://www.koyasu.org/royal/neta81.html
```

http://home.att.ne.jp/zeta/sano/dajyare/d-hist.htm
http://www5e.biglobe.ne.jp/~kajilin/tencho-100man-en-hairimasu.html
http://www5d.biglobe.ne.jp/~katumi/newpage19.htm

References

Baayen, H. R. (2008) *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambriedge University Press.

Bailey, T. and Hahn, U. (2005) "Phoneme similarity and confusability." Journal of Memory and Language 52, pp. 339–362.

Berkley, D. (1994) "The OCP and gradient data." Studies in the Linguistic Sciences 24, pp. 59–72.

Bladon, R. A. W. and Lindblom, B. (1981) "Modeling the judgment of vowel quality differences." Journal of the Acoustical Society of America 69, pp. 1414–1422.

Coetzee, A. and Pater, J. (2008) "Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic." Natural Language and Linguistic Theory 26, pp. 289–337.

Cutler, A. and Otake, T. (2002) "Rhythmic categories in spoken-word recognition." Journal of Memory and Language 46, pp. 296–322.

Diehl, R., Lindblom, B. and Creeger, C. (2004) "Increasing realism of auditory representations yields further insights into vowel phonetics." Proceedings of the 15th International Congress of Phonetic Sciences 2, pp. 1381–1384.

Fleischhacker, H. (2000) "Cluster dependent epenthesis asymmetries." In Albright, A. and Cho, T. (eds.) *UCLA Working Papers in Linguistics 5*, (pp. 71–116). Los Angeles: UCLA.

Fleischhacker, H. (2005) *Similarity in Phonology: Evidence from Reduplication and Loan Adaptation*. Doctoral dissertation, UCLA.

Frisch, S. (2000) "Temporally organized lexical representations as phonological units." In Broe, M. and Pierrehumbert, J. (eds.) *Language Acquisition and the Lexicon: Papers in Laboratory Phonology V*, (pp. 283–298). Cambridge: Cambridge University Press.

Frisch, S., Pierrehumbert, J. and Broe, M. (2004) "Similarity avoidance and the OCP." Natural Language and Linguistic Theory 22, pp. 179–228.

Hirahara, T. and Kato, H. (1992) "The effect of F0 on vowel identification." In Tohkura, Y., Vatikiotis-Bateson, E. V. and Sagisaka, Y. (eds.) *Speech Perception, Production and Linguistic Structure*, (pp. 89–112). Tokyo: Ohmsha.

Hura, S., Lindblom, B. and Diehl, R. (1992) "On the role of perception in shaping phonological assimilation rules." Language and Speech 35, pp. 59–72.

Jun, J. (2004) "Place assimilation." In Hayes, B., Kirchner, R. and Steriade, D. (eds.) *Phonetically based Phonology*, (pp. 58–86). Cambridge: Cambridge University Press.

Kawahara, S. (2007) "Half-rhymes in Japanese rap lyrics and knowledge of similarity." Journal of East Asian Linguistics 16, pp. 113–144.

Kawahara, S. (2009a) "Faithfulness, correspondence, and perceptual similarity: Hypotheses and experiments." Journal of the Phonetic Society of Japan 13, pp. 52–61.

Kawahara, S. (2009b) "Probing knowledge of similarity through puns." In Shinya, T. (ed.) *Proceedings of Sophia Linguistics Society*, volume 23, (pp. 110–137). Tokyo: Sophia University Linguistics Society.

Kawahara, S. and Shinohara, K. (2009) "The role of psychoacoustic similarity in Japanese puns: A corpus study." Journal of Linguistics 45, pp. 111–138.

Kawahara, S. and Shinohara, K. (to appear) "Phonetic and psycholinguistic prominences in pun formation: Experimental evidence for positional faithfulness." In McClure, W. and den Dikken, M. (eds.) *Japanese/Korean Linguistics 18*. Stanford: CSLI.

Keating, P. A. and Huffman, M. (1984) "Vowel variation in Japanese." Phonetica 41, pp. 191–207.

Klein, W., Plomp, R. and Pols, C. W. (1970) "Vowel spectra, vowel spaces, and vowel identification." Journal of the Acoustical Society of America 48, pp. 999–1009.

Kohler, K. (1990) "Segmental reduction in connected speech in German: phonological facts and phonetic explanations." In Hardcastle, W. J. and Marchal, A. (eds.) *Speech Production and Speech Modelling*, (pp. 69–92). Dordrecht: Kluwer.

Kubozono, H. (1999) *Nihongo no Onsei: Gendai Gengogaku Nyuumon 2 [Japanese Phonetics: An Introduction to Modern Linguisitcs 2].* Tokyo: Iwanami.

LaCharité, D. and Paradis, C. (2005) "Category preservation and proximity versus phonetic approximation in loanword adaptation." Linguistic Inquiry 36, pp. 223–258.

Liljencrants, J. and Lindblom, B. (1972) "Numerical simulation of vowel quality systems: The role of perceptual contrast." Language 48, pp. 839–862.

Lindblom, B. (1962) "Phonetic aspects of linguistic explanation." Studia Linguistica XXXII, pp. 137–153.

Lindblom, B. (1986) "Phonetic universals in vowel systems." In Ohala, J. and Jaeger, J. (eds.) *Experimental Phonology*, (pp. 13–44). Orlando: Academic Press.

Malécot, A. (1956) "Acoustic cues for nasal consonants: An experimental study involving a tape-splicing technique." Language 32, pp. 274–84.

Minkova, D. (2003) *Alliteration and Sound Change in Early English*. Cambridge: Cambridge University Press.

Mohr, B. and Wang, W. S. (1968) "Perceptual distance and the specification of phonological features." Phonetica 18, pp. 31–45.

Nishi, K., Strange, W., Akahane-Yamada, R., Kubo, R. and Trent-Brown, S. A. (2008) "Acoustic and perceptual similarity of Japanese and American English vowels." Journal of the Acoustical Society of America 124, pp. 576–588.

Otake, T. and Cutler, A. (2001) "Recognition of (almost) spoken words: Evidence from word play in Japanese." Proceedings of European Conference on Speech Communication and Technology 7, pp. 464–468.

Pierrehumbert, J. B. (1993) "Dissimilarity in the Arabic verbal roots." Proceedings of North East Linguistic Society 23, pp. 367–381.

Pols, L. (1983) "Three mode principle component analysis of confusion matrices, based on the identification of Dutch consonants, under various conditions of noise and reverberation." Speech Communication 2, pp. 275–293.

R Development Core Team (1993-2010) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Shattuck-Hufnagel, S. (1986) "The representation of phonological information during speech production planning: Evidence from vowel errors in spontaneous speech." Phonology Yearbook 3, pp. 117–149.

Shatzman, K. and Kager, R. (2007) "A role for phonotactic constraints in speech perception." Proceedings of the 16th International Congress of Phonetic Sciences (pp. 1409–1412).

Shepard, R. (1972) "Psychological representation of speech sounds." In David, E. E. and Denes,

P. B. (eds.) *Human Communication: A Unified View*, (pp. 67–113). New York: McGraw-Hill.
Shinohara, S. (2004) “A note on the Japanese pun, dajare: Two sources of phonological similarity.” Ms. Laboratoire de Psycheologie Experimentale.
Smits, R., Ten Bosch, L. and Collier, R. (1996) “Evaluation of various sets of acoustic cues for the perception of prevocalic stop consonants.” Journal of the Acoustical Society of America 100, pp. 3852–3864.
Steriade, D. (2000) “Paradigm uniformity and the phonetics-phonology boundary.” In Broe, M. B. and Pierrehumbert, J. B. (eds.) *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, (pp. 313–334). Cambridge: Cambridge University Press.
Steriade, D. (2001/2008) “The phonology of perceptibility effect: The P-map and its consequences for constraint organization.” In Hanson, K. and Inkelas, S. (eds.) *The nature of the word*, (pp. 151–179). Cambridge: MIT Press. Originally circulated in 2001.
Steriade, D. (2003) “Knowledge of similarity and narrow lexical override.” In Nowak, P. M., Yoquelet, C. and Mortensen, D. (eds.) *Proceedings of the 29th annual meeting of the Berkeley Linguistics Society*, (pp. 583–598). Berkeley: BLS.
Steriade, D. and Zhang, J. (2001) “Context-dependent similarity.” A talk given at CLS 37.
Stevens, K. and Blumstein, S. (1978) “Invariant cues for place of articulation in stop consonants.” Journal of the Acoustical Society of America 64, pp. 1358–1368.
Takieli, M. and Cullinan, W. (1979) “The perception of temporally segmented vowels and consonant-vowel syllables.” Journal of Speech and Hearing Research (pp. 103–121).
Terbeek, D. (1977) “A cross-language multidimensional scaling study of vowel perception.” UCLA Working Papers in Phonetics. Paper No.3 .
Trubetzkoy, N. S. (1939/1969) *Grundzüge der phonologie*. Güttingen: Vandenhoeck and Ruprecht [Translated by Christiane A. M. Baltaxe 1969, University of California Press].
Vance, T. (1987) *An Introduction to Japanese Phonology*. New York: SUNY Press.
Winitz, H., Scheib, M. E. and Reeds, J. A. (1972) “Identification of stops and vowels for the burst portion of /p,t,k/ isolated from conversation speech.” Journal of the Acoustical Society of America 51, pp. 1309–1317.
Winter, S. (2003) *Empirical Investigations into the Perceptual and Articulatory Origins of Cross-Linguistic Asymmetries in Place Assimilation*. Doctoral dissertation, Ohio State University.
Yu, A. (2007) “Understanding near mergers: The case of morphological tone in Cantonese.” Phonology 24, pp. 187–214.
Zuraw, K. (2007) “The role of phonetic knowledge in phonological patterning: Corpus and survey evidence from Tagalog reduplication.” Language 83, pp. 277–316.
Zwicky, A. (1976) “This rock-and-roll has got to stop: Junior’s head is hard as a rock.” In Mufwene, S., Walker, C. and Steever, S. (eds.) *Proceedings of Chicago Linguistic Society 12*, (pp. 676–697). Chicago: CLS.
Zwicky, A. and Zwicky, E. (1986) “Imperfect puns, markedness, and phonological similarity: With fronds like these, who needs anemones?” Folia Linguistica 20, pp. 493–503.

5. Shinohara, Kazuko and Shigeto Kawahara (2010) Syllable intrusion in Japanese puns, dajare. Proceedings of the 10th meeting of Japan Cognitive Linguistic Association.

Syllable intrusion in Japanese puns, *dajare*

Kazuko Shinohara[1] and Shigeto Kawahara[2]
[1]Tokyo University of Agriculture and Technology, [2]Rutgers University

1. Introduction

This paper reports a corpus-based study of a particular pattern in Japanese puns. In Japanese puns, speakers pair two similar or identical phrases to create an expression. In a series of recent projects, we have been analyzing Japanese puns—known as *dajare*—to investigate linguistic knowledge of similarity (see Kawahara 2009a and Kawahara's website for overviews). This paper reports on a subproject of this larger enterprise, which specifically looks at cases in which one phrase contains an internal extra internal syllable, which does not appear in the other corresponding phrase. The first aim of this paper is descriptive: we investigate what kinds of syllables Japanese speakers allow to intrude in this way when creating puns. Second, through the analysis of such patterns, we argue that in composing puns, Japanese speakers define the measure of similarity of a given syllable with Ø based on various phonetic and psycholinguistic considerations. Finally, we point out parallels between pun pairing patterns and some sound patterns in natural languages.

In analyzing Japanese pun patterns, we have been testing two general theses about linguistic organization proposed by Steriade (2001/2008), which are shown in (1) and (2). (See Kawahara 2009b for a review of a recent body of work building on (1) and (2).)

(1) Speakers attempt to minimize the differences between corresponding segments in verbal art patterns and in phonology.
(2) The measure of similarity has psychoacoustic or perceptual grounds.

Steriade (2003) has supported these theses by analyzing Romanian half rhymes. Building on her work, our previous studies also have supported these theses through the analysis of half rhymes in Japanese rap lyrics (Kawahara 2007) and consonant mismatches and vocalic mismatches in Japanese imperfect puns (Kawahara and Shinohara 2009, 2010). In this paper, we further support these theses by analyzing syllable intrusion in Japanese puns.

Specifically, in this paper we argue that (i) speakers minimize the differences between corresponding segments in the syllable intrusion patterns, (ii) the measure of similarity has psychoacoustic or perceptual grounds, (iii) the psycholinguistic non-salience of affixal elements may also affect pun pairing patterns, and (iv) we find parallels between pun pairing patterns and phonological patterns in natural languages.

2. Background: Japanese puns (*dajare*)

In composing puns (*dajare*), Japanese speakers create expressions using identical or similar sounding words/phrases. The correspondence between the two elements in Japanese puns can be perfect or imperfect. In perfect puns, an identical sound string appears twice. An example appears in (3), in which the same sound sequence [arumikan] is repeated twice. (In this paper, we represent Japanese sentences with *romaji*, the Romanization of Japanese, for the sake of exposition.)

(3)

arumikan-no	ue-ni	aru	mikan
aluminum.can-GEN	top-LOC	exist	orange

'An orange on an aluminum can.'

Imperfect puns, on the other hand, involve similar, but not identical, sound sequences, as in (4)-(6), where underlining indicates the corresponding similar sound sequences, and bold face indicates mismatched sounds. The example in (4) includes a consonantal mismatch, where [m] and [n] are a pair of different, but similar, sounds (Kawahara and Shinohara 2009). The example in (5) involves a vocalic mismatch between [i] and [e] (Kawahara and Shinohara 2010). The example in (6) has an internal extra syllable [ku] in the second word, which is not included in the first word that stands in correspondence.

(4) okosa**m**a-o okosa**n**aide
kid-ACC don't.wake.up
'Don't wake up the kid.'

(5) Ha**i**deggaa-no zense-wa ha**e** dek-ka?
Heidegger-GEN previous.life-TOP fly be-Q
'Was Heidegger a fly in his previous life?'

(6) Shopan-no sho**ku**pan.
Chopin-GEN bread
'Chopin's bread.'

In all the cases in (3)-(6), the two words/phrases in correspondence are overtly expressed. In addition to these examples, we find puns where one of the phrases is not overtly expressed, as in (7).

(7) match**o-ga** uri-no shoojo
macho-NOM specialty-GEN girl
'A girl who is proud of being a macho.'

There are not two corresponding words or phrases in (7). Instead, the entirety of (7) implicitly corresponds with the name of a story *Matchi-uri-no shoojo* 'Little Match Girl'. In our previous studies as well as the current study, we have set aside these cases in which only one element is overtly expressed, although we do not wish to imply that the patterns like (7) are not interesting to analyze.

In summary, there are several types of Japanese puns.[i] Building on our previous work (Kawahara and Shinohara 2009, 2010), the present paper focuses on the type illustrated by the example in (6), that is, the cases of syllable intrusion. Some other examples of syllable intrusion are shown in (8)-(9).

(8) bundoki-o bundo**tto**ki.
protractor-ACC take.away
'Take away the protractor (from him).'

(9) ribaundo shinai yooni ribaa-**de** undoo.
rebound do.not so.that river-LOC exercise
'I will exercise in the river so that I won't gain weight again.'

3. Goals of the study

Recall that our general enterprise aims to test the theses in (1) and (2) by analyzing Japanese puns. These theses are general linguistic principles, and we can recast them more specifically as (10) and (11), as applied to puns.

(10) When pun-makers compose puns, they attempt to minimize the differences between corresponding elements in puns.
(11) The measure of similarity between corresponding elements has psychoacoustic or perceptual grounds.

Our previous studies on Japanese puns have supported (10) and (11) (Kawahara and Shinohara 2009, 2010). In this study, we further test these theses by studying the cases of syllable intrusion in Japanese imperfect puns. We observe that when syllables are intruded, speakers prefer those that are similar to Ø; in other words, the more similar to Ø an intruding syllable is, the more frequently it is observed. We thus argue for the following three points:

(12) Speakers minimize the differences between corresponding segments in the syllable intrusion pattern: specifically, they minimize the differences between the intruded syllables and Ø.
(13) The measure of similarity has psychoacoustic or perceptual grounds.
(14) The psycholinguistic non-salience of affixal elements may affect pun pairing.

4. Method

First to achieve our descriptive goal—to investigate what kinds of syllable intrusion patterns Japanese speakers allow in composing puns—we analyzed puns with syllable intrusion: puns in which one phrase contains an extra internal syllable (intruding syllable) which is not contained in the other phrase. We collected examples from 17 summary websites of *dajare* (see appendix 1), and elicited examples from several native speakers by asking them to compose puns out of the blue. We found about 3,200 examples with several types of mismatches (consonantal mismatch, vocalic mismatch, syllable intrusion, metathesis, and others). Among them, we selected examples of syllable intrusion where one phrase internally contains an extra syllable that is not included in the other phrase. We did not include cases in which one phrase is a subset of the other phrase, e.g. *<u>buta</u>-ga <u>buta</u>reta* 'A pig was hit', because these cases may involve pun domains which do not simply coincide with word boundaries. We also excluded examples where an intruded vowel forms a diphthong with the preceding vowel or is identical to the preceding vowel with no intervening consonant. This process resulted in a total of 149 examples.

Based on the 149 examples, we counted the number of each vowel [a], [i], [u], [e], [o] to investigate which vowel is most frequently intruded and compared the intruded vowel with adjacent vowels. We analyzed only vowels, not consonants, because vowels are (psycho-)acoustically more salient than consonants and because the number of samples we obtained was not large enough to analyze consonants. Japanese has only five vowels but many more consonants—testing the behavior of intruding consonants would require a bigger database; however, see Kawahara and Shinohara 2009: section 4.4 for discussion of consonants that are likely to correspond with Ø in Japanese puns.

To calculate the reliability of our estimates, because the distributions of these vowels are unknown, we calculated 95% bootstrap intervals using a bootstrap method (Efron and Tibshirani 1993). We created 50,000 samples using resampling with replacement, and calculated 95% percentiles over those 50,000 samples. The simulation was done using R (R Development Core Team 1993-2010): the code is illustrated in appendix 2.

5. Results

We found three major patterns of syllable intrusion patterns: [1] copy vowel intrusion, [2] high vowel intrusion, and [3] affixal vowel intrusion. First, about 60% (89 out of 149) of the intruded vowels are copies from one or both of the adjacent syllables. If copied, all kinds of vowels are allowed, as in (15)-(19). The preceding vowels are copied in (15)-(17), and the following vowel is copied in (18). In (19), the intruded vowel comes from either the preceding or the following vowel.[ii]

(15) kanashibari-no kanash**ii** h**i**bari
be.bound.hand.and.foot -GEN sad lark
'A sad lark is bound hand and foot.'

(16) Asaka Mitsuyo-ni asa kamits**u**k**u**-yo
Asaka Mitsuyo-LOC morning bite-PARTICLE
'I bite Mitsuyo Asaka in the morning.'

(17) esukareetaa-de **eree** tsukare-ta
escalator-LOC very tired-PAST
'I got very tired on the escalator.'

(18) Tomakomai-ni-wa tomat**o** k**o**-mai
Tomakomai-LOC-TOP tomato come-not
'Tomatoes will not come to Tomakomai.'

(19) azarashi-ga **a**m**a**z**a**rashi
seal-NOM rain.exposed
'The seal is left in the rain.'

Second, if not copied, high vowels /i, u/ often appear in intruded syllables (46 out of 60 non-copy vowels; 76.7%), as in (20)-(21).

(20) Shootokutaishi-o shoodoku sh**i**tai-shi
Shootokutaishi-ACC sterilize want.to.do-PARTICLE
'I want to sterilize Shootokutaishi.'

(21) Shopan-no shok**u**pan.
Chopin-GEN bread
'Chopin's bread.'

Finally, if the intruding vowels are neither copied nor high, then they must usually be affixal vowels (11 out of 14 non-copy, non-high vowels; 78.6%). In (22)-(24), -*de* and -*da* are all affixes.

(22) ribaundo shinai yooni ribaa-**de** undoo
rebound do.not so.that river-LOC exercise
'I will exercise in the river so that I won't gain weight again.'

(23) Kinshichoo-e iku-no-wa kinshi-**de** choo
Kinshichoo-to go-GEN-TOP prohibited-CONJ PARTICLE
'It is prohibited to go to Kinshichoo.'

(24) Koronbusu mite koron-**da** busu
Columbus see fall.down-PAST ugly.girl

'An ugly girl who saw Columbus and fell down.'

Table 1 shows the counts of these three types of vowels (copy, high, and affixal).

Table 1: The distribution of intruded vowels. The instances of affixes are subsets of non-high, non-copy vowels.

	[u]	[i]	[o]	[e]	[a]	Total
Copy	19	16	22	3	29	89
Non-copy	26	20	4	8	2	60
Affix	-	-	(2)	(7)	(2)	

To summarize then, we observe the following three patterns of syllable intrusion:

[1] Copying of adjacent vowels.
[2] High vowels [i, u].
[3] Affixal vowel intrusion.

Out of 149 examples, there are only three examples that do not fit in any of these categories (two [o] and one [e]). Table 2 provides the 95% bootstrap confidence intervals to test the generalizations in [1]-[3]. They show that the patterns in [1]-[3] are generally statistically reliable. First, the 95% confidence intervals for copied vowels do not overlap with zero except for the one for [e], which means that vowel copying did not presumably occur by chance. Second, among non-copied vowels, the 95% bootstrap intervals for high vowels do not overlap with zero. Third, the 95% bootstrap confidence interval for the affixal [e] does not overlap with zero, although those for [o] and [a] do.

Table 2: The 95% bootstrap intervals of the intruded vowels.

	[u]	[i]	[o]	[e]	[a]
Copy	11-27	9-24	14-31	0-7	20-39
Non-copy, high	17-35	12-28	-	-	-
Non-copy, non-high, affix	-	-	0-5	2-12	0-5
Non-copy, non-high, non-affix	-	-	0-5	0-3	0-0

6. Discussion

6.1. Phonetic and psycholinguistic grounding

Three patterns of syllable intrusion exist: [1] copy vowel intrusion, [2] high vowel intrusion, and [3] affixal vowel intrusion, although the third generalization is not as statistically secure as the first two. All three patterns ([1]-[3]) seem to have psychoacoustic or psycholinguistic grounds: intruding syllables cause differences between corresponding phrases in puns; therefore, the principles of similarity maximization in (10) and (11) predict that the less salient the intruded syllable is—i.e., the closer it is to Ø—the better. We now argue that the observed syllable intrusion patterns make phonetic or psycholinguistic sense in this regard.

First, copy vowels: human auditory systems are sensitive to changes (Delgutte 1997). Therefore, whereas adding a non-copy vowel would be noticeable, adding a copy vowel would allow that vowel to perceptually blend into their environment, making the copy vowel less noticeable. In other words, our auditory system is more likely to detect a vowel if that vowel involves a change from surrounding vowels

(This hypothesis must be tested experimentally.)

Second, there are several reasons that high vowels are perceptually non-intrusive. First high vowels [i, u] are the shortest vowels in Japanese and in other languages (Lehiste 1970) because the lower the vowel, the longer the jaw has to travel. Table 3 summarizes the results of previous production studies on the durations of the five vowels in Japanese, all of which show that the two high vowels are the shortest vowels in Japanese. Due to the shortness of high vowels, they are least perceptually intrusive among (non-copied) vowels (see also Steriade 2001/2008). Second, high vowels can devoice between voiceless consonants and other environments in Japanese, becoming less audible (Tsuchida 1997), which would make high vowels even less intrusive. On the other hand, non-high vowels rarely if ever get devoiced, maintaining their periodic intensity. Third, high vowels may be less intense to begin with than non-high vowels because high vowels involve narrower apertures, although cross-linguistic measures of this correlation are not always consistent (see Parker 2002: Chapter 4).

Table 3: Phonetic durations of Japanese vowels. The data from Han (1962) are ratios with respect to the duration of [u]. Otherwise they are in milliseconds.

[u]	[i]	[o]	[e]	[a]	sources
1.00	1.17	1.26	1.37	1.44	Han 1962
62	70	88	93	99	Sagisaka 1985 (in real words)
58	61	71	79	86	Sagisaka 1985 (in nonce words)
58.3	69.8	77.7	80.0	83.7	Campbell 1992
56.8	67.5	75.4	85.7	82.3	Arai, Warner and Greenberg 2001

Finally, affixal elements are non-salient psycholinguistically, and hence they have a possibility of being ignored in pun pairing patterns. For example, Jarvella and Meijers (1983) show that Dutch listeners can make quicker same/different judgments about roots than about inflection forms. Affixal elements are therefore psycholinguistically non-salient (see Beckman 1997; Smith 2002 for summaries of evidence for psychological non-salience of affixal elements). To summarize then, those vowels that can appear in the syllable intrusion patterns in Japanese are those that are perceptually or psycholinguistically less salient.

6.2. Parallels with linguistic patterns

There may exist phonetic and psycholinguistic grounding of syllable intrusion patterns. These groundings have interesting parallels with phonological patterns in natural language. First, many languages use copy vowels for epenthesis (insertion of vowels that are not present in the underlying representations) (Kawahara 2004; Kitto and de Lacy 1999; Uffman 2006). For example, when Japanese speakers adapt foreign words, they insert a vowel after a word-final consonant. When the word-final consonant is [h] (or its allophones), copy epenthesis occurs, as in (25) (Kawahara 2004). Another example comes from Kolami (Zou 1991: 463) shown in (26), where copying takes place to break up tri-consonantal clusters. In (25) and (26), the epenthesized vowels are copies of the preceding vowels.

(25) a. Bach → [b**a**hh**a**]
b. Gogh → [g**o**hh**o**]
c. Zürich → [tʃuur**i**çç**i**]

(26) a. /ayk+t/ → [**a**y**a**kt] 'swept away' cf. /ayk/ → [ayk]
b. /erk+t/ → [**e**r**e**kt] 'lit (fire)' cf. /erk/ → [erk]

Second, high vowels are commonly used as epenthetic vowels as well. In Japanese loanword adaptation, when the word-final consonants are not [h], high vowels are commonly inserted: [i] after [tʃ] and [u] after other consonants, as in (27) (Katayama 1998: 39-40) (they exceptionally insert [o] after coronal stops, presumably because Japanese phonology does not allow coronal stops preceded by high vowels.).

(27)	a.	church	→	[tʃaatʃ**i**]
	b.	spark	→	[supaak**u**]

Other languages, for example, Turkish, also have high vowel epenthesis, as shown in (28) (Clements and Sezer 1982: 243-244; Howe and Pulleyblank 2004: 11; Steriade 2001/2008). In Turkish nominative forms where there is no vocalic suffix, the stem-final consonant clusters are broken up by high vowel epenthesis.

(28)	NOM.SG	3.POSS	Gloss
[i]	fik^{j}**i**r	fik^{j}r-i	'idea'
[u]	koj**u**n	kojn-u	'bosom'

The use of copy vowels and high vowels as epenthetic vowels makes phonetic sense if speakers are minimizing the perceptual disparities between inputs and outputs and if copy and high vowels are perceptually non-intrusive.

Third, recall that in the syllable intrusion patterns, Japanese speakers sometimes allow affixal vowels to intrude, i.e. speakers can treat affixal elements almost as if they are not there. We again find a parallel in natural language patterns, specifically in the context of reduplication. In reduplication, languages prefer to copy material from roots, rather than from affixes (Nelson 2003: 39; Spaelti 1997: Chapter 4; Urbanczyk 2007: 490). In other words, affixal segments are more easily ignored than root segments in reduplication, just as in pun pairing patterns. For instance, in Axininca Campa, reduplication mainly targets root morphemes, as exemplified in (29), rather than copying from the many affixes present (McCarthy and Prince 1993: 63; Urbanczyk 2007: 490).

(29)	Root	Reduplicated form	Gloss
a.	kawosi	noŋ-**kawosi-kawosi**-wai-t-aki	'bathe'
b.	t^{h}aaŋki	noŋ-**t^{h}aaŋki-t^{h}aaŋki**-wai-t-aki	'hurry'
c.	kint^{h}a	noŋ-**kint^{h}a-kint^{h}a**-wai-t-aki	'tell'

7. Conclusion

When creating puns, Japanese speakers are willing to intrude syllables with copy and high vowels, and to a lesser extent, with affixal vowels. Based on this observation and our previous work, we argued that (i) speakers attempt to minimize the difference between corresponding segments in syllable intrusion in puns (the difference between Ø and the intruded syllable), (ii) the measure of similarity has psychoacoustic or perceptual grounds, (iii) the psycholinguistic non-salience of affixal elements may affect pun pairing patterns, and (iv) we find parallels between pun pairing patterns and natural language patterns. More generally, our work supports the theses proposed by Steriade (2001/2008) concerning how similarity affects phonological organization (see (1)/(2)).

The present study also suggests by way of a case study that perceptual aspects of sounds can be an interesting topic in cognitive linguistics. Although collaboration between cognitive linguists and phonologists/phoneticians has not been much pursued, we suggest that a collaborative effort between these researchers may create a new and interesting field of study.

Acknowledgements
This project is partially supported by a Research Council Grant from Rutgers University to the second author. We are grateful to Aaron Braver, Kelly Garvey, Lara Greenberg, Kazu Kurisu, Jeremy Perkins, Shanna Lichtman, and Kyoko Yamaguchi for their comments on earlier versions of this paper.

Notes:

i There are also other types of puns such as those that involve metathesis, phrasal boundary mismatches, and accentual mismatches, which would be all interesting to analyze (Kawahara 2009a). See the second author's website (http://www.rci.rutgers.edu/~kawahara/pun.html) for remaining agendas for this project.
ii We find that copying from preceding vowels is more common than copying from following vowels, as shown in Table A, although the 95% bootstrap confidence intervals overlap for [u, i, e]. To the extent that this difference in directionality is reliable, it may show that copying from preceding vowels is less intrusive than copying from following vowels.

Table A: The directionality of copying. The values in parentheses show 95% confidence intervals.

	[u]	[i]	[o]	[e]	[a]
From preceding vowel	11 (5-17)	8 (3-14)	15 (8-22)	3 (0-7)	18 (11-26)
From following vowel	4 (1-8)	5 (1-10)	3 (0-7)	0 (0-0)	2 (0-5)
From both	4 (1-8)	3 (1-10)	4 (1-8)	0 (0-0)	9 (1-8)

Appendix 1: The list of the websites used for data collection

http://www.dajarenavi.net/

http://dajare.jp/

http://www1.tcn-catv.ne.jp/h.fukuda/

http://dajare.noyokan.com/museum/coin.html

http://dajare.noyokan.com/music/index.html

http://www.ipc-tokai.or.jp/˜y-kamiya/Dajare/

http://www.geocities.co.jp/Milkyway-Vega/8361/umamoji3.html.

http://www.webkadoya.com/noumiso/aho/dajare1.htm

http://www.bekkoame.ne.jp/˜novhiko/joke.htm#1

http://planetransfer.com/natsumi/oyajigag/

http://planetransfer.com/natsumi/keijiban/

http://planetransfer.com/natsumi/touko/

http://karufu.net/joke/joke.html

http://www.koyasu.org/royal/neta81.html

http://home.att.ne.jp/zeta/sano/dajyare/d-hist.htm

http://www5e.biglobe.ne.jp/˜kajilin/tencho-100man-en-hairimasu.html

http://www5d.biglobe.ne.jp/˜katumi/newpage19.htm

Appendix 2: the bootstrap code

```
# This program uses a Bootstrap method and calculates 95% bootstrap confidence intervals of all the elements in a sample. The number of the original sample is n; resampling is repeated m times. To save space, this code is not complete. Contact the second author to obtain the whole code.
```

```
X<-read.csv("intrusion_bootstrap.csv",header=T)
x<-X[,1] # the list of elements
p<-X[,2] # the probabilities of the elements
n<-149 # specify the size of the original sample
m<-50000 # specify the number of resampling you like

     my.boot.u1<-numeric(0)# create a hash for [u1] etc...
     my.boot.i2<-numeric(0)# 1=copy, 2=non-copy, 3=affix
     my.boot.a3<-numeric(0)
     …
for(i in 1:m) {
     y<-sample(x,n,replace=TRUE,prob=p) #resampling
     my.boot.u1[i]<-length(grep("u1",y)) # count the number of elements and store it
     my.boot.i2[i]<-length(grep("i2",y))
     my.boot.a3[i]<-length(grep("a3",y))
     …    }

Y<-matrix(nrow=13,ncol=2) # prepare a matrix to store results

Y[1,1-2]<-quantile(my.boot.u1,p=c(0.025,0.975))  # calculate the confidence intervals
Y[2,1-2]<-quantile(my.boot.i2,p=c(0.025,0.975)) # and store them in Y
Y[3,1-2]<-quantile(my.boot.a3,p=c(0.025,0.975))
…
```

References

Arai, T., Warner, N., & Greenberg, S. (2001). OGI tagengo denwa onsei koopasu-ni okeru nihongo shizen hatsuwa onsei no bunseki [Analysis of spontaneous Japanese in OGI multi-language telephone speech corpus]. Nihon Onkyoo Gakkai Shunki Happyoukai vol.1 [The Spring Meeting of the Acoustical Society of Japan], 361-362.

Beckman, J. (1997). Positional faithfulness, positional neutralization, and Shona vowel harmony. Phonology, 14, 1-46.

Campbell, N. (1992). Segmental elasticity and timing in Japanese. In Y. Tohkura, E. V. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), Speech perception, production and linguistic structure (p. 403-18). Ohmsha.

Clements, G.N. & Sezer, E. (1982). Vowel and consonant disharmony in Turkish. In H. van der Hulst & N. Smith (Eds.), The structure of phonological representation (p. 213-255). Dordrecht: Foris.

Delgutte, B. (1997). Auditory neural processing of speech. In W. Hardcastle & J. Laver (Eds.), The handbook of phonetic sciences (p.507-538). Oxford: Blackwell.

Efron, B., & Tibshirani, R. J. (1993). An introduction to bootstrapping. Boca Raton: Chapman and Hall/CRC.

Fleischhacker, H. (2005). Similarity in phonology: Evidence from reduplication and loan adaptation. Doctoral dissertation, University of California, Los Angeles.

Han, M. (1962). The feature of duration in Japanese. Onsei no Kenkyuu [Studies in Phonetics], 10, 65-80.

Hawkins, J.& Cutler, A. (1988). Psycholinguistic factors in morphological asymmetry. In J. A. Hawkins, (Ed.), Explaining language universals (p. 280-317). Oxford: Basil Blackwell.

Howe, D., & Pulleyblank, D. (2004). Harmonic scales as faithfulness. Canadian Journal of Linguistics, 49, 1-49.

Jarvella, R., and Meijers, G. (1983). Recognizing morphemes in spoken words: Some evidence for a stem-organized mental lexicon. In G. B. Flores d'Arcais and R. Jarvella (Eds.), The process of language understanding (p. 81-113). New York: John Wiley and Sons.

Katayama, M. (1998). Optimality Theory and Japanese loanword phonology. Doctoral dissertation, University of California, Santa Cruz.

Kawahara, S. (2004). The locality of echo epenthesis: Comparison with reduplication. In K. Moulton & M. Wolf (Eds.), Proceedings of North East Linguistic Society (Vol. 34, p. 295-309). Amherst: GLSA.

Kawahara, S. (2007). Half-rhymes in Japanese rap lyrics and knowledge of similarity. Journal of East Asian Linguistics, 16, 113-144.

Kawahara, S. (2009a). Probing knowledge of similarity through puns. In T. Shinya (Ed.), Proceedings of Sophia Linguistics Society (Vol. 23, p. 110-137). Tokyo: Sophia University Linguistics Society.

Kawahara (2009b). Faithfulness, correspondence, and perceptual similarity: Hypotheses and experiments. Journal of the Phonetic Society of Japan, 13, 52-61.

Kawahara, S., & Shinohara, K. (2009). The role of psychoacoustic similarity in Japanese puns: A corpus study. Journal of Linguistics, 45, 111-138.

Kawahara, S., & Shinohara, K. (2010). Calculating vocalic similarity through puns. Journal of the Phonetic Society of Japan 14.

Kitto, C., & de Lacy, P. (1999). Correspondence and epenthetic quality. In C. Kitto & C. Smallwood (Eds.), Proceedings of AFLA VI: The sixth meeting of the Austronesian Formal Linguistics Association (Toronto Working Papers in Linguistics) (p. 181-200). Toronto: Department of Linguistics, University of Toronto.

Lehiste, I. (1970). Suprasegmentals. Cambridge: MIT Press.

McCarthy, J. & Prince, A. (1993). Prosodic morphology I. Constraint interaction and satisfaction. Ms., University of Massachusetts, Amherst and Rutgers University.

Nelson, N. (2003). Asymmetric anchoring. Doctoral dissertation, Rutgers University.

Parker, S. (2002). Quantifying the sonority hierarchy. Doctoral dissertation, University of Massachusetts, Amherst.

R Development Core Team (1993-2010). “R: A language and environment for statistical computing.” A software, available at http://www.R-project.org.

Sagisaka, Y. (1985). Onsei gousei-no tame-no inritsu seigyo-no kenkyuu [A study on prosodic controls for speech synethesis]. Doctoral dissertation, Waseda University.

Smith, J. (2002). Phonological augmentation in prominent positions. Doctoral dissertation, University of Massachusetts, Amherst.

Spaelti, P. (1997). Dimensions of variation in multi-pattern reduplication. Doctoral dissertation, University of California, Santa Cruz.

Steriade, D. (2001/2008). The phonology of perceptibility effect: The P-map and its consequences for constraint organization. In K. Hanson & S. Inkelas (Eds.), The nature of the word (p. 151-179). Cambridge: MIT Press. (originally circulated in 2001)

Steriade, D. (2003). Knowledge of similarity and narrow lexical override. Proceedings of the Twenty-ninth Meeting of the Berkeley Linguistics Society. 583–598.

Tsuchida, A. (1997). Phonetics and phonology of Japanese vowel devoicing. Doctoral dissertation, Cornell University.

Uffman, C. (2006) Epenthetic vowel quality in loanwords: Empirical and formal issues. Lingua, 116, 1079-1111.

Urbanczyk, S. (2007). Reduplication. In Paul de Lacy (ed.) Cambridge handbook of phonology. Cambridge: Cambridge University Press.

Zou, K. (1991). Assimilation as spreading in Kolami. In F. Ingemann (Ed.) 1990 Mid-America Linguistics Conference Papers. 461-475.

6. Kawahara, Shigeto and Kazuko Shinohara (to appear) Phonetic and psycholinguistic prominences in pun formation. In M. den Dikken and W. McClure (eds.) Japanese/Korean Linguistics 18. CSLI.

Phonetic and psycholinguistic prominences in pun formation: Experimental evidence for positional faithfulness

SHIGETO KAWAHARA
Rutgers University

KAZUKO SHINOHARA
Tokyo University of Agriculture and Technology

1. Introduction

This paper addresses two issues. The first issue is similarity effects in phonology. We commonly observe that speakers maximize the similarity between corresponding segments (e.g. input and output). In particular, several previous studies have noted that speakers can simplify their articulation so long as its consequence is perceptually non-conspicuous (Huang, 2001; Hura et al., 1992; Johnson, 2003; Kawahara, 2006; Kohler, 1990; Steriade, 2001/2008). For example, Japanese speakers can devoice geminates when they occur with another voiced obstruent (Nishimura, 2003, 2006). Kawahara (2006) argues that devoicing of a geminate occurs because it is perceptually non-conspicuous i.e. voiced geminates and voiceless geminates are “perceptually similar enough”. This example illustrates another point: various contextual factors contribute to the measure of similarity. In the Japanese example, geminates can devoice, but singletons do not (when they co-occur within another voiced obstruent). Based on acoustic and perception experiments, Kawahara (2006) argues that this asymmetry arises because a voicing contrast is less perceptible in geminates than in singletons. In other words, the perceptibility of a voicing contrast depends on whether the contrast is hosted by a singleton consonant or a geminate consonant. In general, the position of a contrast matters for the perceptibility of the contrast (Steriade, 2001/2008, and others). To summarize, this paper supports two themes about the similarity effects in phonology: speakers minimize the differences between corresponding elements, and the measure of similarity depends on contextual factors.

The second aim of this paper is to bear on the controversy between positional faithfulness theory and positional markedness theory. Some phonological contrasts are maintained in some positions but neutralized elsewhere (the situation referred to as "positional neutralization") (Trubetzkoy, 1939/1969). For example, Tamil allows mid vowels and rounded vowels only in initial syllables (Beckman, 1997, p.6). In the current framework of Optimality Theory (Prince and Smolensky, 1993/2004), two major theories exist to account for positional neutralization patterns. On the one hand, the positional faithfulness theory posits that speakers prohibit changes in phonetically or psycholinguistically prominent positions (Beckman, 1997). On the other hand, positional markedness theory posits that speakers exert a strong pressure against having a particular contrast/structure in non-prominent positions (Zoll, 1998). Evidence for either position has been put forth in the recent OT literature (positional faithfulness: Casali 1997; Kawahara 2006; Kawahara and Hara to appear; Lombardi 1999; Steriade 2001/2008, among others; positional markedness: de Lacy 2000; Itô and Mester 2003; Prince and Tesar 2004; Smith 2002; Zhang 2004, among others). To bear on this debate, this paper provides independent experimental support for the positional faithfulness theory.

To address these two questions—the issue of similarity effects in phonology and the controversy between the positional faithfulness theory and positional markedness theory—this paper analyzes Japanese puns (*dajare*). Punning is a common practice in Japanese, at least for some speakers. They create sentences using two identical or similar sounding words or phrases, as in *buta-ga butareta* 'A pig was hit', *aizusan-no aisu* 'Ice cream from Aizu' and *okosama-o okosanaide* 'Don't wake up a kid'. Paired words can contain identical sound sequences as in the first example, but they can also contain non-identical pairs of sounds ([z] vs. [s] in the second example, and [m] vs. [n] in the third example). Speakers nevertheless attempt to maximize the similarity between the corresponding words in Japanese imperfect puns (Cutler and Otake, 2002; Kawahara, 2009; Kawahara and Shinohara, 2009; Shinohara, 2004). Our experiments below show that the positions of mismatches affect the wellformedness of imperfect puns—speakers disprefer mismatches in certain phonological positions, in our case in initial syllables and long vowels. We argue that these positional effects are grounded in phonetic and psycholinguistic prominences of these phonological positions, and that positional faithfulness, not positional markedness, can account for our observation.

Finally, before closing this introductory discussion, a remark on our theoretical context is in order. We would like to situate our work in a larger theoretical context, which is growing in interest in using verbal art patterns to probe our linguistic knowledge especially by way of an experimental/corpus-based method (Fabb, 1997; Fleischhacker, 2000, 2005; Itô et al., 1996; Kawahara, 2007, 2009; Kawahara and Shinohara, 2009; Shinohara, 2004; Steriade, 2003; Yip, 1999; Zwicky, 1976; Zwicky and Zwicky, 1986, among others). To the extent that our arguments successfully address the above-mentioned phonological questions, this paper supports a general approach which addresses phonological questions through experimental studies of verbal art.

2. Experiment 1

2.1. Introduction and background

The first experiment tested whether speakers avoid mismatches in initial positions. If speakers attempt to maximize the similarity between corresponding words in puns, we expect that they do avoid mismatches in initial positions, because initial syllables play an important role in word recognition and hence mismatches in these positions would be perceptually salient. Here we briefly review the evidence for the psycholinguistic prominence of initial syllables (the following summary builds on Beckman 1997; see Beckman 1997; Hawkins and Cutler 1988; Smith 2002 for more comprehensive reviews). First, hearing initial portions of words helps listeners to retrieve the whole words in short-term memory recall tasks (Horowitz et al., 1968, 1969; Nooteboom, 1981). Second, in “tip-of-the-tongue” phenomena, speakers can only vaguely remember the word they are trying to pronounce, but cannot remember its exact phonological shape, and in such cases speakers can guess the first sounds more accurately than non-initial sounds (Brown, 1991; Brown and MacNeill, 1966). Third, in tip-of-the-tongue situations, initial sounds help retrieve the whole word (Freedman and Landauer, 1966). Fourth, listeners are faster when detecting mispronunciations in non-initial positions (Cole, 1973; Cole and Jakimik, 1980)—once they hear initial syllables, that input activates words starting with those syllables, and hence the listeners can anticipate what is coming next. Finally, sound symbolism—particular images associated with particular sounds—is stronger word-initially than non-word-initially, at least in Japanese (Bruch, 1986; Kawahara et al., 2008a).

Because of their psycholinguistic prominence, initial syllables exhibit a privileged status in phonology as well (Beckman, 1997). For example, in Sino-Japanese, while initial syllables can contain a variety of consonants, second syllables only allow [t] and [k] (Kawahara et al., 2002; Tateishi, 1990). Viewed from the perspective of Optimality Theory (Prince and Smolensky,

TABLE 1 A correspondence theoretic illustration of the parallel between phonology and pun formation. The top figure (a)=phonological input-output correspondence. The bottom figure (b)=surface-to-surface correspondence in pun.

a. Phonological input-output correspondence

$$\begin{array}{lcccccc} \text{Input} & / & \mathbf{s}_i & a_j & \mathbf{s}_k & u_l & / \\ & & | & | & | & | & \\ \text{Output} & [& \mathbf{s}_i & a_j & \mathbf{t}_k & u_l &] \end{array}$$

b. Pun formation (surface-to-surface correspondence)

$$\begin{array}{lcccccc} \text{Word 1} & [& \mathbf{s}_i & a_j & \mathbf{s}_k & u_l &] \\ & & | & | & | & | & \\ \text{Word 2} & [& \mathbf{s}_i & a_j & \mathbf{t}_k & u_l &] \end{array}$$

1993/2004), if there were an underlying form like /sasu/ (*as per* Richness of the Base), then speakers avoid changing the initial [s] but not the second [s] (perhaps to [satu]).[1] In other words, speakers avoid making changes particularly in initial syllables.

To the extent that phonological patterns and pun patterns are governed by the same principles, we would expect that speakers avoid mismatches in initial syllables in puns as well. Correspondence Theory (McCarthy and Prince, 1995) helps us to illustrate this prediction.[2] As in Table 1(a), in Sino-Japanese phonology speakers allow changes in word-internal positions (/s_k/ $\rightarrow$ [t_k]), but do not allow changes in word-initial positions (/s_i/ $\rightarrow$ [s_i]). If a parallel exists between phonological patterns and pun patterns, then speakers should disprefer mismatches in word-initial positions in puns, as in (b). In both cases, the identity restriction should be stronger word-initially than word-internally. The following experiment supports this prediction.

[1] Kawahara et al. (2002) develop an analysis of these patterns using positional faithfulness constraints. Initial syllables are protected by special faithfulness constraints, which dominate markedness constraints that collectively rule out all consonants but [t, k]. These markedness constraints dominate general faithfulness constraints for Sino-Japanese, which results in neutralization of all consonants to [t, k] in non-initial syllables. A positional-markedness based analysis is also possible, which would use constraints that prohibit consonants other than [t, k] in non-initial syllables.

[2] Our formalization based on Correspondence Theory is not new. Several authors have proposed a correspondence theory of rhymes and other language games (Holtman, 1996; Itô et al., 1996; Steriade, 2003; Steriade and Zhang, 2001; Yip, 1999).

2.2. Method

In order to control for factors other than positional effects, we performed wellformedness judgment experiments. The first experiment tested whether speakers avoid mismatches in initial positions. The stimuli were minimal pairs that contain a pair of sounds that minimally differ in voicing ([t-d], [d-t], [k-g], [g-k], [s-z], [z-s]).[3] To control for the phonological distance between the punning constituents, the stimuli all had the same structure, [X-particle Y]. The punning constituents X and Y were all three syllables long. In one condition the mismatch occurred in the initial syllables (e.g. **s***asetsu-ni* **z***asetsu* ‘I gave up making a left turn’), and in another condition, the mismatch occurred in the second syllables (*hi***s***ashi-ni hi***z***ashi* ‘Sunlight on the sun roof’).[4] (Due to space limitation, we cannot provide a full list of the stimuli—please contact the authors to obtain the list.) Additional filler items were interwoven with the target questions. The participants rated both the funniness and the acceptability of each pun sentence on a 1-to-4 scale for both questions. We were interested in the second question, but we included the first question, so that the participants would tease apart these questions. We put the funniness rating before the wellformedness rating to make it clear that the wellformedness rating should not be based on funniness. The questionnaire started with two sample questions, with one example which is clearly an example of a Japanese pun (*arumikan-no ue-ni aru mikan* ‘An orange on a can’) and one example which clearly is not (*hana-yori dango* ‘Foods are more important than cherry blossom’). The latter example does not involve a pair of similar words/phrases, and hence it is not a good example of a pun. A total of 37 speakers participated in this study, but we excluded eight of them because they did not consider the first example *arumikan-no ue-ni aru mikan* as a good pun or considered *hana-yori dango* as a perfect pun.

[3] We could not control for matches in accents for two reasons. First, we could not find enough minimal pairs if we controlled accents. Second, lexical accents are subject to high interspeaker variability due to dialectal and generational differences, and therefore it was impossible to find minimal pairs that match in accents for all speakers.

We chose a [voice] mismatch because we found it easiest to create the stimuli this way. Replicating our result with other featural mismatches would strengthen our claim about position sensitivity, but we will leave it for future research.

[4] We also included pairs which contained mismatches in final syllables in our experiment (reported in Kawahara et al. 2008b). These syllables behaved just like initial syllables. This result is a bit unexpected because it is known that initial syllables are psycholinguistically more prominent than final syllables, at least in a lexical retrieval task (Nooteboom, 1981). It may be that recency effects (Gupta, 2005; Gupta et al., 2005) are playing a role here—speakers remember the final syllables of the first word most vividly when they find the second punning word, and therefore they avoid mismatches in final syllables because of the vivid memory. See also Brown & McNell (1966) for evidence that speakers remember word-final segments as much as word-initial segments in tip-of-the-tongue phenomena. See also Walter (2002) for some evidence that final syllables are positionally strong in phonology.

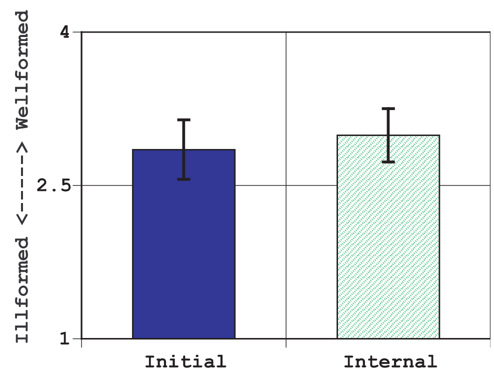

FIGURE 1 Wellformedness of puns with initial mismatches and those with internal mismatches. The error bars = 95% CIs calculated across 29 speakers.

2.3. Results and discussion

Figure 1 illustrates the wellformedness ratings of puns with initial mismatches and those with internal mismatches. As observed, speakers judged mismatches in initial syllables less acceptable than those in non-initial syllables (average ratings: initial 2.85 vs. internal 2.99). One may argue that this difference is too small to be conclusive. Indeed speakers have different standards about pun-wellformedness, and so the effect of the positional difference may look small with respect to relatively large variability. However, the difference is robust within each speaker, and hence statistically significant according to a non-parametric within-subject Wilcoxon signed-rank test ($z = 2.59, p = .01$) (we used this within-subject non-parametric test because we could not assume normality).

3. Experiment 2

3.1. Introduction and background

Experiment 1 supports the principle of positional faithfulness in that speakers avoid mismatches in initial positions, but a question arises whether we observe other kinds of positional effects. The second experiment thus tested whether speakers avoid mismatches in long vowels. Long vowels are, by definition, phonetically long. Different long vowels are more different from each other than different short vowels (Steriade, 2003)—an [aa]-[ii] pair is more different than an [a]-[i] pair. A change in long vowels would be more perceptible also because speakers hyperarticulate long vowels more than short vowels in Japanese, and as a result, long vowels are more (psycho-)acoustically dispersed than short vowels (Hirata and Tsukada, 2003; Hisagi et al., 2008).

Just as in initial syllables, we observe that in phonology speakers avoid

TABLE 2 A correspondence theoretic illustration of the parallel between phonology and pun formation.

a. Phonological input-output correspondence

$$\begin{array}{lcccccc} \text{Input} & / & \mathrm{t}_i & \mathbf{\tilde{a}\tilde{a}}_j & \mathrm{t}_k & \mathbf{\tilde{a}}_l & / \\ & & | & | & | & | & \\ \text{Output} & [& \mathrm{t}_i & \mathbf{\tilde{a}\tilde{a}}_j & \mathrm{t}_k & \mathbf{a}_l &] \end{array}$$

b. Pun formation (surface-to-surface correspondence)

$$\begin{array}{lcccccc} \text{Word 1} & [& \mathrm{t}_i & \mathbf{\tilde{a}\tilde{a}}_j & \mathrm{t}_k & \mathbf{\tilde{a}}_l &] \\ & & | & | & | & | & \\ \text{Word 2} & [& \mathrm{t}_i & \mathbf{\tilde{a}\tilde{a}}_j & \mathrm{t}_k & \mathbf{a}_l &] \end{array}$$

long vowel mismatches. Hindi for example allows a surface nasality contrast in long vowels, but not in short vowels (Steriade 1994 and references cited therein). A hypothetical underlying /tããtã/ would map to [tããta]. As illustrated in Table 2(a), in phonology speakers avoid making changes—or neutralizing contrasts—more in long vowels (/ãã$_j$/ → [ãã$_j$]) than in short vowels (/ã$_l$ → [a$_l$]). Similarly, we expect that speakers avoid mismatches in long vowels more than in short vowels in imperfect puns, as in (b).

3.2. Method

The method is almost identical to Experiment 1, except that we had four practice questions. In addition to the two examples used in the previous experiment, we had *manjuu-o mittsu moratta Akechi Mitsuhide-ga “A, kechi, mittsu hidee”* ‘Akechi Mitsuhide was given three pieces of manjuu, and said “that’s mean, only three?”’—an example of a good pun—and *dakara, kore-wa zura dewa arimasen* ‘I am telling you that this is not a wig’—an example of a non-pun sentence. The design had three fully crossed factors: 10 vowel combinations ([a-i], [a-u], [a-e], [a-o], [i-u], [i-e], [i-o], [u-e], [u-o], [e-o]) × 2 orders (e.g. [a-i] vs. [i-a]) × 2 lengths (short vs. long). An example of a crucial pair was: *jook***uu***-no jook***aa** ‘A joker in the sky’ vs. *ripp***u***-ga ripp***a** ‘The lips are fine’. Additional fillers were added and interwoven with the target items. 26 speakers participated in the study. All the participants judged the sample good puns as good puns and sample non-puns as non-puns, and hence the data from all the participants were included in the analysis.

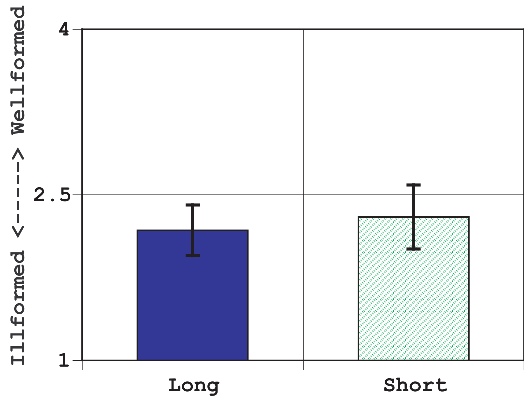

FIGURE 2 Wellformedness of puns with long vowel mismatches and short vowel mismatches. The error bars represent 95% CIs across 26 speakers.

3.3. Results and discussion

Figure 2 illustrates the results. Speakers rated those with long mismatches as worse than short mismatches (average ratings: long 2.09 vs. short 2.30) and the difference is statistically significant according to a within-subject Wilcoxon signed-rank test ($z = 2.93, p < .01$). Mismatches in long vowels are perceptually salient because of their long duration and their hyperarticulated nature, and hence avoided by the participants.

4. Conclusion

4.1. Summary

In summary, speakers avoid mismatches in initial syllables and long vowels in Japanese imperfect puns, just as in phonology. We thus find the same principle both in phonology and pun formation. In this regard we find non-trivial parallels between phonology and verbal art patterns.

4.2. Bearing on the positional faithfulness vs. markedness debate

The principle of positional faithfulness can explain our results, because we observe that speakers avoid mismatches in strong positions in puns, and the avoidance of mismatches in strong positions is what positional faithfulness demands (Beckman, 1997). Positional markedness on the other hand has nothing to say about the results because it evaluates the wellformedness of one form only, but it does not demand anything about the relation between two forms (Zoll, 1998).[5] Overall, therefore, our experiments provide independent support for the principle of positional faithfulness that speakers avoid mismatches in phonetically and psycholinguistically strong positions.

4.3. Concluding discussion

Before closing this paper, we would like to address one final issue. One may argue that our argument is based on “para-linguistic patterns”. However, we find non-trivial parallels between pun patterns and phonology (Kawahara, 2009; Kawahara and Shinohara, 2009), and we would miss the parallels if we treated them separately. In other words, to the extent that we find parallels between pun patterns and phonology, which we hope to have shown that we do in this paper, it is effective to use verbal art patterns to investigate our knowledge of similarity. To conclude, our paper supports the general strategy to probe our linguistic knowledge through the analysis of verbal art patterns.

Acknowledgments

The experiments reported in this paper are a part of a larger project, which investigates knowledge of similarity through puns, as outlined in Kawahara (2009). Further information about this general project can be found at the first author’s website. An earlier version of Experiment 1 was performed as BA thesis research by Nobuhiro Yoshida at Tokyo University of Agriculture and Technology. Experiment 1 was also presented as Kawahara, Shinohara, & Yoshida (2008b). We are grateful to the audience at Sophia University (07/19/2008), the Language Communication, and Cognition (Brighton University, 08/04/2008), and the 18th meeting of Japanese/Korean Linguistics (City University of New York, 11/13/2008). We finally would like to thank Kazu Kurisu, Kyoko Yamaguchi, and Betsy Wang for comments on earlier versions of this paper. This project is partly funded by a Research Council Grant from Rutgers University to the first author. The usual disclaimer applies. This paper supersedes the section 3 of Kawahara (2009).

References

Beckman, Jill. 1997. Positional faithfulness, positional neutralization, and Shona vowel harmony. *Phonology* 14(1):1–46.

Brown, Alan. 1991. A review of the tip-of-the-tongue experience. *Psychological Bulletin* 109(2):204–223.

Brown, Roger and David MacNeill. 1966. The ‘tip of the tongue’ phenomenon. *Journal of Verbal Learning and Verbal Behavior* 5(4):325–337.

Bruch, Julie. 1986. Expressive phonemes in Japanese. *Kansas Working Papers in Linguistics* 11:1–8.

---

[5] We do not wish to imply that positional markedness constraints are not necessary—they do not explain our results.

Casali, Roderic F. 1997. Vowel elision in hiatus contexts: Which vowel goes? *Language* 73(3):493–533.

Cole, Ronald. 1973. Listening for mispronunciations: A measure of what we hear during speech. *Perception & Psychophysics* 13:153–156.

Cole, Ronald and Jola Jakimik. 1980. How are syllables used to recognize words? *Journal of Acoustical Society of America* 67(3):965–970.

Cutler, Anne and Takashi Otake. 2002. Rhythmic categories in spoken-word recognition. *Journal of Memory and Language* 46(2):296–322.

de Lacy, Paul. 2000. Markedness in prominent positions. In O. Matushansky, A. Costa, J. Martin-Gonzalez, L. Nathan, and A. Szczegielniak, eds., *HUMIT 2000: MITWPL 40*, pages 53–66. Cambridge, Mass.: MIT Working Papers in Linguistics.

Fabb, Nigel. 1997. *Linguistics and Literature: Language in the Verbal Arts of the World*. Oxford: Basil Blackwell.

Fleischhacker, Heidi. 2000. Cluster dependent epenthesis asymmetries. In A. Albright and T. Cho, eds., *UCLA Working Papers in Linguistics 5*, pages 71–116. Los Angeles: UCLA.

Fleischhacker, Heidi. 2005. *Similarity in Phonology: Evidence from Reduplication and Loan Adaptation*. Doctoral dissertation, UCLA.

Freedman, Jonathan and Thomas Landauer. 1966. Retrieval of long-term memory: "tip-of-the-tongue" phenomenon. *Psychonomic Science* 4(8):309–310.

Gupta, Prahlad. 2005. Primacy and recency in nonword repetition. *Memory* 13(3):318–324.

Gupta, Prahlad, John Lipinski, Brandon Abbs, and Po-Han Lin. 2005. Serial position effects in nonword repetition. *Journal of Memory and Language* 53:141–162.

Hawkins, John and Anne Cutler. 1988. Psycholinguistic factors in morphological asymmetry. In J. A. Hawkins, ed., *Explaining Language Universals.*, pages 280–317. Oxford: Basil Blackwell.

Hirata, Yukari and Kimiko Tsukada. 2003. The effects of speaking rates and vowel length on formant movements in Japanese. In A. Agwuele, W. Warren, and S.-H. Park, eds., *Proceedings of the 2003 Texas Linguistic Society Conference*, pages 73–85. Somerville, MA: Cascadilla Press.

Hisagi, Miwako, Kanae Nishi, and Winifred Strange. 2008. Acoustic properties of Japanese and English vowels: Effects of phonetic and prosodic context. In M. Endo-Hudson, S.-A. Jun, P. Sells, P. M. Clancy, S. Iwasaki, and S. Sung-Ock, eds., *Japanese/Korean Linguistics 13*. Stanford: CSLI.

Holtman, Astrid. 1996. *A Generative Theory of Rhyme: An Optimality Approach*. Doctoral dissertation, Utrecht Institute of Linguistics.

Horowitz, Leonard, Peter Chilian, and Kenneth Dunnigan. 1969. Word fragments and their reintegrative powers. *Journal of Experimental Psychology* 80(2):392–394.

Horowitz, Leonard, Margeret White, and Douglas Atwood. 1968. Word fragments as aids to recall: The organization of a word. *Journal of Experimental Psychology* 76(2):219–226.

Huang, Tsan. 2001. The interplay of perception and phonology in tone 3 sandhi in Chinese Putonghua. In E. Hume and K. Johnson, eds., *Ohio State University Working Papers in Linguistics 55: Studies on the Interplay of Speech Perception and Phonology*, pages 23–42. Columbus: OSU Working Papers in Linguistics.

Hura, Susan, Björn Lindblom, and Randy Diehl. 1992. On the role of perception in shaping phonological assimilation rules. *Language and Speech* 35:59–72.

Itô, Junko, Yoshihisa Kitagawa, and Armin Mester. 1996. Prosodic faithfulness and correspondence: Evidence from a Japanese argot. *Journal of East Asian Linguistics* 5:217–294.

Itô, Junko and Armin Mester. 2003. *Japanese Morphophonemics*. Cambridge: MIT Press.

Johnson, Keith. 2003. *Acoustic and Auditory Phonetics: 2nd Edition*. Malden and Oxford: Blackwell.

Kawahara, Shigeto. 2006. A faithfulness ranking projected from a perceptibility scale: The case of voicing in Japanese. *Language* 82(3):536–574.

Kawahara, Shigeto. 2007. Half-rhymes in Japanese rap lyrics and knowledge of similarity. *Journal of East Asian Linguistics* 16(2):113–144.

Kawahara, Shigeto. 2009. Probing knowledge of similarity through puns. In T. Shinya, ed., *Proceedings of Sophia Linguistics Society*, vol. 23, pages 110–137. Tokyo: Sophia University Linguistics Society.

Kawahara, Shigeto and Yurie Hara. to appear. Hiatus resolution in Hiroshima Japanese. In M. Abdurrahman, A. Schardl, and M. Walkow, eds., *Proceedings of North East Linguistic Society 38*. Amherst: GLSA.

Kawahara, Shigeto, Kohei Nishimura, and Hajime Ono. 2002. Unveiling the unmarkedness of Sino-Japanese. In W. McClure, ed., *Japanese/Korean Linguistics 12*, pages 140–151. Stanford: CSLI.

Kawahara, Shigeto and Kazuko Shinohara. 2009. The role of psychoacoustic similarity in Japanese puns: A corpus study. *Journal of Linguistics* 45(1):111–138.

Kawahara, Shigeto, Kazuko Shinohara, and Yumi Uchimoto. 2008a. A positional effect in sound symbolism: An experimental study. In *Proceedings of the Japan Cognitive Linguistics Association 8*, pages 417–427. Tokyo: JCLA.

Kawahara, Shigeto, Kazuko Shinohara, and Nobuhiro Yoshida. 2008b. Positional effects in Japanese imperfect puns. A talk presented at Language, Communication, and Cognition (Brighton Univresity, August 4th, 2008).

Kohler, Klaus. 1990. Segmental reduction in connected speech in German: phonological facts and phonetic explanations. In W. J. Hardcastle and A. Marchal, eds., *Speech Production and Speech Modelling*, pages 69–92. Dordrecht: Kluwer.

Lombardi, Linda. 1999. Positional faithfulness and voicing assimilation in Optimality Theory. *Natural Language and Linguistic Theory* 17(2):267–302.

McCarthy, John J. and Alan Prince. 1995. Faithfulness and reduplicative identity. In J. Beckman, L. Walsh Dickey, and S. Urbanczyk, eds., *University of Massachusetts Occasional Papers in Linguistics 18*, pages 249–384. Amherst: GLSA.

Nishimura, Kohei. 2003. *Lyman’s Law in loanwords*. MA thesis, Nagoya University.

Nishimura, Kohei. 2006. Lyman's Law in loanwords. *Phonological Studies [Onin Kenkyuu]* 9:83–90.

Nooteboom, Sieb. 1981. Lexical retrieval from fragments of spoken words: Beginnings vs. endings. *Journal of Phonetics* 9:407–424.

Prince, Alan and Paul Smolensky. 1993/2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden and Oxford: Blackwell.

Prince, Alan and Bruce Tesar. 2004. Learning phonotactic distributions. In R. Kager, J. Pater, and W. Zonneveld, eds., *Constraints in Phonological Acquisition*, pages 245–291. Cambridge: Cambrdige University Press.

Shinohara, Shigeko. 2004. A note on the Japanese pun, dajare: Two sources of phonological similarity. ms. Laboratoire de Psycheologie Experimentale.

Smith, Jennifer. 2002. *Phonological Augmentation in Prominent Positions*. Doctoral dissertation, University of Massachusetts, Amherst.

Steriade, Donca. 1994. Positional neutralization and the expression of contrast. ms. University of California, Los Angeles.

Steriade, Donca. 2001/2008. The phonology of perceptibility effect: The p-map and its consequences for constraint organization. In K. Hanson and S. Inkelas, eds., *The nature of the word*, pages 151–179. Cambridge: MIT Press. originally circulated in 2001.

Steriade, Donca. 2003. Knowledge of similarity and narrow lexical override. In P. M. Nowak, C. Yoquelet, and D. Mortensen, eds., *Proceedings of the twenty-ninth annual meeting of the Berkeley Linguistics Society*, pages 583–598. Berkeley: BLS.

Steriade, Donca and Jie Zhang. 2001. Context-dependent similarity. A talk given at CLS 37.

Tateishi, Koichi. 1990. Phonology of Sino-Japanese morphemes. In G. Lamontagne and A. Taub, eds., *University of Massachusetts Occasional Papers in Linguistics 13*, pages 209–235. Amherst: GLSA.

Trubetzkoy, Nikolai S. 1939/1969. *Grundzüge der phonologie*. Güttingen: Vandenhoeck and Ruprecht [Translated by Christiane A. M. Baltaxe 1969, University of California Press].

Walter, Mary-Ann. 2002. Final position, prominence, and licensing of contrasts. A handout for a talk delivered at 2nd Annual Conference on Contrast and Complexity in Phonology. Toronto, May 3-5.

Yip, Moira. 1999. Reduplication as alliteration and rhyme. *Glot International* 4(8):1–7.

Zhang, Jie. 2004. The role of contrast-specific and language specific phonetics in contour tone distribution. In B. Hayes, R. Kirchner, and D. Steriade, eds., *Phonetically-based Phonology*, pages 157–190. Cambridge: Cambridge University Press.

Zoll, Cheryl. 1998. Positional asymmetries and licensing. ms. MIT.

Zwicky, Arnold. 1976. This rock-and-roll has got to stop: Junior's head is hard as a rock. In S. Mufwene, C. Walker, and S. Steever, eds., *Proceedings of Chicago Linguistic Society 12*, pages 676–697. Chicago: CLS.

Zwicky, Arnold and Elizabeth Zwicky. 1986. Imperfect puns, markedness, and phonological similarity: With fronds like these, who needs anemones? *Folia Linguistica* 20:493–503.

7. Kawahara, Shigeto (2007) Half rhymes in Japanese rap lyrics and knowledge of similarity. Journal of East Asian Linguistics 16.2: 113-144.

J East Asian Linguist
DOI 10.1007/s10831-007-9009-1

Half rhymes in Japanese rap lyrics and knowledge of similarity

Shigeto Kawahara

Received: 19 July 2005 / Accepted: 14 April 2006

Abstract Using data from a large-scale corpus, this paper establishes the claim that in Japanese rap rhymes, the degree of similarity of two consonants positively correlates with their likelihood of making a rhyme pair. For example, similar consonant pairs like {m-n}, {t-s}, and {r-n} frequently rhyme whereas dissimilar consonant pairs like {m-ʃ}, {w-k}, and {n-p} rarely do. The current study adds to a body of literature that suggests that similarity plays a fundamental role in half rhyme formation (A. Holtman, 1996, A generative theory of rhyme: An optimality approach, PhD dissertation. Utrecht Institute of Linguistics; R. Jakobson, 1960, Linguistics and poetics: Language in literature, Harvard University Press, Cambridge; D. Steriade, 2003, Knowledge of similarity and narrow lexical override, Proceedings of Berkeley Linguistics Society, 29, 583–598; A. Zwicky, 1976, This rock-and-roll has got to stop: Junior's head is hard as a rock. Proceedings of Chicago Linguistics Society, 12, 676–697). Furthermore, it is shown that Japanese speakers take acoustic details into account when they compose rap rhymes. This study thus supports the claim that speakers possess rich knowledge of psychoacoustic similarity (D. Steriade, 2001a, Directional asymmetries in place assimilation. In E. Hume, & K. Johnson (Eds.), The role of speech perception in phonology (pp. 219–250). San Diego: Academic Press.; D. Steriade, 2001b, The phonology of perceptibility effects: The P-map and its consequences for constraint organization, ms., University of California, Los Angeles; D. Steriade, 2003, Knowledge of similarity and narrow lexical override, Proceedings of Berkeley Linguistics Society, 29, 583–598).

Keywords Similarity · Half-rhymes · Verbal art · P-map

S. Kawahara (✉)
Department of Linguistics, 226 South College,
University of Massachusetts,
Amherst, MA 01003, USA
e-mail: kawahara@linguist.umass.edu

1 Introduction

Much recent work in phonology has revealed that similarity plays a fundamental role in phonological organization. The list of phenomena that have been argued to make reference to similarity is given in (1).

(1) The effects of similarity in phonology

Input-output mapping: Phonological alternations often take place in such a way that the similarity between input and output is maximized; the rankings of faithfulness constraints (McCarthy and Prince 1995; Prince and Smolensky 2004) are determined by perceptual similarity rankings (Fleischhacker 2001, 2005; Kawahara 2006; Steriade 2001a, b; Zuraw 2005; see also Huang 2001; Hura, Lindblom, and Diehl 1992; Kohler 1990).

Loanword adaptation: Borrowers strive to mimic the original pronunciation of loanwords as much as possible (Adler 2006; Kang 2003; Kenstowicz 2003; Kenstowicz and Suchato 2006; Silverman 1992; Steriade 2001a; Yip 1993).

Paradigm uniformity: Morphologically related words are required to be as similar as possible both phonologically (Benua 1997) and phonetically (Steriade 2000; Yu 2006; Zuraw 2005).

Reduplicative identity: Reduplicants and corresponding bases are required to be as similar as possible; as a consequence, phonological processes can under- or over-apply (McCarthy and Prince 1995; Wilbur 1973).

Gradient attraction: The more similar two forms are, the more pressure they are under to get even more similar (Burzio 2000; Hansson 2001; Rose and Walker 2004; Zuraw 2002).

Similarity avoidance: Adjacent or proximate similar sounds are avoided in many languages, including Arabic (Greenberg 1950; McCarthy 1988), English (Berkley 1994), Japanese (Kawahara, Ono, and Sudo 2006), Javanese (Mester 1986), Russian (Padgett 1992), and others. The effect is also known as Obligatory Contour Principle (OCP; Leben 1973; see also Côté 2004; McCarthy 1986; Yip 1988).

Contrast dispersion: Contrastive sounds are required to be maximally or sufficiently dissimilar (Diehl and Kluender 1989; Flemming 1995; Liljencrants and Lindblom 1972; Lindblom 1986; Padgett 2003; see also Gordon 2002).

As listed in (1), similarity is closely tied to the phonetic and phonological organization of a grammar. Therefore, understanding the nature of linguistic knowledge about similarity is important to current theories of both phonetics and phonology.

Against this background, this paper investigates the knowledge of similarity that Japanese speakers possess. I approach this investigation by examining the patterns of half rhymes found in Japanese rap lyrics.[1] A half rhyme, also known as a partial or

[1] The terms ‘‘rap’’ and ‘‘hip hop’’ are often used interchangeably. However, technically speaking, ‘‘rap’’ is a name for a particular musical idiom. ‘‘Hip hop’’ on the other hand can refer to the overall culture including not only rap songs but also related fashions, arts, dances, and others activities. See Manabe (2006) for the sociological development of rap songs and hip hop culture in Japan, as well as the evolution of technique in rap rhyming in Japanese. See Kawahara (2002) and Tsujimura, Okumura, and Davis (2006) for previous linguistic analyses on rhyme patterns in Japanese rap songs.

imperfect rhyme, is a rhyming pair that falls short of strict identity. To take a few English examples, in *Oxford Town*, *BOB DYLAN* rhymes *so**n*** with *bo**m**b* (Zwicky, 1976); in *Rapper's Delight* by *THE SUGARHILL GANG*, *sto**p*** and *ro**ck*** rhyme (in this paper, song titles are shown in *italics* and artists' names in *CAPITAL ITALICS*). Even though the {n-m} and {p-k} pairs do not involve strictly identical consonants, the rhyming consonants are nevertheless congruent in every feature except place. Previous studies since Jakobson (1960) have noted this increasing production of half rhymes in a given pair of consonants that share similar features: the tendency for half rhymes to involve similar consonants has been observed in many rhyming traditions, including several African languages (Tigrinya, Tigre, Tuareg: Greenberg 1960, pp. 941–942, p. 946), English Mother Goose (Maher 1969), English rock lyrics (Zwicky 1976), German folk songs (Maher 1972), Irish verses (Malone 1987), Romanian half rhymes (Steriade 2003), 18th and 19th-century Russian verses (Holtman 1996, pp. 34–35), and Turkish poems (Malone 1988a). The role of similarity in poetry has been pointed out also for alliterations in languages such as Middle English (Minkova 2003), Early Germanic (Fleischhacker 2005), Irish (Malone 1987, 1988b), and Somali (Greenberg 1960). Finally, imperfect puns are also known to involve pairs of similar consonants in English (Fleischhacker 2005; Zwicky and Zwicky 1986) as well as in Japanese (Shinohara 2004). In a nutshell, cross-linguistically, similar consonants tend to form consonant pairs in a range of verbal art patterns.

This paper establishes the claim that the same tendency is observed in the rhyming patterns of Japanese rap lyrics—the more similar two consonants are, the more likely they are to make a rhyme pair. The folk definition of Japanese rap rhymes consists of two rules: (i) between two lines that rhyme, the line-final vowels are identical, but (ii) onset consonants need not be identical (see, e.g., http:// lyricz.info/howto.html). For example, *KASHI DA HANDSOME* ends two rhyming lines with *m**a**d**e*** 'until' and *d**a**r**e*** 'who' (*My Way to the Stage*), where the corresponding vowel pairs are identical ({a-a}, {e-e}), but the consonant pairs are not ({m-d, d-r}).[2] However, despite the fact the folk definition ignores consonants in rap rhyming, it is commonly the case that similar consonants make rhyme pairs. An illustrative example is given in (2).

(2) *Mastermind* (DJ HASEBE feat. MUMMY-D & ZEEBRA)

a. kettobase **kettobase**
kick it kick it
'Kick it, kick it'

b. kettobashita kashi de **gettomanee**
funky lyrics with get money
'With funky lyrics, get money'

In (2), *MUMMY-D* rhymes *kettobase* 'kick it' and *gettomanee* 'get money', as suggested by the fact that all of the vowels in the two words agree ({e-e}, {o-o}, {a-a}, {e-e}). However, with the exception of the second pair of consonants ({tt-tt}), the consonant pairs ({k-g}, {b-m}, and {s-n}) are not strictly identical.

However, similar to the English half rhymes cited above, these rhyming pairs nevertheless involve similar consonants—all of the pairs agree in place. An example like (2) suggests that similar consonants tend to form half rhymes in Japanese, just as

[2] The quotation of the Japanese rap lyrics in this paper is permitted by the 32nd article of the Copy Right Law in Japan. Thanks are due to Akiko Yoshiga at JASRAC for her courtesy to go over the manuscript and check the legality of the quotations.

they do in many other rhyming traditions. Parallel examples are commonplace: in *Forever*, *LIBRO* rhymes ***d**ai**sh**oo**t**ai* 'big invitation' with ***s**ai**j**oo**k**ai* 'top floor', which involves pairs of similar obstruents ({d-s}, {ʃ-(d)ʒ}, {t-k}).[3] Similarly, in *Project Mou*, *TSURUGI MOMOTAROU* rhymes ***r**en**ch**uu* 'they' and ***n**en**j**uu* 'all year', and this rhyming pair again involves similar consonant pairs ({r-n}, {tʃ-(d)ʒ}).

In order to provide a solid statistical foundation for the claim that similarity plays a role in Japanese rap rhyme formation, this paper presents data from a large-scale corpus showing that if a pair of two consonants share increasingly similar features, they are more likely to form a rhyme pair.

Not only does this paper show that similarity plays a role in half rhyme patterns, it also attempts to investigate what kind of knowledge of similarity is responsible for the similarity pattern observed in Japanese rap lyrics. Specifically, I argue that we need to take acoustic details into account when we compute similarity, i.e., Japanese speakers have a fairly rich sensitivity to detailed acoustic information when they compose rhymes. This conclusion accords with Steriade's (2001a, b) claim that speakers possess what she calls the P-Map—the repository of their knowledge of similarity between sounds.

In short, the present study is a detailed investigation of Japanese speakers' knowledge of similarity through the analysis of half rhymes in Japanese rap songs. More broadly speaking, this work follows a body of literature that analyzes poetry as a source of information about the nature of linguistic knowledge (in addition to the references cited above, see Fabb 1997; Halle and Keyser 1971; Kiparsky 1970, 1972, 1981; Ohala 1986; Yip 1984, 1999, among others).

The rest of this paper proceeds as follows. Section 2 describes how the data were collected and analyzed and gives an overview of the aspects of Japanese rap rhymes. Section 3 presents evidence for the correlation between similarity and rhymability in Japanese rap songs. Section 4 argues that speakers take acoustic details into account when they compose rhymes.

2 Method

2.1 Compiling the database

This subsection describes how the database was compiled for this study. In order to statistically verify the hypothesis that similarity plays a crucial role in the formation of Japanese half rhymes, lyrics in 98 Japanese rap songs were collected. (The complete list of songs is given in Appendix 1.)

In Japanese rap songs, the rhyme domain of two lines is defined as the sequence of syllables at the ends of the lines in which the vowels of the two lines agree. This basic principle is illustrated by the example in (3).

(3) *Hip-hop Gentleman* (DJ MASTERKEY feat. BAMBOO, MUMMY-D & YAMADA-MAN)

a. geijutsu wa **bakuhatsu**
art TOPIC explosion
'Art is explosion.'

[3] [dʒ] is said to allophonically alternate with [ʒ] intervocalically (Hattori 1984, p. 88). However, in conversational speech, both variants occur in both positions—they are more or less free variants of the same phoneme. I therefore treat this phoneme as [dʒ] in this paper.

b.	ikiru-no	ni	**<u>yakudatsu</u>**
	living	for	useful
	'Useful for life.'		

As seen in (3), a rhyme domain can—and usually does—extend over more than a single syllable. The onset of a rhyme domain is very frequently signaled by a boundary tone. The boundary tone is usually a high tone (H) followed by a low tone (L) (with subsequent spreading of L), and this boundary tone can replace lexical tones.[4] For example, in (3), *bakuhatsu* 'explosion' and *yakudatsu* 'useful' are pronounced with HLLL contours even though their lexical accentual patterns are LHHH and LHHL, respectively. Given a rhyme pair like ***bakuhatsu*** vs. ***yakudatsu*** in (3), the following consonant pairs were extracted: {b-y}, {k-k}, {h-d}, {ts-ts}.

Another noteworthy fact about Japanese rhymes is the occasional presence of an ''extrametrical'' syllable, i.e., a word-final syllable in one line occasionally contains a vowel that does not have a correspondent vowel in the other line (Kawahara 2002; Tsujimura, Okumura, and Davis 2006). An example is shown in (4), in which the final syllable [hu] is excluded from the rhyme domain; the extrametrical elements are denoted by < >.

(4) *Koko Tokyo* (AQUARIUS feat. S-WORD, BIG-O & DABO)

a.	biru	ni	umaru	**<u>kyoo</u>**	
	building	DATIVE	buried	today	
	'Buried in the buildings today.'				
b.	Ashita	ni	wa	wasurechimaisouna	**<u>kyoo<hu></u>**
	tomorrow	by	TOPIC	forget-Antihonorific	fear
	'The fear that we might forget by tomorrow.'				

Some other examples of extrametrical syllables include ***<u>kuruma</u>*** 'car' vs. *ooban-**<u>buruma</u>**<i>* 'luxurious behavior' (by *ZEEBRA* in *Golden MIC*), ***<u>iiyo</u>****<ru>* 'talk to' vs. ***<u>iiyo</u>*** 'good' (by *GO* in *Yamabiko 44-go*), and ***<u>hyoogenryo</u>****<ku>* 'writing ability' vs. ***<u>foomeeshon</u>*** 'formation' (by *AKEEM* in *Forever*).[5] Extrametrical syllables whose vowels appear in only one line of the pair were ignored in compiling the database for the current study.

Some additional remarks on how the rhyme data were compiled are in order. First, when the same phrase was repeated several times, the rhyme pairs in that phrase were counted only once, in order to prevent such repeated rhyme pairs from having unfairly large effects on the overall corpus. Second, only pairs of onset consonants were

[4] Other types of boundary tones can sometimes be used. In *Jibetarian*, for example, *SHIBITTO* uses HLH boundary tones to signal the onset of rhyme domains. *LIBRO* uses LH boundary tones in *Taidou*.

[5] As implied by the examples given here, it is usually high vowels that can be extrametrical. My informal survey has found many instances of high extrametrical vowels but no mid or low extrametrical vowels (a systematic statistical study is yet to be performed). This observation correlates with the fact that Japanese high vowels can devoice and tend to become almost inaudible in some environments (Tsuchida 1997), and also with the claim that high vowels are perceptually closest to zero and hence are most likely to correspond with zero in phonology, i.e., they are most prone to be deleted and epenthesized cross-linguistically (Davis and Zawaydeh 1997; Howe and Pulleyblank 2004; Kaneko and Kawahara 2002; Pulleyblank 1998; Tranel 1999).

To the extent that this correlation is valid, we may be able to examine the extent to which each consonant is perceptually similar to zero by looking at the likelihood of consonants rhyming with zero. Unfortunately, in compiling the database, rhyme pairs in which one member is zero were often ignored. This topic thus needs to be addressed in a new separate study.

counted in this study. Coda consonants were ignored because they place-assimilate to the following consonant (Itô 1986). Third, the compilation of the data was based on surface forms. For instance, [ʃi], which is arguably derived from /si/, was counted as [ʃ]. The decision followed Kawahara et al. (2006) who suggest that similarity is based on surface forms rather than phonemic forms (see Kang 2003; Kenstowicz 2003; Kenstowicz and Suchato 2006; Steriade 2000 for supporting evidence for this view; see also Sect. 4; cf. Jakobson 1960; Kiparsky 1970, 1972; Malone 1987).

2.2 A measure of rhymability

The procedure outlined above yielded 10,112 consonantal pairs. The hypothesis tested in this paper is that the more similar two consonants are, the more frequently they make a rhyme pair—in other words, it is not the case that segments are randomly paired up as a rhyming pair. In order to bear out this notion, we need to have a measure of rhymability between two segments, i.e., a measure of how well two consonants make a rhyme pair. For this purpose, the so-called O/E values are used, which are observed frequencies relativized with respect to expected frequencies.

To explain the notion of O/E values, an analogy might be useful. Suppose that there are 100 cars in our sample; 10 of them (10%) are red cars, and 60 of them (60%) are owned by a male. We expect *ceteris paribus* that the number of cars that are both red and owned by a male is 6 (=.1 × .6 × 100). This value is called the expected frequency of the {red-male} pair (or the E value)—mathematically, the expected number for pairs of $\{x\ y\}$ can be calculated as: $\mathrm{E}(x, y) = P(x) \times P(y) \times N$ (where N is the total number of pairs). Suppose further that in reality the car types are distributed as in Table 1, in which there are nine cars that are both red and owned by a male. We call this value the observed frequency (or the O value) of the {red-male} pair.

In Table 1 then, {red} and {male} combine with each other more often than expected (O = 9 > E = 6). More concretely, the {red-male} pairs occur 1.5 times as often as the expected value (=9/6). This value—the O value divided by the E value—is called the O/E value. An O/E value thus represents how frequently a particular pair actually occurs relative to how often it is expected to occur.[6] If O/E > 1, it means that the pair occurs more often than expected (overrepresented); if O/E < 1, it means that the pair occurs less often than expected (underrepresented). To take another example, the {red-female} pair has an O/E value of .25 (O = 1, E = 4). This low O/E value indicates that the combination of {red} and {female} is less frequent than expected.

In this way, O/E values provide a measure of the combinability of two elements, i.e., the frequency of two elements to co-occur. I therefore use O/E values as a measure of the rhymability of two consonants: the higher the O/E value, the more often the two consonants will rhyme. Parallel to the example discussed above, O/E values greater than 1 indicate that the given consonant pairs are combined as rhyme pairs more frequently than expected while O/E values less than 1 indicate that the given consonant pairs are combined less frequently than expected. Based on the O values obtained from the entire collected corpus (i.e., counts from the 20,224

[6] This measure has been used by many researchers in linguistics, and the general idea at least dates back to Trubetzkoy (1939[1969]): "[t]he absolute figures of actual phoneme frequency are only of secondary importance. Only the relationship of these figures to the theoretically expected figures of phoneme frequency is of real value" (1969, p. 264).

Table 1 The observed frequencies of car types

	Red	Non-red	Sum
Male	9	51	60
Female	1	39	40
Sum	10	90	100

rhyming consonants), the E values and O/E values for all consonant pairs were calculated. The complete list of O/E values is given in Appendix 3.

3 Results

The results of the O/E analysis show that rhymability correlates with similarity: the more similar two consonants are, the more frequently they rhyme. The analysis of this section is developed as follows: in Sect. 3.1, I provide evidence for a general correlation between similarity and rhymability. In Sect. 3.2 and Sect. 3.3, I demonstrate that the correlation is not simply a general tendency but in fact that all of the individual features indeed contribute to similarity as well. Building on this conclusion, Sect. 3.4 presents a multiple regression analysis which investigates how much each feature contributes to similarity.

3.1 Similarity generally correlates with rhymability

First, to investigate the general correlation between similarity and rhymability, as a first approximation, the similarity of consonant pairs was estimated in terms of how many feature specifications they each had in common (Bailey and Hahn 2005; Klatt 1968; Shattuck-Huffnagel and Klatt 1977; Stemberger 1991): other models of similarity are discussed below. Seven features were used for this purpose: [sonorant], [consonantal], [continuant], [nasal], [voice], [palatalized], and [place]. [Palatalized] is a secondary place feature, which distinguishes plain consonants from palatalized consonants. Given these features, for example, the {p-b} pair shares six feature specifications (all but [voi]), and hence it is treated as similar by this system. On the other hand, a pair like {ʃ-m} agrees only in [cons], and is thus considered dissimilar.

In assigning feature specifications to Japanese sounds, I made the following assumptions. Affricates were assumed to be [±cont], i.e., they disagree with any other segments in terms of [cont]. [r] was treated as [+cont] and nasals as [−cont]. [h] was considered a voiceless fricative, not an approximant. Sonorants were assumed to be specified as [+voi] and thus agree with voiced obstruents in terms of [voi]. Appendix 2 provides a complete matrix of these feature specifications.

The prediction of the hypothesis presented in Sect. 2 is that similarity, as measured by the number of shared feature specifications, positively correlates with rhymability, as measured by O/E values. In calculating the correlation between these two factors, the linear order of consonant pairs—i.e., which line they were in—was ignored. Furthermore, segments whose total O values were less than 10 (e.g., [p^j], [n^j]) were excluded since their inclusion would result in too many zero O/E values. Additionally, I excluded the pairs whose O values were zero since most of these pairs contained a member which was rare to begin with; for example, [p^j] and [m^j] occurred only twice in the corpus. Consequently, 199 consonantal pairs entered into the subsequent analysis.

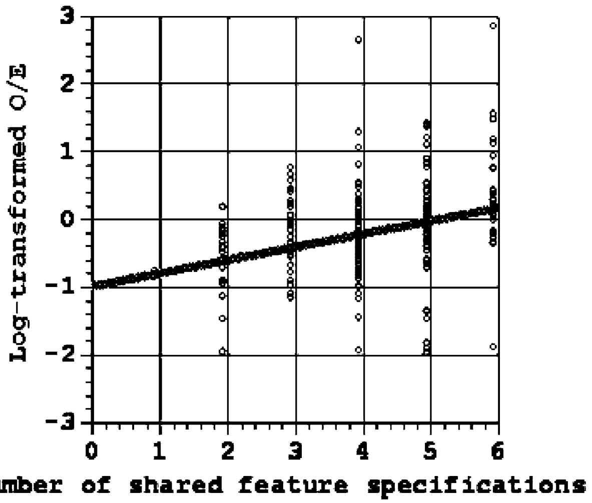

Fig. 1 The correlation between the number of shared feature specifications (similarity) and the O/E values (rhymability). The line represents a linear regression line

Figure 1 illustrates the overall result, plotting on the *y*-axis the natural log-transformed O/E values[7] for different numbers of shared feature specifications. The line represents an OLS regression line. Figure 1 supports the claim that there is a general correlation between similarity and rhymability: as the number of shared featural specifications increases, so do the O/E values.

To statistically analyze the linear relationship observed in Fig. 1, the Spearman correlation coefficient r_s, a non-parametric numerical indication of the strength of linear correlation, was calculated. The result revealed that r_s is .34, which significantly differs from zero ($p < .001$, two-tailed): similarity and rhymability do indeed correlate with each other.

Figure 1 shows that the scale of rhymability is gradient rather than categorical. It is not the case that pairs that are more dissimilar than a certain threshold are completely banned; rather, dissimilar consonant pairs are just less likely to rhyme than pairs that are more similar to one another. This gradiency differs from what we observe in other poetry traditions, in which there is a similarity threshold beyond which too dissimilar rhyme pairs are categorically rejected (see Holtman 1996; Zwicky 1976 for English; Steriade 2003 for Romanian). The source of the cross-linguistic difference is not entirely clear—perhaps it is related to the folk definition of Japanese rhymes that consonants are "ignored." This issue merits future cross-linguistic investigation. Although the gradient pattern in Fig. 1 seems specific to the Japanese rap rhyme patterns, it parallels the claim by Frisch (1996), Frisch, Pierrehumbert, and Broe (2004), and Stemberger (1991, pp. 78–79) that similarity is inherently a gradient notion.

In order to ensure that the correlation observed in Fig. 1 does not depend on a particular model of similarity chosen here, similarity was also measured based on the

[7] It is conventional in psycholinguistic work to use log-transformed value as a measure of frequency (Coleman and Pierrehumbert 1997; Hay et al. 2003). People's knowledge about lexical frequencies is better captured as log-transformed frequencies than raw frequencies (Rubin 1976; Smith and Dixon 1971). The log-transformation also increases normality of the distribution of the O/E values, which is useful in some of the statistical tests in the following sections.

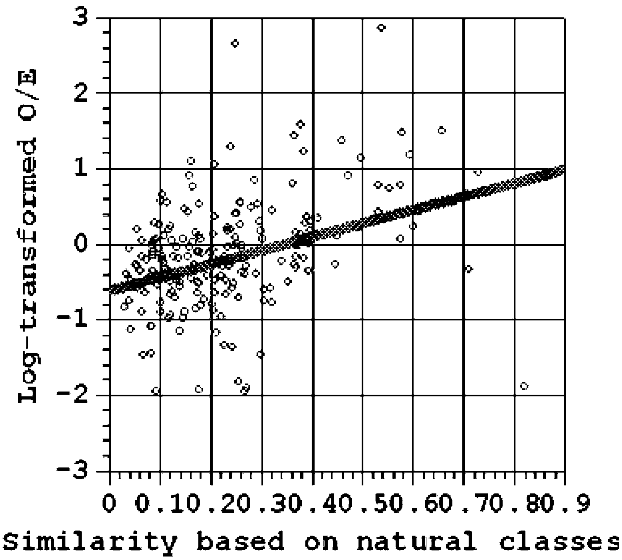

Fig. 2 The correlation between rhymability and similarity calculated based on the number of shared natural classes. The line represents a linear regression line

number of shared natural classes, following Frisch (1996) and Frisch et al. (2004). In their model, similarity is defined by the equation in (5):

$$(5) \qquad \text{Similarity} = \frac{\textit{shared natural classes}}{\textit{shared natural classes} + \textit{non-shared natural classes}}$$

A similarity matrix was computed based on (5), using a program written by Adam Albright (available for download from http://web.mit.edu/albright/www/). The scatterplot in Fig. 2 illustrates how this type of similarity correlates with rhymability. It plots the natural log-transformed O/E values against the similarity values calculated by (5).

Again, there is a general trend in which rhymability positively correlates with similarity as shown by the regression line with a positive slope. With this model of similarity, the Spearman correlation r_s is .37, which is again significant at .001 level (two-tailed).

Although this model yielded a higher r_s value than the previous model (.34), we must exercise caution in comparing these values—r_s is sensitive to the variability of the independent variable, which depends on the number of ties (Myers and Well 2003, p. 508; see also pp. 481–485). However, there are many more ties in the first model than in the second model, and therefore the goodness of fit cannot be directly compared between the two models.

Finally, one implicit assumption underlying Figs. 1 and 2 is that each feature contributes to similarity by the same amount. This assumption might be too naive, and Sects. 3.4 and 4.1 re-examine this assumption.

In summary, the analyses based on the two ways of measuring similarity both confirm the claim that there is a general positive correlation between similarity and rhymability. The following two subsections show that for each feature the pairs that share the same specification are overrepresented.

Table 2 The effects of major class features on rhymability

(i) [son]

C2 / C1	Sonorant	Obstruent
Sonorant	O=375 O/E=1.66	O=630 O/E=.81
Obstruent	O=666 O/E=.82	O=2960 O/E=1.05

(ii) [cons]

C2 / C1	Glide	Non-glide
Glide	O=8 O/E=1.69	O=154 O/E=.98
Non-glide	O=131 O/E=.98	O=4338 O/E=1.00

3.2 Major class features and manner features

We have witnessed a general correlation between similarity and rhymability. Building on this result, I now demonstrate that the correlation is not only a general tendency but that all of the individual features in fact contribute to the similarity.

Tables 2 and 3 illustrate how the agreement in major class and manner features contributes to overrepresentation. In these tables, C1 and C2 represent the first and second consonant in a rhyme pair. Unlike the analysis in Sect. 3.1, the order of C1 and C2 is preserved, but this is for statistical reasons and is not crucial for the claim made in this subsection: the O/E values of the {+F, −F} pair and those of the {−F, +F} pairs are always nearly identical. The statistical significance of the effect of each feature was tested by a Fisher's Exact test.

One final note: there were many pairs of identical consonants, partly because rappers very often rhyme using two identical sound sequences, as in ***saa kitto*** 'then probably' vs. ***saakitto*** 'circuit' (by *DELI* in *Chika-Chika Circuit*). It was expected, then, that if we calculated similarity effects using this dataset, similarity effects

Table 3 The effects of the manner features on rhymability

(i) [cont]

C2 / C1	Continuant	Non-cont
Continuant	O=598 O/E=1.27	O=804 O/E=.86
Non-cont	O=781 O/E=.86	O=1928 O/E=1.07

(ii) [nas]

C2 / C1	Nasal	Oral
Nasal	O=203 O/E=1.85	O=541 O/E=.85
Oral	O=480 O/E=.84	O=3413 O/E=1.03

(iii) [pal]

C2 / C1	Palatal	Non-pal
Palatal	O=185 O/E=2.76	O=313 O/E=.73
Non-pal	O=438 O/E=.79	O=3695 O/E=1.03

(iv) [voi]

C2 / C1	Voiced	Voiceless
Voiced	O=315 O/E=1.25	O=477 O/E=.88
Voiceless	O=496 O/E=.89	O=1252 O/E=1.05

would be obtained simply due to the large number of identical consonant pairs. To avoid this problem, their O values were replaced with the corresponding E values, which would cause neither underrepresentation nor overrepresentation.

Table 2(i) illustrates the effect of [son]: pairs of two sonorants occur 1.66 times more often than expected. Similarly, the {obs-obs} pairs are overrepresented though only to a small extent (O/E = 1.05). The degree of overrepresentation due to the agreement in [son] is statistically significant ($p < .001$). Table 2(ii) illustrates the effect of [cons], which distinguishes glides from non-glides. Although the {glide-glide} pairs are overrepresented (O/E = 1.69), the effect is not statistically significant ($p = .15$). This insignificance is due presumably to the small number of {glide-glide} pairs (O = 8).

Table 3(i) shows that pairs of two continuants (fricatives and approximants) are overrepresented (O/E = 1.27), showing the positive effect of [cont] on rhymability ($p < .001$). Next, as seen in Table 3(ii), the {nasal-nasal} pairs are also overrepresented (O/E = 1.85; $p < .001$). Third, Table 3(iii) shows the effect of [pal] across all places of articulation, which distinguishes palatalized consonants from non-palatalized consonants. Its effect is very strong in that two palatalized consonants occur almost three times as frequently as expected ($p < .001$). Finally, Table 3(iv) shows the effect of [voi] within obstruents, which demonstrates that voiced obstruents are more frequently paired with voiced obstruents than with voiceless obstruents (O/E = 1.25; $p < .001$). In a nutshell, the agreement in any feature results in higher rhymability. All of the overrepresentation effects, modulo that of [cons], are statistically significant at .001 level.

Overall, in both Tables 2 and 3, the {−F, −F} pairs (=the bottom right cells in each table) are overrepresented but only to a small extent. This slight overrepresentation of the {−F, −F} pairs suggests that the two pairs that share [−F] specifications are not considered very similar—the shared negative feature specifications do not contribute to similarity as much as the shared positive feature specifications (see Stemberger 1991 for similar observations).

We have observed in Table 3(iv) that agreement in [voi] increases rhymability within obstruents. In addition, sonorants, which are inherently voiced, rhyme more frequently with voiced obstruents than with voiceless obstruents. Consider Table 4.

As seen in Table 4, the probability of voiced obstruents rhyming with sonorants (.29; s.e. = .01) is much higher than the probability of voiceless obstruents rhyming with sonorants (.14; s.e. = .004). The difference between these probabilities (ratios) is statistically significant (by approximation to a normal distribution, $z = 13.42, p < .001$). In short, sonorants are more rhymable with voiced obstruents than with voiceless obstruents,[8] and this observation implies that voicing in sonorants promotes similarity with voiced obstruents. (See Coetzee and Pater 2005; Côté 2004; Frisch et al. 2004; Stemberger 1991, pp. 94–96; Rose and Walker 2004 and Walker 2003, for evidence that obstruent voicing promotes similarity with sonorants in other languages.)

This patterning makes sense from an acoustic point of view. First of all, low frequency energy is present during constriction in both sonorants and voiced obstruents but not in voiceless obstruents. Second, Japanese voiced stops are spirantized intervocalically, resulting in clear formant continuity (Kawahara 2006). For these reasons, voiced obstruents are acoustically more similar to sonorants than voiceless obstruents are.

[8] There are no substantial differences among three types of sonorants: nasals, liquids, and glides. The O/E values for each of these classes with voiced obstruents are .77, .78, .82, respectively. The O/E values with voiceless obstruents are .49, .44, .43, respectively.

Table 4 The rhyming pattern of sonorants with voiced obstruents and voiceless obstruents

	[+voi] obstruent	[-voi] obstruent	Sum
[+son]	599 (29.1%)	903 (14.4%)	1502
[-son]	1457 (70.9%)	5355 (85.6%)	6812
Sum	2056	6258	8314

However, given the behavior of [voi] in Japanese phonology, the similarity between voiced obstruents and sonorants is surprising. Phonologically speaking, voicing in Japanese sonorants is famously inert: OCP(+voi) requires that there be no more than one ‘‘voiced segment’’ within a stem, but only voiced obstruents, not voiced sonorants, count as ‘‘voiced segments’’ (Itô and Mester 1986; Itô, Mester, and Padgett 1995; Mester and Itô 1989).[9] The pattern in Table 4 suggests that, despite the phonological inertness of [+voice] in Japanese sonorants, Japanese speakers know that sonorants are more similar to voiced obstruents than to voiceless obstruents. This conclusion implies that speakers must be sensitive to acoustic similarity between voiced obstruents and sonorants or at least know that redundant and phonologically inert features can contribute to similarity.

3.3 Place homorganicity

In order to investigate the effect of place homorganicity on rhymability, consonants were classified into five groups: labial, coronal obstruent, coronal sonorant, dorsal, and pharyngeal. This grouping is based on four major place classes, but the coronal class is further divided into two subclasses. This classification follows the previous studies on consonant co-occurrence patterns, which have shown that coronal sonorants and coronal obstruents constitute separate classes, as in Arabic (Frisch et al. 2004; Greenberg 1950; McCarthy 1986, 1988), English (Berkley 1994), Javanese (Mester 1986), Russian (Padgett 1992), Wintu (McGarrity 1999), and most importantly in this context, Japanese (Kawahara et al. 2006). The rhyming patterns of the five classes are summarized in Table 5.

Table 5 shows that two consonants from the same class are all overrepresented. To check the statistical significance of the overrepresentation pattern in Table 5, a non-parametric sign-test was employed since there were only five data points.[10] As a null hypothesis, let us assume that each homorganic pair is overrepresented by chance (i.e., its probability is ½). The probability of the five pairs all being overrepresented is then $(½)^5 = .03$. This situation therefore is unlikely to have arisen by chance, or in other words, we can reject the null hypothesis at $\alpha = .05$ level that the homorganic pairs are overrepresented simply by chance.

[9] One might wonder whether [+voi] in nasals is active in Japanese phonology since nasals cause voicing of following obstruents, as in /sin+ta/ → [sinda] ‘died’, which can be captured as assimilation of [+voi] (Itô et al. 1995). However, post-nasal voicing has been reanalyzed as the effect of the constraint $*N\mathring{C}$, which prohibits voiceless obstruents after a nasal. This analysis has a broader empirical coverage (Hayes and Stivers 1995; Pater 1999) and obviates the need to assume that [+voi] associated with nasals is active in Japanese phonology (Hayashi and Iverson 1998).

[10] An alternative analysis would be to apply a χ^2-test for each homorganic pair (Kawahara et al. 2006; Mester 1986). This approach is not taken here because the O/E values of the homorganic cells are not independent of one another, and hence multiple applications of a χ^2-test are not desirable.

Table 5 The effects of place homorganicity on rhymability

C2 / C1	Labial (p, pʲ, b, bʲ, m, mʲ, Φ, w)	Cor-obs (t, d, ts, s, z, ʃ, ʒ, tʃ, dʒ)	Cor-son (n, nʲ, r, rʲ, j)	Dorsal (k, kʲ, g, gʲ)	Pharyngeal (h, hʲ)
Labial	O=181 O/E=1.50	O=213 O/E=.76	O=183 O/E=1.14	O=152 O/E=.88	O=32 O/E=1.19
Cor-obs	O=191 O/E=.73	O=816 O/E=1.34	O=275 O/E=.79	O=310 O/E=.83	O=53 O/E=.91
Cor-son	O=178 O/E=1.13	O=286 O/E=.78	O=352 O/E=1.68	O=157 O/E=.69	O=24 O/E=.68
Dorsal	O=149 O/E=.89	O=335 O/E=.86	O=142 O/E=.64	O=390 O/E=1.62	O=43 O/E=1.14
Phary	O=33 O/E=1.23	O=61 O/E=.97	O=22 O/E=.62	O=41 O/E=1.07	O=12 O/E=2.06

On the other hand, most non-homorganic pairs in Table 5 are underrepresented. One exception is the {labial-pharyngeal} pairs (O/E = 1.19, 1.23). The overrepresentation pattern can presumably be explained by the fact that Japanese pharyngeals are historically—and arguably underlyingly—labial (McCawley 1968, pp. 77–79; Ueda 1898). Therefore for Japanese speakers these two classes are phonologically related, and this phonological connection may promote phonological similarity between the two classes of sounds (see Malone 1988b; Shinohara 2004 for related discussion of this issue; cf. Steriade 2003). There might be an influence of orthography as well since [h], [p], and [b] are all written by the same letters, with the only distinction made by way of diacritics.

The {dorsal-pharyngeal} pairs are also overrepresented (O/E = 1.14, 1.07), and this pattern makes sense acoustically; [h] resembles [k, g] in being spectrally similar to adjacent vowels.[11] [k, g] extensively coarticulate with adjacent vowels for tongue backness because there is an extensive spatial overlap of gestures of vowels and dorsals. As a result, dorsals are spectrally similar to the adjacent vowels (Keating 1996, p. 54; Keating and Lahiri 1993; Tabain and Butcher 1999). As for [h], the shape of the vocal tract during the production of [h] is simply the interpolation of the surrounding vowels (Keating 1988), and thus the neighboring vowels' oral articulation determines the distribution of energy in the spectrum during [h]. Given the acoustic affinity between

[11] The similarity between [h] and dorsal consonants is further evidenced by the fact that Sanskrit [h] was borrowed as [k] in Old Japanese; e.g., Sanskrit *Maha* 'great' and *Arahân* 'saint' were borrowed as [ma**k**a] and [ara**k**an], respectively (Ueda 1898).

Table 6 Consonant co-occurrence restrictions on pairs of adjacent consonants in Yamato Japanese (based on Kawahara et al., 2006)

C1 \ C2	Labial (p, b, m, Φ, w)	Cor-obs (t, d, ts, s, z, ʃ, ʒ, tʃ, dʒ, ç)	Cor-son (n, r)	Dorsal (k, g)	Pharyngeal (h)
Labial	O=43 O/E=.22	O=374 O/E=1.30	O=312 O/E=1.19	O=208 O/E=1.02	O=0 O/E=.00
Cor-obs	O=434 O/E=1.40	O=247 O/E=.53	O=409 O/E=.97	O=445 O/E=1.35	O=2 O/E=.76
Cor-son	O=132 O/E=1.20	O=182 O/E=1.10	O=69 O/E=.46	O=164 O/E=1.40	O=2 O/E=2.14
Dorsal	O=242 O/E=1.12	O=389 O/E=1.20	O=369 O/E=1.25	O=66 O/E=.29	O=3 O/E=1.63
Phary	O=33 O/E=.60	O= 119 O/E=1.45	O=59 O/E=.79	O=54 O/E=.93	O=0 O/E=.00

[h] and [k, g], the fact that these two groups of consonants are overrepresented provides a further piece of evidence that rhymability is based on similarity.[12]

Another noteworthy finding concerning the effect of place on rhymability is that the coronal consonants split into two classes, i.e., the {cor son-cor obs} pairs are not overrepresented (O/E = .79, .78). This finding parallels Kawahara et al. (2006) finding. They investigate the consonant co-occurrence patterns of adjacent consonant pairs in the native vocabulary of Japanese (Yamato Japanese) and observe that adjacent pairs of homorganic consonants are systematically underrepresented; however, they also find that the {cor son-cor obs} pairs are not underrepresented, i.e., they are treated as two different classes. Kawahara et al. (2006) results are reproduced in Table 6 (the segmental inventory in Table 6 slightly differs from that of Table 5 because Table 6 is based on Yamato morphemes only).

This convergence—coronal sonorants and coronal obstruents constitute two separate classes both in the rap rhyming pattern and the consonant co-occurrence pattern—strongly suggests that the same mechanism (i.e., similarity) underlies both of the patterns. However, similarity manifests itself in opposite directions, i.e., in co-occurrence patterns similarity is avoided whereas in rap rhyming it is favored. In fact, there is a strong negative correlation between the values in Table 5 and the values in Table 6 ($r = -.79$, $t(23) = -5.09$, $p < .001$, two-tailed).

[12] The {labial-cor son} pairs are overrepresented as well (O/E = 1.14, 1.13). I do not have a good explanation for this pattern.

3.4 Multiple regression analysis

We have seen that for all of the features (the weak effect of [cons] notwithstanding), two consonants that share the same feature specification are overrepresented to a statistically significant degree. This overrepresentation suggests that all of the features contribute to similarity to some degree. A multiple regression analysis can address how much agreement in each feature contributes to similarity. Multiple regression yields an equation that best predicts the values of a dependent variable Y by p-number of predictors X by positing a model, $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2, \ldots, \beta_p X_p + \epsilon$ (where ϵ stands for a random error component). For the case at hand, the independent variables are the seven features used in Sect. 3.1–3.3; each feature is dummy-coded as 1 if two consonants agree in that feature, and 0 otherwise. Based on the conclusion in Sect. 3.3, coronal sonorants and coronal obstruents were assumed to form separate classes. For the dependent variable Y, I used the natural-log transformed O/E values. This transformation made the effect of outliers less influential and also made the distribution of the O/E values more normal. Finally, I employed the Enter method, in which all independent variables are entered in the model in a single step. This method avoids potential inflation of Type I error which the other methods might entail.

Regressing the log-transformed O/E on all features yielded the equation in (6).

(6) $\log_e(\mathrm{O/E})$
$-.94 + .62[\mathrm{pal}] + .25[\mathrm{voi}] + .23[\mathrm{nas}] + .21[\mathrm{cont}] + .11[\mathrm{son}]$
$- .008[\mathrm{cons}] - .12[\mathrm{place}]$
(where for each [f], [f] = 1 if a pair agrees in [f], and 0 otherwise)

The equation (6) indicates, for example, that agreeing in [pal] increases $\log_e(\mathrm{O/E})$ by .62, everything else being held constant. The slope coefficients (b_p) thus represent how much each of the features contributes to similarity. In (6), the slope coefficients are significantly different from zero for [pal], [voi], and [cont] and marginally so for [nas] as summarized in Table 7. R^2, a proportion of the variability of $\log_e(\mathrm{O/E})$ accounted for by (6), is .24 ($F(7, 190) = 8.70$, $p < .001$; R^2_{adj} is .21).[13]

The b_p values in Table 7 show that the manner features ([pal, voi, nas, cont]) contribute the most to similarity. The major class features ([cons] and [son]) do not contribute as much to similarity, and the coefficient for [place] does not significantly differ from zero either.

The lack of a significant effect of major features and [place] appears to conflict with what we observed in Sects. 3.2 and 3.3; [son], [cons], and [place] do cause overrepresentation when two consonants agree in these features. Presumably, the variability due to these features overlaps too much with the variability due to the four manner features, and as a result, the effect of [place] and the major class features per se may not have been large enough to have been detected by multiple regression. However, that [place] has a weaker impact on perceptual similarity compared to manner features is compatible with the claim that manner features are

[13] All assumptions for a multiple regression analysis were met reasonably well. The Durbin–Watson statistic was 1.65; the largest Cook's distance was .08; no leverage values were larger than the criterion ($=2(p + 1)/N = .08$); the smallest tolerance (=that of [son]) is .722. See, e.g., Myers and Well (2003, pp. 534–548, pp. 584–585) for relevant discussion of these values.

Table 7 The results of the multiple regression analysis

	b_p	s.e.(b_p)	t(192)	p (two-tailed)
Intercept	-.94	.20	-4.74	<.001
[pal]	.62	.09	6.60	<.001
[voi]	.25	.09	2.66	<.01
[nas]	.23	.14	1.73	.09
[cont]	.21	.10	2.17	.03
[son]	.11	.11	1.03	.31
[cons]	-.008	.14	-.55	.59
[place]	-.12	.11	-1.17	.25

more perceptually robust than place features as revealed by similarity judgment experiments (Bailey and Hahn 2005, p. 352; Peters 1963; Walden and Montgomery 1975) and identification experiments (Benkí 2003; Miller and Nicely 1955; Wang and Bilger 1973).[14]

To summarize, the multiple regression analysis has revealed the degree to which the agreement of each feature contributes to similarity: (i) [pal] has a fairly large effect, (ii) [cont], [voi], and [nas] have a medium effect, and (iii) [son], [cons] and [place] have too weak an effect to be detected by multiple regression. This result predicts that [pal] should have a stronger impact on similarity than other manner features, and this prediction should be verified by a perceptual experiment in a future investigation (see also Sect. 4.1 for some relevant discussion).

4 Discussion: acoustic similarity and rhymability

In Sect. 3, distinctive features were used to estimate similarity. I argue in this section that measuring similarity in terms of the sum of shared feature specifications or the number of shared natural classes does not suffice; Japanese speakers actually take acoustic details into account when they compose rhymes. I present three kinds of arguments for this claim. First, there are acoustically similar pairs of consonants whose similarity cannot be captured in terms of features (Sect. 4.1). Second, I show that the perceptual salience of features depends on its context (Sect. 4.2). Third, I show that the perceptual salience of different features is not the same, as anticipated in the multiple regression analysis (Sect. 4.3).

4.1 Similarity not captured by features

First of all, there are pairs of consonants which are acoustically similar but whose similarity cannot be captured in terms of features. We have already seen in Sect. 3.3 that the {dorsal-pharyngeal} pairs are more likely to rhyme than expected, and I argued that this overrepresentation is due to the spectral similarity of these pairs of

[14] In Tuareg half-rhymes, a difference in [place] can be systematically ignored (Greenberg 1960); this pattern also accords with the weak effect of [place] identified here.

consonants; note that there is nothing in the feature system that captures the similarity between dorsals and pharyngeals.[15]

Another example of this kind is the {k^j-tʃ} pair, which is highly overrepresented (O/E = 3.91): concrete rhyme examples include ***ch**ooshi* 'condition' vs. ***ky**oomi* 'interest' (by *565* in *Tokyo Deadman Walking*), *kai**ch**oo* 'smooth' vs. *kan**ky**oo* 'environment' and (*kan*)***ky**oo* 'environment' vs. (*sei*)***ch**oo* 'growth' (by *XBS* in *64 Bars Relay*), and *gozen**ch**uu* 'morning' vs. *goren**ky**uu* 'five holidays in a row' (by *TSURUGI MOMOTAROU* in *Project Mou*). Similarly, the {k^j-dʒ} pair is also highly overrepresented (O/E = 3.39), e.g., *too**j**oo* 'appearance' vs. *too**ky**oo* 'Tokyo' (by *RINO LATINA* in *Shougen*), ***ky**uudai* 'stop conversing' vs. ***j**uutai* 'serious injury' (by *DABO* in *Tsurezuregusa*), and *nen**j**uu* 'all year' vs. *sen**ky**uu* 'Thank you' (by *TSURUGI MOMOTAROU* in *Project Mou*).

These overrepresentation patterns are understandable in light of the fact that [k^j] is acoustically very similar to [tʃ] and to a lesser extent, to [dʒ] (Blumstein 1986; Chang, Plauché, and Ohala 2001; Guion 1998; Keating and Lahiri 1993; Ohala 1989). First, velar consonants have longer aspiration than non-velar consonants (Hirata and Whiton 2005, p. 1654; Maddieson 1997, p. 631), and their long aspiration acoustically resembles affrication. Second, [k^j]'s F2 elevation due to palatalization makes [k^j] sound similar to coronals, which are also characterized by high F2.

The similarity of these pairs cannot be captured in terms of distinctive features: the {k^j-dʒ} pair and the {k^j-tʃ} pair agree in only three and four distinctive features, respectively, but the average O/E values of such pairs is .87 and 1.05. An anonymous reviewer has pointed out that all of these consonants contain a [+palatalized] feature which plays a morphological role in onomatopoetic words (Hamano 1998; Mester and Itô 1989), and that the Japanese kana orthography has a special symbol for palatalization; (s)he suggests that these morphological and orthographic factors might promote the similarity of the {k^j-dʒ} and {k^j-tʃ} pairs. This may indeed be the case, but agreement in [palatalized] does not fully explain why we observe high O/E values for the {k^j-tʃ} and {k^j-dʒ} pairs. As revealed by the multiple regression analysis, the contribution to similarity by [+palatalized] is $e^{.62} = 1.86$, which does not by itself explain the high O/E values. The high rhymability of these pairs thus instantiates yet another case in which similarity cannot be measured solely in terms of distinctive features—Japanese speakers have awareness for detailed acoustic information such as the duration of aspiration and the raising of F2 caused by palatalization.

4.2 The perceptual salience of a feature depends on its context

The second piece of evidence that computations of similarity must take acoustic details into account is the fact that the salience of the same feature can differ depending on its context. Here, evidence is put forth in terms of the perceptual salience of [place].

Cross-linguistically, nasals are more prone to place assimilation than oral consonants (Cho 1990; Jun 1995; Mohanan 1993). Jun (1995) argues that this is because place cues are more salient for oral consonants than for nasal consonants.

[15] No recent models of distinctive features have a feature that distinguishes {dorsal, pharyngeal} from other places of articulation (Clements and Hume 1995; McCarthy 1988; Sagey 1986). *The Sound Pattern of English* characterizes velars, uvulars, pharyngeals as [+back] consonants (Chomsky and Halle 1968, p. 305), but subsequent research abandoned this consonantal feature since these consonants do not form a natural class in phonological patterns.

Boersma (1998, p. 206) also notes that ‘‘[m]easurements of the spectra … agree with confusion experiments (for Dutch: Pols 1983), and with everyday experience, on the fact that [m] and [n] are acoustically very similar, and [p] and [t] are farther apart. Thus, place information is less distinctive for nasals than it is for plosives.’’ Ohala and Ohala (1993, pp. 241–242) likewise suggest that ‘‘[nasal consonants’] place cues are less salient than those for comparable obstruents.’’ The reason why nasals have weaker place cues than oral stops is that formant transitions into and out of the neighboring vowels are obscured due to coarticulatory nasalization (Jun 1995, building on Malécot 1956; see also Hura et al. 1992, p. 63; Mohr and Wang 1968).

The fact that a place distinction is less salient for nasals than it is for obstruents is reflected in the rhyming pattern. The {m-n} pair is the only minimal pair of nasals that differs in place,[16] and its O/E value is 1.41, which is a rather strong overrepresentation. Indeed the rhyme pairs involving the [m]~[n] pair are very common in Japanese rap rhymes: *yonaka* ‘midnight’ vs. *omata* ‘crotch’ (by *AKEEM* in *Forever*), *kaname* ‘core’ vs. *damare* ‘shut up’ (by *KOHEI JAPAN* in *Go to Work)*, and *hajimari* ‘beginning’ vs. *tashinami* ‘experience’ (by *YOSHI* in *Kin Kakushi kudaki Tsukamu Ougon*).[17]

Minimal pairs of oral consonants differing in place are less common than the {m-n} pair: all minimal pairs of oral consonants show O/E values smaller than 1.41—{p-t}: 1.09, {p-k}: 1.07, {t-k}: .93, {b-d}: 1.25, {b-g}: 1.25, {d-g}: 1.36. To statistically validate this conclusion, a sign-test was performed: the fact that the nasal minimal pair has an O/E value larger than all oral minimal pairs is unlikely to have arisen by chance ($p < .05$).

We may recall that O/E values correlate with similarity, and therefore this result shows that the {m-n} pair is more similar than any minimal pairs of oral consonants that differ in place. This finding in turn implies that place cues are less salient for nasals than for oral consonants. To conclude, then, in measuring (dis-)similarity due to [place], Japanese speakers take into account the fact that [place] distinctions in nasal consonants are weak, due to the blurring of formant transitions caused by coarticulatory nasalization. This finding constitutes yet another piece of evidence that Japanese speakers use their knowledge about acoustic details in composing rap rhymes.

4.3 The saliency of features is not homogeneous

The final argument that similarity cannot be measured solely in terms of distinctive features rests on the fact that some features are more perceptually salient than other features, as suggested previously in Sect. 3.4 (see also Bailey and Han 2005; Benkí 2003; Miller and Nicely 1955; Mohr and Wang 1968; Peters 1963; Walden and Montgomery 1975; Wang and Bilger 1973). What concerns us here is the

[16] The other nasal segment in Japanese (a so-called moraic nasal, [N]) appears only in codas and hence was excluded from this study.

[17] Nasals pairs with disagreeing place specifications are common in English rock lyrics (Zwicky 1976), English imperfect puns (Zwicky and Zwicky 1986), and English and German poetry (Maher 1969, 1972).

perceptibility of [voi]. Some recent proposals hypothesize that [voi] only weakly contributes to consonant distinctions (cf. Coetzee and Pater 2005). Steriade (2001b) for example notes that a [voi] distinction is very often sacrificed to satisfy markedness requirements (e.g., a prohibition against coda voiced obstruents) and argues that frequent neutralization of [voi] owes to its low perceptibility. Kenstowicz (2003) observes that in loanword adaptation, when the recipient language has only a voiceless obstruent and a nasal, the voiced stops of the donor language invariably map to the former, i.e., we find [d] borrowed as [t] but not as [n] (see also Adler 2006 for a similar observation regarding Hawaiian). Furthermore, Kawahara et al. (2006) found that in Yamato Japanese, pairs of adjacent consonants that are minimally different in [voi] behave like pairs of adjacent identical consonants in that they are not subject to co-occurrence restrictions. Kawahara et al. (2006) hypothesize that pairs of consonants that differ only in [voi] do not differ enough to be regarded as non-identical, at least for the purpose of co-occurrence restriction patterns.

These phonological observations accord with previous psycholinguistic findings as well: the Multi-Dimensional Scaling (MDS) analyses of Peters (1963) and Walden and Montgomery (1975) revealed that voicing plays a less important role in distinguishing consonants than other manner features. Bailey and Han (2005, p. 352) found that manner features carry more weight than voicing in listener's similarity judgments. Wang and Bilger's (1973) confusion experiment in a noisy environment revealed the same tendency. In short, the psychoacoustic salience of [voi] is weaker than that of other manner features.

With these observations in mind, we expect that pairs of consonants that are minimally different in [voi] should show greater rhymability than minimal pairs that disagree in other manner features. The multiple regression analysis in Sect. 3.4 revealed that [voi], [cont], and [nas] had a fairly comparable effect on rhymability, but it is not possible to statistically compare the slope coefficients of these features since all of the estimates are calculated on the basis of one sample (the estimates are not independent of one another). Therefore, in order to compare the effect of [voi], [cont], and [nas], I compare the O/E values of the minimal pairs defined in terms of these features. The data are shown in (7).[18]

(7)	[cont]		[nasal]		[voi]	
	{p-Φ}:	1.44	{b-m}:	1.30	{p-b}:	1.98
	{t-s}:	1.01	{d-n}:	1.41	{t-d}:	2.04
	{d-z}:	1.19			{tʃ-dʒ}:	4.11
					{k-g}:	1.44
					{k^j-g^j}:	16.5
					{s-z}:	3.07

[18] The effect of the [palatalized] feature is not consistent (e.g., {p-p^j}: 0.00, {n}-{n^j}: 3.10, {r-r^j}: 0.78, {k-k^j}: 0.71, {g-g^j}: 1.76). Hence it is difficult to make a consistent comparison between the [voi] and [palatalized] features; however, we may recall that the multiple regression analysis shows that [palatalized] has an outstandingly strong effect on similarity.

The O/E values of the minimal pairs that differ in [voi] are significantly larger than the O/E values of the minimal pairs that differ in [cont] or [nasal] (by a non-parametric Wilcoxon Rank-Sum (Mann–Whitney) test: Wilcoxon $W = 15.5$, $z = 2.65$, $p < .01$).[19] Indeed even a cursory survey reveals that the minimal pairs differing in [voi] are extremely common in Japanese rap songs, e.g., *toogarashi* 'red pepper' vs. *tookaraji* 'not too far' (by *NANORUNAMONAI* in *Kaerimichi)*, (*i*)*ttokuze* 'I'll say it' vs. *tokusen* 'special' (by *OSUMI* in *Don't Be Gone*), and (*shin*)*jitai* 'want to believe' vs. (*jibun*) *shidai* 'depend on me' (by *AI* in *Golden MIC*).

In summary, minimal pairs that differ only in [voi] are treated as quite similar to each other by Japanese speakers. This characteristic of [voi] in turn suggests that in measuring similarity, Japanese speakers do not treat [voi] and other manner features homogeneously, i.e., Japanese speakers have awareness of the varying perceptual salience of different features.

4.4 Summary: knowledge of similarity

I have argued that Japanese speakers take acoustic details into account when they measure similarity in composing rap rhymes. Given that the use of featural similarity does not provide adequate means to compute similarity in the rhyme patterns, we might dispense entirely with featural similarity in favor of (psycho)acoustic similarity. Ultimately, this move may be possible but not within the present analysis: it would require a complete similarity matrix of Japanese sounds, which must be obtained through extensive psycholinguistic experiments (e.g., identification experiments under noise, similarity judgment tasks, etc.; see Bailey and Hahn 2005 for a recent overview). Constructing the acoustic similarity matrix and using it for another analysis of the Japanese rap patterns are both interesting topics for future research topics but impossible goals for this paper.

The conclusion that speakers possess knowledge of similarity based on acoustic properties fits well with Steriade's (2001a, b, 2003) P-map hypothesis. She argues that language users possess detailed knowledge of psychoacoustic similarity, encoded in the grammatical component called the P-map. The P-map exists outside of a phonological grammar, but phonology deploys the information it stores (Fleischhacker 2001, 2005; Kawahara 2006; Zuraw 2005). The analysis of rap rhyme patterns has shown that speakers do possess such knowledge of similarity and deploy it in composing rhymes, and future research could reasonably concern itself with determining whether this knowledge of similarity interacts—or even shapes—phonological patterns.

5 Conclusion

The rhymability of two consonants positively correlates with their similarity. Furthermore, the knowledge of similarity brought to bear by Japanese speakers includes detailed acoustic information. These findings support the more general idea

[19] Voicing disagreement is much more common than nasality disagreement in other languages' poetry as well (for English, see Zwicky 1976; for Romanian, see Steriade 2003). Shinohara (2004) also points out that voicing disagreement is commonly found in Japanese imperfect puns, as in *Tokoya wa dokoya* 'where is the barber?' Zwicky and Zwicky (1986) make a similar observation in English imperfect puns.

that speakers possess rich knowledge of psychoacoustic similarity (Steriade 2001a, b, 2003 and earlier psycholinguistic work cited above).

One final point before closing: we cannot attribute the correlation between rhymability and similarity in Japanese rap songs to a conscious attempt by Japanese rap lyrists to make rhyming consonants as similar as possible. Crucially, the folk definition of rhymes says explicitly that consonants play no role in the formation of rhymes. Therefore, the patterns of similarity revealed in this study cannot be attributed to any conscious, conventionalized rules. Rather, the correlation between similarity and rhymability represents the spontaneous manifestation of unconscious knowledge of similarity.

The current study has shown that the investigation of rhyming patterns reveals such unconscious knowledge of similarity that Japanese speakers possess. Therefore, more broadly speaking, the current findings give support to the claim that investigation of paralinguistic patterns can contribute to the inquiry into our collective linguistic knowledge and should be undertaken in tandem with analyses of purely linguistic patterns (Bagemihl 1995; Fabb 1997; Fleischhacker 2005; Goldston 1994; Halle and Keyser 1971; Holtman 1996; Itô, Kitagawa, and Mester 1996; Jakobson 1960; Kiparsky 1970, 1972, 1981; Maher 1969, 1972; Malone 1987, 1988a, b; Minkova 2003; Ohala 1986; Shinohara 2004; Steriade 2003; Yip 1984, 1999; Zwicky 1976; Zwicky and Zwicky 1986 among many others).

Acknowledgements I am grateful to Michael Becker, Kathryn Flack, Becca Johnson, Teruyuki Kawamura, John Kingston, Daniel Mash, John McCarthy, Joe Pater, Caren Rotello, Kazuko Shinohara, Lisa Shiozaki, Donca Steriade, Kiyoshi Sudo, Kyoko Takano, and Richard Wright for their comments on various manifestations of this work. Portions of this work have been presented at the following occasions: UMass Phonology Reading Group (6/17/2005), Tokyo University of Agriculture and Technology (7/26/2005), University of Tokyo (8/15/2005), the LSA Annual Meeting at Albuquerque (1/06/2006), the University of Tromsø (3/21/2006), and the first MIT-UMass joint phonology meeting (MUMM; 5/6/2006). I would like to thank the audiences at these occasions. Comments from three anonymous reviewers of *JEAL* have substantially improved the contents and the presentation of the current work. All remaining errors and oversights are mine.

Appendix 1: Songs consulted for the analysis (alphabetically orderd)

The first names within the parentheses are artist (or group) names. Names after "feat." are those of guest performers.

1. 180 Do (LUNCH TIME SPEAX)
2. 64 Bars Relay (SHAKKAZOMBIE feat. DABO, XBS, and SUIKEN)
3. 8 Men (I-DEA feat. PRIMAL)
4. A Perfect Queen (ZEEBRA)
5. Akatsuki ni Omou (MISS MONDAY feat. SOWELU)
6. All Day All Night (XBS feat. TINA)
7. Area Area (OZROSAURUS)
8. Asparagus Sunshine (MURO feat. G.K.MARYAN and BOO)
9. Bed (DOUBLE feat. MUMMY-D and KOHEI JAPAN)
10. Bouhuuu (ORIGAMI)
11. Brasil Brasil (MACKA-CHIN and SUIKEN)
12. Breakpoint (DJ SACHIO feat. AKEEM, DABO, GOKU, K-BOMB, and ZEEBRA)
13. Brother Soul (KAMINARI-KAZOKU feat. DEN and DEV LARGE)

14. Bureemen (MABOROSHI feat. CHANNEL, K.I.N, and CUEZERO)
15. Change the Game (DJ OASIS feat. K-DUB SHINE, ZEEBRA, DOOZI-T, UZI, JA KAMIKAZE SAIMON, KAZ, TATSUMAKI-G, OJ, ST, and KM-MARKIT)
16. Chika-Chika Circit (DELI feat. KASHI DA HANDSOME, MACKA-CHIN, MIKRIS, and GORE-TEX)
17. Chiishinchu (RINO-LATINA II feat. UZI)
18. Chou to Hachi (SOUL SCREAM)
19. Crazy Crazy (DJ MASTERKEY feat. Q and OSUMI)
20. Do the Handsome (KIERU MAKYUU feat. KASHI DA HANDSOME)
21. Doku Doku (TOKONA-X, BIGZAM, S-WORD, and DABO)
22. Don't be Gone (SHAKKAZOMBIE)
23. Eldorado Throw Down (DEV LARGE, OSUMI, SUIKEN, and LUNCH TIME SPEAX)
24. Enter the Ware (RAPPAGARIYA)
25. Final Ans. (I-DEA feat. PHYS)
26. Forever (DJ TONK feat. AKEEM, LIBRO, HILL-THE-IQ, and CO-KEY)
27. Friday Night (DJ MASTERKEY feat. G.K. MARYAN)
28. Go to Work (KOHEI JAPAN)
29. Golden MIC—remix version (ZEEBRA feat. KASHI DA HANDSOME, AI, DOOZI-T, and HANNYA)
30. Good Life (SUITE CHIC feat. FIRSTKLAS)
31. Hakai to Saisei (KAN feat. RUMI and KEMUI)
32. Hakushu Kassai (DABO)
33. Happy Birthday (DJ TONK feat. UTAMARU)
34. Hard to Say (CRYSTAL KAY feat. SPHERE OF INFLUENCE and SORA3000)
35. Hazu Toiu Otoko (DJ HAZU feat. KENZAN)
36. Hima no Sugoshikata (SUCHADARA PARR)
37. Hip Hop Gentleman (DJ MASTERKEY feat. BAMBOO, MUMMY-D, and YAMADA-MAN)
38. Hi-Timez (HI-TIMEZ feat. JUJU)
39. Hitoyo no Bakansu (SOUL SCREAM)
40. Ikizama (DJ HAZU feat. KAN)
41. Itsunaroobaa (KICK THE CAN CREW)
42. Jidai Tokkyuu (FLICK)
43. Junkai (MSC)
44. Kanzen Shouri (DJ OASIS feat. K-DUB SHINE, DOOZI-T, and DIDI)
45. Kikichigai (DJ OASIS feat. UTAMARU and K-DUB SHINE)
46. Kin Kakushi Kudaki Tsukamu Ougon (DJ HAZIME feat. GAKI RANGER)
47. Kitchen Stadium (DJ OASIS feat. UZI and MIYOSHI/ZENZOO)
48. Koi wa Ootoma (DABO feat. HI-D)
49. Koko Tokyo (AQUARIUS feat. S-WORD, BIG-O, and DABO)
50. Konya wa Bugii Bakku (SUCHADARA PARR feat. OZAWA KENJI)
51. Koukai Shokei (KING GIDDORA feat. BOY-KEN)
52. Lesson (SHIMANO MOMOE feat. MUMMY-D)
53. Lightly (MAHA 25)

54. Lights, Camera, Action (RHYMESTER)
55. Mastermind (DJ HASEBE feat. MUMMY-D and ZEEBRA)
56. Microphone Pager (MICROPHONE PAGER)
57. Moshimo Musuko ga Umaretara (DJ MITSU THE BEATS feat. KOHEI JAPAN)
58. Mousouzoku DA Bakayarou (MOUSOUZOKU)
59. My Honey—DJ WATARAI remix (CO-KEY, LIBRO, and SHIMANO MOMOE)
60. My Way to the Stage (MURO feat. KASHI DA HANDSOME)
61. Nayame (ORIGAMI feat. KOKORO)
62. Nigero (SHIIMONEETAA & DJ TAKI-SHIT)
63. Nitro Microphone Underground (NITRO MICROPHONE UNDERGROUND)
64. Norainu (DJ HAZU feat. ILL BOSSTINO)
65. Noroshi—DJ TOMO remix (HUURINKAZAN feat. MACCHO)
66. Ohayou Nippon (HANNYA)
67. On the Incredible Tip (MURO feat. KASHI DA HANDSOME, JOE-CHO, and GORIKI)
68. One (RIP SLYME)
69. One Night (I-DEA feat. SWANKY SWIPE)
70. One Woman (I-DEA feat. RICE)
71. Ookega 3000 (BUDDHA BRAND feat. FUSION CORE)
72. Phone Call (DJ OASIS feat. BACKGAMMON)
73. Prisoner No1.2.3. (RHYMESTER)
74. Project Mou (MOUSOUZOKU)
75. R.L. II (RINO-LATINA II)
76. Rapgame (IQ feat. L-VOCAL, JAYER, and SAYAKO)
77. Revolution (DJ YUTAKA feat. SPHERE OF INFLUENCE)
78. Roman Sutoriimu (SIIMONEETAA and DJ TAKI-SHIT)
79. Shougen (LAMP EYE feat. YOU THE ROCK, G.K. MARYAN, ZEEBRA, TWIGY, and DEV LARGE)
80. Sun oil (ALPHA feat. UTAMARU and NAGASE TAKASHI)
81. Take ya Time (DJ MASTERKEY feat. HI-TIMEZ)
82. Tengoku to Jikoku (K-DUB SHINE)
83. Ten-un Ware ni Ari (BUDDHA BRAND)
84. The Perfect Vision (MINMI)
85. The Sun Rising (DJ HAZIME feat. LUNCH TIME SPEAX)
86. Tie Break ’80 (MACKA-CHIN)
87. Tokyo Deadman Walking (HANNYA feat. 565 and MACCHO)
88. Tomodachi (KETSUMEISHI)
89. Tomorrow Never Dies (ROCK-TEE feat. T.A.K. THE RHHHYME)
90. Toriko Rooru (MONTIEN)
91. Tsurezuregusa (DABO feat. HUNGER and MACCHO)
92. Ultimate Love Song(Letter) (I-DEA feat. MONEV MILS and KAN)
93. Unbalance (KICK THE CAN CREW)
94. Whoa (I-DEA feat. SEEDA)
95. Yakou Ressha (TWIGY feat. BOY-KEN)
96. Yamabiko 44-go (I-DEA feat. FLICK)

97. Yoru to Kaze (KETSUMEISHI)
98. Zero (DJ YUTAKA feat. HUURINKAZAN)

Appendix 2: Feature matrix

Table 8 Feature matrix. Affricates are assumed to disagree with any other segments in terms of [cont]

	son	cons	cont	nasal	voice	place	palatalized
p	-	+	-	-	-	lab	-
p^j	-	+	-	-	-	lab	+
b	-	+	-	-	+	lab	-
b^j	-	+	-	-	+	lab	+
Φ	-	+	+	-	-	lab	-
m	+	+	-	+	+	lab	-
m^j	+	+	-	+	+	lab	+
w	+	-	+	-	+	lab	-
t	-	+	-	-	-	cor	-
d	-	+	-	-	+	cor	-
s	-	+	+	-	-	cor	-
z	-	+	+	-	+	cor	-
ʃ	-	+	+	-	-	cor	+
ts	-	+	±	-	-	cor	-
tʃ	-	+	±	-	-	cor	+
dʒ	-	+	±	-	+	cor	+
n	+	+	-	+	+	cor	-
n^j	+	+	-	+	+	cor	+
r	+	+	+	-	+	cor	-
r^j	+	+	+	-	+	cor	+
j	+	-	+	-	+	cor	+
k	-	+	-	-	-	dors	-
k^j	-	+	-	-	-	dors	+
g	-	+	-	-	+	dors	-
g^j	-	+	-	-	+	dors	+
h	-	+	+	-	-	phary	-
h^j	-	+	+	-	-	phary	+

Appendix 3: A complete chart of the O/E values and O values

Table 9 A complete chart of the O/E values and O values

	p	pʲ	b	bʲ	Φ	m	mʲ	w	t	d	s	z	ʃ	ts	tʃ	dʒ	n	nʲ	r	rʲ	j	k	kʲ	g	gʲ	h	hʲ
p	5.65	0	1.98	0	1.44	.7	0	.86	1.09	.70	.91	.97	.88	.45	1.14	0.22	.32	0	.79	0	.63	1.07	2.32	1.02	0	1.62	0
	14	*0*	*15*	*0*	*1*	*8*	*0*	*2*	*17*	*6*	*10*	*4*	*7*	*1*	*5*	*1*	*4*	*0*	*14*	*0*	*2*	*26*	*3*	*8*	*0*	*9*	*0*
pʲ		0	0	0	0	0	0	0	0	0	0	0	0	0	17.8	0	0	0	0	0	24.9	0	0	0	0	0	0
		0	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*1*	*0*	*0*	*0*	*0*	*0*	*1*	*0*	*0*	*0*	*0*	*0*	*0*
b			*2.41*	0	.94	1.30	0	1.26	.52	1.25	.72	.95	.49	.29	.96	.51	.62	0	1.36	0	.72	.75	.76	1.25	2.82	1.59	2.07
			56	*0*	*2*	*46*	*0*	*9*	*25*	*33*	*24*	*12*	*12*	*2*	*13*	*7*	*24*	*0*	*74*	*0*	*7*	*56*	*3*	*30*	*3*	*27*	*1*
bʲ				398	0	0	0	2.81	0	0	2.87	0	0	0	0	0	0	0	1.76	0	9.95	0	0	0	0	5.67	0
				1	*0*	*0*	*0*	*1*	*0*	*0*	*1*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*1*	*0*	*1*	*0*	*0*	*0*	*0*	*1*	*0*
Φ					20.5	.62	0	1.53	.92	.42	.65	0	.90	1.59	1.62	0	0	0	1.20	0	1.13	.88	0	.45	0	4.51	22.6
					4	*2*	*0*	*1*	*4*	*1*	*2*	*0*	*2*	*1*	*2*	*0*	*0*	*0*	*6*	*0*	*1*	*6*	*0*	*1*	*0*	*7*	*1*
m						2.8	0	1.11	.66	1.00	.77	.73	.41	.38	.44	.52	1.41	0	1.00	0	.89	.91	.67	.74	1.24	.62	0
						150	*0*	*12*	*48*	*40*	*39*	*14*	*15*	*4*	*9*	*11*	*83*	*0*	*83*	*0*	*13*	*103*	*4*	*27*	*2*	*16*	*0*
mʲ							0	0	0	0	0	0	9.87	0	0	0	0	0	0	146	0	0	0	0	0	0	0
							0	*0*	*0*	*0*	*0*	*0*	*1*	*0*	*0*	*0*	*0*	*0*	*0*	*1*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
w								8.17	.82	.99	.48	0	0	0	0	.70	2.01	8.4	1.13	0	.67	.79	0	1.48	0	.57	0
								18	*12*	*8*	*5*	*0*	*0*	*0*	*0*	*3*	*24*	*1*	*19*	*0*	*2*	*18*	*0*	*11*	*0*	*3*	*0*
t									2.85	2.04	1.01	.69	.24	.14	.22	.39	.86	1.26	.64	.30	.86	.93	.49	.93	0	.63	1.01
									278	*110*	*69*	*18*	*12*	*2*	*6*	*11*	*68*	*1*	*71*	*1*	*17*	*142*	*4*	*46*	*0*	*22*	*1*
d										3.46	.92	1.19	.40	0	.79	.51	1.41	0	1.05	1.08	1.19	.80	.67	1.36	.83	1.04	0
										103	*35*	*17*	*11*	*0*	*12*	*8*	62	*0*	*65*	2	*13*	*67*	*3*	*37*	*1*	*20*	*0*
s											4.56	3.07	.80	4.14	.46	.55	.55	1.79	.73	.42	.57	.75	.35	1.10	.65	1.10	0
											220	*56*	*28*	*41*	*9*	*11*	*31*	*1*	*57*	*1*	*8*	*80*	*2*	*38*	*1*	*27*	*0*

Table 9 continued

z												7.10 *49*	.15 *2*	1.07 *4*	.14 *1*	.13 *1*	.57 *12*	0 *0*	.67 *20*	1.12 *1*	.57 *3*	.59 *24*	.93 *2*	1.22 *16*	0 *0*	.65 *6*	0 *0*
ʃ													7.20 *184*	.14 *1*	2.4 *34*	3.73 *54*	.37 *15*	2.47 *1*	.49 *28*	2.32 *4*	.79 *8*	.63 *49*	2.17 *9*	.32 *8*	2.69 *3*	.84 *15*	1.97 *1*
ts														16.3 *33*	.75 *3*	.24 *1*	0 *0*	0 *0*	.43 *7*	0 *0*	0 *0*	1.46 *32*	0 *0*	.70 *5*	0 *0*	1.00 *5*	0 *0*
tʃ															11.1 *87*	4.11 *33*	.13 *3*	0 *0*	.38 *12*	2.10 *2*	.36 *2*	.53 *23*	3.91 *9*	.43 *6*	1.62 *1*	.61 *6*	0 *0*
dʒ																6.81 *56*	1.00 *23*	0 *0*	.62 *20*	2.05 *2*	.52 *3*	.54 *24*	3.39 *8*	.35 *5*	3.16 *2*	.99 *10*	3.48 *1*
n																	3.06 *198*	3.10 *2*	1.17 *106*	0 *0*	1.43 *23*	.62 *77*	.30 *2*	.80 *32*	0 *0*	.46 *13*	0 *0*
nʲ																		0 *0*	1.10 *1*	36.6 *1*	0 *0*	0 *0*	0 *0*	0 *0*	0 *0*	0 *0*	0 *0*
r																			2.71 *347*	.78 *3*	.75 *17*	.58 *101*	.22 *2*	.87 *49*	.80 *2*	.58 *23*	0 *0*
rʲ																				77.4 *9*	2.93 *2*	.38 *2*	0 *0*	0 *0*	13.3 *1*	1.67 2	0 *0*
j																					9.95 *40*	.58 *18*	1.82 *3*	.80 *8*	0 *0*	.99 *7*	4.97 *1*
k																						2.11 *501*	.71 *9*	1.44 *111*	.59 2	1.16 63	1.94 *3*
kʲ																							14.8	.49	16.5	1.04	12.1

Table 9 continued

																							10	*2*	*3*	3	*1*
g																								1.76 *44*	0 *0*	.80 *14*	0 *0*
g^j																									20.5 *1*	0 *0*	0 *0*
h																										4.2 *52*	0 *0*
h^j																											0 0

The upper values in each cell represent O/E values. The lower values, shown in *italics*, represent O values

References

Adler, A. (2006). Faithfulness and perception in loanword adaptation: A case study from Hawaiian. *Lingua, 116*, 1021–1045.

Bagemihl, B. (1995). Language games and related areas. In J. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 697–712). Oxford: Blackwell.

Bailey, T., & Hahn, U. (2005). Phoneme similarity and confusability. *Journal of Memory and Language, 52*, 339–362.

Benkí, J. (2003). Analysis of English nonsense syllable recognition in noise. *Phonetica, 60*, 129–157.

Benua, L. (1997). *Transderivational identity: Phonological relations between words*. PhD dissertation, University of Massachusetts, Amherst.

Berkley, D. (1994). The OCP and gradient data. *Studies in the Linguistic Sciences. 24*, 59–72.

Blumstein, S. (1986). On acoustic invariance in speech. In J. Perkell & D. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 178–193). Hillsdale: Lawrence Erlbaum.

Boersma, P. (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.

Burzio, L. (2000). Segmental contrast meets output–to–output faithfulness. *The Linguistic Review,* 17, 367–384.

Chang, S., Plauché, M., & Ohala, J. (2001). Markedness and consonant confusion asymmetries. In E. Hume & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 79–101). San Diego: Academic Press.

Cho, Y.-M. (1990). *Parameters of consonantal assimilation*. PhD dissertation, Stanford University.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.

Clements, N., & Hume, E. (1995). The internal organization of speech sounds. In J. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 245–306). Oxford: Blackwell.

Coetzee, A., & Pater, J. (2005). Lexically gradient phonotactics in Muna and optimality theory, ms., University of Massachusetts, Amherst.

Coleman, J., & Pierrehumbert, J. (1997). Stochastic phonological grammars and acceptability. *Computational Phonology* (pp. 49–56). Somerset: Association for Computational Linguistics.

Côté, M.-H. (2004). Syntagmatic distinctness in consonant deletion. *Phonology, 21*, 1–41.

Davis, S., & Zawaydeh, B. A. (1997). Output configurations in phonology: Epenthesis and syncope in Cairene Arabic. In *Optional viewpoints* (pp.45–67). Bloomington: Indiana University of Linguistics Club.

Diehl, R., & Kluender, K. (1989). On the objects of speech perception. *Ecological Psychology, 1*, 121–144.

Fabb, N. (1997). *Linguistics and literature: Language in the verb arts in the world*. Oxford: Basil Blackwell.

Fleischhacker, H. (2001). Cluster–dependent epenthesis asymmetries. *UCLA Working Papers in Linguistics, 7*, 71–116.

Fleischhacker, H. (2005). *Similarity in phonology: Evidence from reduplication and loan adaptation*, PhD dissertation, University of California, Los Angeles.

Flemming, E. (1995). *Auditory representations in phonology*. PhD dissertation, University of California, Los Angeles.

Frisch, S. (1996). *Similarity and frequency in phonology*. PhD dissertation, Northwestern University.

Frisch, S., Pierrehumbert, J., & Broe, M. (2004). Similarity avoidance and the OCP. *Natural Language and Linguistic Theory, 22*, 179–228.

Goldston, C. (1994). The phonology of rhyme, ms., California State University, Fresno.

Gordon, M. (2002). A phonetically-driven account of syllable weight. *Language, 78*, 51–80.

Greenberg, J. (1950). The patterning of root morphemes in semitic. *Word, 6*, 162–181.

Greenberg, J. (1960). A survey of African prosodic systems. In S. Diamond (Ed.), *Culture in history: Essays in honor of Paul Radin* (pp. 925–950). New York: Octagon Books.

Guion, S. (1998). The role of perception in the sound change of velar palatalization. *Phonetica, 55*, 18–52.

Halle, M., & Keyser, S. (1971). *English stress, its form, its growth, and its role in verse*. New York: Harper and Row.

Hamano, S. (1998). *The sound-symbolic system of Japanese*. Stanford: CSLI Publications.

Hansson, G. (2001). The phonologization of production constraints: Evidence from consonant harmony. *Proceedings of Chicago Linguistic Society, 37*, 187–200.

Hattori, S. (1984). *Onseigaku* [*Phonetics*]. Tokyo: Iwanami.

Hay, J., Pierrehumbert, J., & Beckman, M. (2003). Speech perception, well-formedness, and the statistics of the lexicon. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in Laboratory Phonology VI* (pp. 58–74). Cambridge: Cambridge University Press.

Hayashi, E., & Iverson, G. (1998). The non-assimilatory nature of postnasal voicing in Japanese. *Journal of Humanities and Social Sciences, 38*, 27–44.

Hayes, B., & Stivers, T. (1995). Postnasal voicing, ms., University of California: Los Angeles.

Hirata, Y., & Whiton, J. (2005). Effects of speech rate on the singleton/geminate distinction in Japanese. *Journal of the Acoustical Society of America, 118*, 1647–1660.

Holtman, A. (1996). *A generative theory of rhyme: An optimality approach*. PhD dissertation, Utrecht Institute of Linguistics.

Howe, D., & Pulleyblank D. (2004). Harmonic scales as faithfulness. *Canadian Journal of Linguistics, 49*, 1–49.

Huang, T. (2001). The interplay of perception and phonology in Tone 3 Sandhi in Chinese Putonghua. *Ohio State University Working Papers in Linguistics, 55*, 23–42.

Hura, S., Lindblom, B., & Diehl, R. (1992). On the role of perception in shaping phonological assimilation rules. *Language and Speech, 35*, 59–72.

Itô, J. (1986). *Syllable theory in prosodic phonology*. PhD dissertation, University of Massachusetts, Amherst.

Itô, J., Kitagawa, Y., & Mester, A. (1996). Prosodic faithfulness and correspondence: Evidence from a Japanese Argot. *Journal of East Asian Linguistics, 5*, 217–294.

Itô, J., & Mester, A. (1986). The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry, 17*, 49–73.

Itô, J., Mester, A., & Padgett, J. (1995). Licensing and underspecification in optimality theory. *Linguistic Inquiry, 26*, 571–614.

Jakobson, R. (1960). *Linguistics and poetics*: *Language in literature*. Cambridge: Harvard University Press.

Jun, J. (1995). *Perceptual and articulatory factors in place assimilation*. PhD dissertation, University of California, Los Angeles.

Kaneko, I., & Kawahara, S. (2002). Positional faithfulness theory and the emergence of the unmarked: The case of Kagoshima Japanese. *ICU [International Christian University] English Studies, 5*, 18–36.

Kang, Y. (2003). Perceptual similarity in loanword adaptation: English post-vocalic word final stops in Korean. *Phonology, 20*, 219–273.

Kawahara, S. (2002). Aspects of Japanese rap rhymes: What they reveal about the structure of Japanese, *Proceedings of Language Study Workshop*.

Kawahara, S. (2006). A faithfulness ranking projected from a perceptibility scale. *Language, 82*, 536–574.

Kawahara, S., Ono, H., & Sudo, K. (2006). Consonant co-occurrence restrictions in Yamato Japanese. *Japanese/Korean Linguistics, 14*, 27–38.

Keating, P. (1988). Underspecification in phonetics. *Phonology, 5*, 275–292.

Keating, P. (1996). The phonology–phonetics interface. *UCLA Working Papers in Phonetics, 92*, 45–60.

Keating, P., & Lahiri, A. (1993). Fronted vowels, palatalized velars, and palatals. *Phonetica, 50*, 73–107.

Kenstowicz, M. (2003). Salience and similarity in loanword adaptation: A case study from Fijian, ms., MIT.

Kenstowicz, M., & Suchato, A. (2006). Issues in loanword adaptation: A case study from Thai. *Lingua, 116*, 921–949.

Kiparsky, P. (1970). Metrics and morphophonemics in the Kalevala. In D. Freeman (Ed.), *Linguistics and literary style* (pp. 165–181). Holt: Rinehart and Winston, Inc.

Kiparsky, P. (1972). Metrics and morphophonemics in the Rigveda. In M. Brame (Ed.), *Contributions to generative phonology* (pp. 171–200). Austin: University of Texas Press.

Kiparsky, P. (1981). The role of linguistics in a theory of poetry. In D. Freeman (Ed.), *Essays in modern stylistics* (pp. 9–23). London: Methuen.

Klatt, D. (1968). Structure of confusions in short-term memory between English consonants. *Journal of the Acoustical Society of America, 44*, 401–407.

Kohler, K. (1990). Segmental reduction in connected speech: Phonological facts and phonetic explanation. In W. Hardcastle, & A. Marchals (Eds.), *Speech production and speech modeling* (pp. 69–92). Dordrecht: Kluwer Publishers.

Leben, W. (1973). *Suprasegmental phonology*. PhD dissertation, MIT.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems. *Language, 48*, 839–862.

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. Ohala & J. Jaeger (Eds.), *Experimental phonology* (pp. 13–44). Orlando: Academic Press.

Maddieson, I. (1997). Phonetic universals. In W. Hardcastle & J. Laver (Eds.), *The handbook of phonetics sciences* (pp 619–639). Oxford: Blackwell.
Maher, P. (1969). English-speakers' awareness of distinctive features. *Language Sciences, 5*, 14.
Maher, P. (1972). Distinctive feature rhyme in German folk versification. *Language Sciences, 16*, 19–20.
Malécot, A. (1956). Acoustic cues for nasal consonants: An experimental study involving tape-splicing technique. *Language, 32*, 274–284.
Malone, J. (1987). Muted euphony and consonant matching in Irish. *Germanic Linguistics, 27*, 133–144.
Malone, J. (1988a) Underspecification theory and turkish rhyme. *Phonology, 5*, 293–297.
Malone, J. (1988b) On the global phonologic nature of classical Irish alliteration. *Germanic Linguistics, 28*, 93–103.
Manabe, N. (2006). Globalization and Japanese creativity: Adaptations of Japanese language to rap. *Ethnomusicology, 50*, 1–36.
McCarthy, J. (1986). OCP effects: Gemination and antigemination. *Linguistic Inquiry, 17*, 207–263.
McCarthy, J. (1988). Feature geometry and dependency: A review. *Phonetica, 43*, 84–108.
McCarthy, J., & Prince, A. (1995). Faithfulness and reduplicative identity. *University of Massachusetts Occasional Papers in Linguistics, 18*, 249–384.
McGarrity, L. (1999). A sympathy account of multiple opacity in wintu. *Indiana University Working Papers in Linguistics, 1*, 93–107.
McCawley, J. (1968). *The phonological component of a grammar of Japanese*. The Hague: Mouton.
Mester, A. (1986). *Studies in tier structure*. PhD dissertation, University of Massachusetts, Amherst.
Mester, A., & Itô, J. (1989). Feature predictability and underspecification: Palatal prosody in Japanese mimetics. *Language, 65*, 258–293.
Miller, G., & Nicely, P. (1955). An analysis of perceptual confusions among some English consonants. *Journal of Acoustical Society of America, 27*, 338–352.
Minkova, D. (2003). *Alliteration and sound change in early English*. Cambridge: Cambridge University Press.
Mohanan, K. P. (1993). Fields of attraction in phonology. In J. Goldsmith (Ed.), *The last phonological rule* (pp. 61–116). Chicago: University of Chicago Press.
Mohr, B., & Wang, W. S. (1968). Perceptual distance and the specification of phonological features. *Phonetica, 18*, 31–45.
Myers, J., & Well, A. (2003). *Research design and statistical analysis*. Mahwah: Lawrence Erlbaum Associates Publishers.
Ohala, J. (1986). A consumer's guide to evidence in phonology. *Phonology Yearbook, 3*, 3–26.
Ohala, J. (1989). Sound change is drawn from a pool of synchronic variation. In L. Breivik, & E. Jahr (Eds.), *Language change: Contributions to the study of its causes* (pp. 173–198). Berlin: Mouton de Gruyter.
Ohala, J., & Ohala, M. (1993). The phonetics of nasal phonology: Theorems and data. In M. Huffman, & R. Krakow (Eds.), *Nasals, nasalization, and the velum* (pp. 225–249). Academic Press.
Padgett, J. (1992). OCP subsidiary features. *Proceedings of North East Linguistic Society, 22*, 335-346.
Padgett, J. (2003). Contrast and post-velar fronting in Russian. *Natural Language and Linguistic Theory, 21*, 39–87.
Pater, J. (1999). Austronesian nasal substitution and other NC̥ Effects. In R. Kager, H. van der Hulst, & W. Zonneveld (Eds.), *The prosody-morphology interface* (pp. 310–343). Cambridge: Cambridge University Press.
Peters, R. (1963). Dimensions of perception for consonants. *Journal of the Acoustical Society of America, 35,* 1985–1989.
Pols, L. (1983). Three mode principle component analysis of confusion matrices, based on the identification of Dutch consonants, under various conditions of noise and reverberation. *Speech Communication, 2*, 275–293.
Prince, A., & Smolensky, P. (2004). *Optimality theory: Constraint interaction in generative grammar*. Oxford: Blackwell,
Pulleyblank, D. (1998). Yoruba vowel patterns: Deriving asymmetries by the tension between opposing constraints, ms., University of British Columbia.
Rose, S., & Walker, R. (2004). A typology of consonant agreement as correspondence. *Language, 80*, 475–531.
Rubin, D. (1976). Frequency of occurrence as a psychophysical continuum: Weber's fraction, Ekman's fraction, range effects, and the phi-gamma hypothesis. *Perception and Psychophysics, 20*, 327–330.

Sagey, E. (1986). *The representations of features and relations in non–liner phonology*. PhD dissertation, MIT.
Shattuck-Hufnagel, S., & Klatt, D. (1977). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Behavior, 18*, 41–55.
Shinohara, S. (2004). A note on the Japanese pun, *Dajare*: Two sources of phonological similarity, ms., Laboratoire de Psychologie Experimentale, CNRS (Centre national de la recherche scientifique).
Silverman, D. (1992). Multiple scansions in loanword phonology: Evidence from Cantonese. *Phonology, 9*, 289–328.
Smith, R., & Dixon, T. (1971). Frequency and judged familiarity of meaningful words. *Journal of Experimental Psychology, 88*, 279–281.
Stemberger, J. (1991). Radical underspecification in language production. *Phonology, 8*, 73–112.
Steriade, D. (2000). Paradigm uniformity and the phonetics/phonology boundary. In M. Broe & J. Pierrehumbert (Eds.), *Papers in laboratory phonology V* (pp. 313–334). Cambridge: Cambridge University Press.
Steriade, D. (2001a) Directional asymmetries in place assimilation. In E. Hume, & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 219–250). San Diego: Academic Press.
Steriade, D. (2001b) The phonology of perceptibility effects: The P-map and its consequences for constraint organization, ms., University of California, Los Angeles.
Steriade, D. (2003). Knowledge of similarity and narrow lexical override. *Proceedings of Berkeley Linguistics Society, 29*, 583–598.
Tabain, M., & Butcher, A. (1999). Stop consonants in Yanyuwa and Yindjibarndi: Locus equation data. *Journal of Phonetics, 27*, 333–357.
Tranel, B. (1999). Optional schwa deletion: On syllable economy in french. In *Formal perspectives on Romance linguistics: Selected papers from the 28th linguistic symposium on Romance linguistics* (pp. 271–288). University of Park.
Trubetzkoy, P. N. (1939). *Principes de phonologie [Principles of phonology]*. Klincksieck, Paris. Translated by Christiane Baltaxe (1969), University of California Press, Berkeley and Los Angeles.
Tsuchida, A. (1997). The *phonetics and phonology of Japanese vowel devoicing*. PhD dissertation, Cornell University.
Tsujimura, N., Okumura, K., & Davis, S. (2006). Rock rhymes in Japanese rap rhymes. *Japanese/Korean Linguistics, 15*.
Ueda, K. (1898). P-Onkoo [On the sound P], *Teikoku Bungaku, 4–1*.
Walden, B., & Montgomery, A. (1975). Dimensions of consonant perception in normal and hearing-impaired listeners. *Journal of Speech and Hearing Research, 18,* 445–455.
Walker, R. (2003). Nasal and oral consonant similarity in speech errors: Exploring parallels with long-distance nasal agreement, ms., University of Southern California.
Wang, M., & Bilger, R. (1973). Consonant confusions in noise: A study of perceptual identity. *Journal of Acoustical Society of America, 54*, 1248–1266.
Wilbur, R. (1973). *The phonology of reduplication*. PhD dissertation, University of Illinois, Urbana-Champaign.
Yip, M. (1984). The development of Chinese verse: A metrical analysis. In M. Aronoff & R. T. Oehrle (Eds.), *Language Sound Structure* (pp. 345–368). Cambridge: MIT Press.
Yip, M. (1988). The obligatory contour principle and phonological rules: A loss of identity. *Linguistic Inquiry, 19*, 65–100.
Yip, M. (1993). Cantonese loanword phonology and optimality theory. *Journal of East Asian Linguistics, 1*, 261–292.
Yip, M. (1999). Reduplication as alliteration and rhyme. *Glot International, 4*(8), 1–7.
Yu, A. (2006). In defense of phonetic analogy: Evidence from Tonal morphology, ms., University of Chicago.
Zuraw, K. (2002). Aggressive reduplication. *Phonology, 19*, 395–439.
Zuraw, K. (2005). The role of phonetic knowledge in phonological patterning: Corpus and survey evidence from tagalog reduplication, ms., University of California, Los Angeles.
Zwicky, A. (1976). This rock-and-roll has got to stop: Junior's head is hard as a rock. *Proceedings of Chicago Linguistics Society, 12*, 676–697.
Zwicky, A., & Zwicky, E. (1986). Imperfect puns, markedness, and phonological similarity: With fronds like these, who needs anemones? *Folia Linguistica, 20,* 493–503.

Discography

Given below is the list of songs whose lyrics or intonational properties are cited in the main text. The list is ordered by song title. In cases in which artists belong to a group, the group name appears first, and the names of individual artists appear in parentheses. Artists name after ''feat.'' are guest vocalists.

''64 Bars Relay,'' SHAKKAZOMBIE (IGNITION MAN, OSUMI) feat. DABO, XBS, and SUIKEN (1999) in *Journey of Foresight*, Cutting Edge.

''Chika-Chika Circuit,'' DELI feat. MACKA-CHIN, MIKRIS, KASHI DA HANDSOME, and GORE-TEX (2003) in *Chika-Chika Maruhi Daisakusen*, Cutting Edge.

''Don't be Gone,'' DJ HAZIME feat. SHAKKAZOMBIE (IGNITION MAN, OSUMI) (2004) in *Ain't No 'stoppin' the DJ*, Cutting Edge.

''Forever,'' DJ TONK feat. AKEEM, LIBRO, HILL-THE-IQ, and CO-KEY (2000) in *Reincarnations*, Independent Label.

''Go to Work,'' KOHEI JAPAN (2000) in *The Adventure of Kohei Japan*, Next Level Recordings.

''Golden MIC,'' ZEEBRA feat. KASHI DA HANDSOME, AI, DOOZI-T, and HANNYA (2003) in *Tokyo's Finest*, Pony Canyon.

''Hip Hop Gentleman,'' DJ MASTERKEY feat. BAMBOO, MUMMY-D, and YAMADA-MAN (2001) in *Daddy's House vol. I*, The Life Entertainment.

''Jibetarian,'' ORIGAMI (NANORUNAMONAI, SHIBITTO) (2004) in *Origami*, Blues Interactions, Inc.

''Kaerimichi,'' ORIGAMI (NANORUNAMONAI, SHIBITTO) (2005) in *Melhentrips*, Blues Interactions, Inc.

''Kin Kakushi Kudaki Tsukamu Ougon,'' DJ HAZIME feat. GAKI RANGER (DJ HI-SWITCH, POCHOMUKIN, YOSHI) (2004) in *Ain't No Stoppin' the DJ*, Cutting Edge.

''Koko Tokyo,'' ACQUARIS (DELI and YAKKO) feat. S-WORD, BIG-O, and DABO (2003) in *Oboreta Machi*, Cutting Edge.

''Mastermind,'' DJ HASEBE feat. MUMMY-D and ZEEBRA (2000), Warner Music Japan.

''My Way to the Stage,'' MURO feat. KASHI DA HANDSOME (2002) in *Sweeeet Baaaad A*s Encounter*, Toys Factory.

''Oxford Town,'' DYLAN, BOB (1990) in *The Freewheelin' Bob Dylan*, Sony.

''Project Mou,'' MOUSOUZOKU (565, DEN, HANNYA, KAMI, KENTASRAS, OOKAMI, MASARU, TSURUGI MOMOTAROU, ZORRO) (2003) in *Project Mou*, Universal J.

''Rapper's Delight,'' THE SUGARHILL GANG (WONDER MIKE, BIG BANK HANK, MASTER GEE) (1991), Sugar Hill.

''Shougen,'' LAMP EYE feat. YOU THE ROCK, G.K. MARYAN, ZEEBRA, TWIGY, and DEV LARGE (1996) Polister.

''Taidou,'' LIBRO (1998) in *Taidou*, Blues Interactions, Inc.

''Tokyo Deadman Walking,'' HANNYA feat. 565 and MACCHO (2004) in *Ohayou Nippon*, Future Shock.

''Tsurezuregusa,'' DABO feat. HUNGER and MACCHO (2002) in *Platinum Tongue*, Def Jam Japan.

''Yamabiko 44-go,'' I-DEA feat. FLICK (GO, KASHI DA HANDSOME) (2004) in *Self-expression, Blues Interactions*, Inc.