音声研究　第13巻第３号
2009(平成21)年12月
●－●頁

Journal of the Phonetic Society
of Japan, Vol.13 No.3
December 2009, pp.00–00

Calculating Vocalic Similarity through Puns

Shigeto KAWAHARA* and Kazuko SHINOHARA**

ダジャレから見る母音の近似性

川原繁人＊・篠原和子＊＊

要旨：近年の音韻論・音声学研究において，音の対応は知覚的に近似するものの間で起こりやすいことが，洒落や韻などの分析をもとに指摘されている（Steriade 2003）。これは日本語のダジャレの音対応分析によっても確認され，特に子音の対応には知覚的近似性の知識が使われていることが示された（Kawahara and Shinohara 2009）。本稿はこれらの研究に基づき，新たに日本語のダジャレにおける母音の対応を分析する。ダジャレのコーパスデータにみられる母音の近似性行列の特徴を分析したところ，弁別素性（distinctive features）に基づく音韻的近似性ではこの特徴は説明しきれず，知覚的近似性を考慮して初めて説明可能となることが判明した。ゆえに本研究の結果からは，弁別素性に基づく音韻論的対応仮説よりも知覚的近似性にもとづく音声学的対応仮説の方が音の対応をより適切に説明できる，という結論が得られる。

Key words: (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 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, Tekieli and Cullinan 1979, Winitz et al. 1972), are rarely if ever present in

* Rutgers University（ラトガーズ大学　Assistant professor）
** Tokyo University of Agriculture and Technology（東京農工大学　准教授）

nasals because intraoral air-pressure 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 and 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 and 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 this 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 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 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 (Fig. 1). This analysis is a multivariate analysis, which finds uncorrelated dimensions that account

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

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

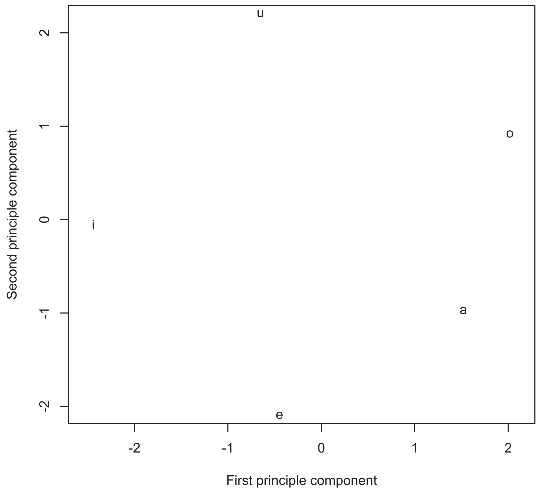

Fig. 1 A principle component analysis of the vowel distance matrix which is computed based on combinability in puns (Table 3).

for the variability in the data as much as possible (see Baayen 2008, Chapter 5)[7]. The two components in Fig. 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 Fig. 1, the five vowels are configured in a way that resembles a vowel space (the parallel becomes clearer if we rotate Fig. 1 clockwise by 45 degrees). In other words, the relative configurations between the elements are comparable between Fig. 1 and a vowel space.

4. Discussion

Fig. 1 resembles a 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).

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 and Huffman (1984) (shown as “K&H”, averaged over seven speakers), and Hirahara and 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 and Kato’s, using the following equation (Liljencrants and Lindblom 1972)[9]:

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

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

Table 4 The feature specifications of the five vowels.

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

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

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 and 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

provide its two-dimensional representation in Fig. 2 to compare it with Fig. 1.

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 and 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]).

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 and 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 and 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 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

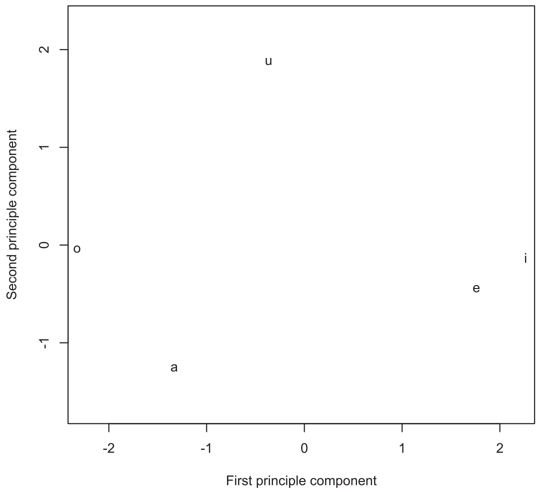

Fig. 2　A principle component analysis of the vowel distance matrix in Table 6.

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

[+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., /k^ja/-/ka/, /k^ju/-/ku/ vs. */k^je/-/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.

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,

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

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

Acknowledgements

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.

Notes

1) These examples (*hop* vs. *stop* and *rock* vs. *stop*) are from *Rapper's Delight* by the Sugerhill Gang.
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).
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.
6) Examples include *uchi-no-tabi to uchuu-no tabi* 'A space trip with my sock' and *kookeeki-ni kuukeeki* 'Cake eaten at the time of good economy.'
7) PCA is similar to a Multi-Dimensional Scaling (MDS) analysis, which has been used to create a vowel distance diagram (Lindblom 1962).
8) A 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 and 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).

References

Baayen, H.R. (2008) *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge 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.274–284.

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.

Tekieli, 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," *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.

(Received Nov. 6, 2008, Accepted Dec. 20, 2009)