MULTI-DIMENSIONAL ANALYSIS OF SONORITY: PERCEPTION, ACOUSTICS, AND PHONOLOGY

Masahiko Komatsu [1,2], *Shinichi Tokuma* [3], *Won Tokuma* [4] *and Takayuki Arai* [1]

[1] Sophia University, Tokyo 102-8554, Japan
[2] University of Alberta, Edmonton T6G 2E7, Canada
[3] Chuo University, Tokyo 192-0393, Japan
[4] Seijo University, Tokyo 157-8511, Japan
`koma2@splab.ee.sophia.ac.jp`

ABSTRACT

Sonority is an important notion in phonology, but its definition has been controversial. Our analysis showed that sonority can be located in a multi-dimensional perceptual space, and that the dimensions of the space have correspondence to both acoustic parameters and phonological features. In the experiment, a confusion matrix was calculated from the results of consonant perception test for LPC residual signals made from /Ca/ syllables. It is considered that LPC residual signals contain only suprasegmental information and thus the confusion pattern for the signals indicates the consonants' similarities in suprasegmental domain. This confusion matrix was analyzed with MDS. The result showed that the consonants can be modeled in a 3-dimensional perceptual space according to their sonority. Its dimensions could be related to acoustic measurements (length, pitch, RMS, HNR) and phonological features ([voice], [sonorant], [continuant]). The result also showed that sonority can be mostly defined within the suprasegmental domain.

1. INTRODUCTION

Sonority is a term that refers to "the loudness of a sound relative to that of other sounds with the same length, stress, and pitch" [1]. Since the late 19th century, the notion of sonority has been used to describe cross-linguistic preferences for certain types of syllable structures and syllable contacts. In phonology, sonority is an essential concept to discuss syllables, and a "sonority scale" is assumed such that phonemes with higher sonority stand closer to the center of the syllable and those with lower sonority stand closer to the margin (Sonority Sequencing Principle). Sonority has also attracted interests of phoneticians (e.g. [2] for acoustic approach). In spite of its importance, however, the definition of sonority is still controversial. (See [3] for an overview on related studies.)

An important aspect of sonority is that it is a "multivariate" property, which has been suggested by phonetic studies. Ohala and Kawasaki [4] demonstrate that the sonority scale cannot handle some types of syllable formation, and they argue that "[the sonority scale] is one-dimensional where it should have as many dimensions as there are acoustic-auditory parameters that can be used to form lexical contrasts." Kawai [5] reports that the sonority rank can be best predicted with pitch and band-pass filtered power measurements based on multiple linear regression analyses.

Most phonologists define sonority with a set of distinctive feature specifications rather than define it with a single primitive feature, although they assume a one-dimensional sonority scale. For example, Blevins [6] proposes a sonority scale defined with seven features. It is also of note that Flemming [7] incorporates a multi-dimensional auditory space into a current phonological theory, though he does not discuss sonority. In his theory, each dimension of the space corresponds to one or more phonological features. Thus, it can be said that sonority consists of multiple features, and that phonological features can be represented with a multi-dimensional auditory space. However, it is not clear how Blevins' features correspond to acoustic-auditory parameters or how sonority is described in Flemming's type space.

This paper aims to find the phonological/phonetic basis of sonority. In this paper, we start with phonetic analyses to investigate whether sonority can be located in a multi-dimensional auditory or perceptual space and whether it has acoustic correlates; and finally discuss the correspondence of phonological features to the perceptual space.

In our previous experiment [8], listeners identified major classes of consonants (i.e., categories corresponding to a sonority scale) in the modified LPC residual signal at the identification rate of 66.4%, while their identification of consonants as a whole was 20.0%. This means that consonants in the modified LPC residual contain sufficient information on sonority perception but little on the perception of other features. Therefore, in this paper we analyze the perceptual data of [8].

2. ANALYSIS OF EXPERIMENTAL DATA

2.1. Stimulus set and procedure of the perceptual experiment

The stimuli and procedure of the perceptual experiment conducted in our earlier study [8] are described here.

In [8], we used LPC residual signals made from 17 Japanese /C/+/a/ syllables. To create the residual signals, original samples were selected from ATR Japanese database [9] (downsampled to 16kHz). From these samples, residual signals were made by 22nd-order LPC (using Hamming window; frame: 512 points, 75% overlap), adjusting their power at each frame to match their original samples. The residual signals were made with a spectral tilt of −6dB/octave so that they sound like human speech rather than noise. The output residual signals were expected to retain most of the information on pitch, amplitude, and harmonics-to-noise ratio of the original signals, but not the spectral information.

15 native speakers of Japanese (4 males and 11 females; 19-22 years old) participated in the perceptual experiment. The

experiment contained 51 stimulus presentations (17 syllables × 3 repetitions). The presentation order was randomized for each participant. In this experiment, the participants were asked to identify the /Ca/ syllables which they listened to through headphones, by clicking with the mouse the right syllable appearing on a PC display. This experimental setup allows participants to proceed at their own pace. This experiment was carried out in a CALL system room at Sophia University, which can accommodate several participants at one time.

Table 1 shows the consonant inventory in the stimulus set, which includes phonetic and phonological descriptions of each consonant. Consonants with higher sonority ranks are regarded more "sonorous." "–" indicates that the sonority rank cannot be determined for those consonants. (See Section 4 for the detailed explanation.) Major classes are the phonological categories corresponding to a sonority scale [3].

Table 1: List of consonants in the stimulus set.

roman-ized trans. & IPA		phonetic			phonological	
		manner	place	voice	sonor-ity rank	major class
p	[p]	plosive	bilabial	voiceless	1	obstru-ent
t	[t]	plosive	alveolar	voiceless		
k	[k]	plosive	velar	voiceless		
b	[b]	plosive	bilabial	voiced	2	
d	[d]	plosive	alveolar	voiced		
g	[ɡ]	plosive	velar	voiced		
s	[s]	fricative	alveolar	voiceless	3	
sy	[ʃ]	fricative	alveopalatal	voiceless		
h	[h]	fricative	glottal	voiceless		
ty	[tʃ]	affricate	alveopalatal	voiceless	–	
z	[dz]	affricate	alveolar	voiced	–	
zy	[dʒ]	affricate	alveopalatal	voiced		
m	[m]	nasal	bilabial	voiced	5	nasal
n	[n]	nasal	alveolar	voiced		
r	[ɾ]	flap	alveolar	voiced	–	liquid
y	[j]	approximant	palatal	voiced	6	glide
w	[w]	approximant	labiovelar	voiced		

2.2. Multi-dimensional scaling (MDS) of the perceptual data

To construct a multi-dimensional space from the perceptual data, we used an MDS analysis. This method creates the optimal spatial representation of the data based on their perceptual similarity, and further enables us to relate the perceptual data to acoustic and phonological parameters.

First, the confusion matrix (17 stimuli × 17 responses) was obtained from the perceptual experiments, by summing up the scores across all participants and trials. This confusion matrix serves as the similarity matrix of consonants.

Then, this matrix was fed to an MDS procedure of SPSS Version 11.0J (ALSCAL command, non-metrical, asymmetric, 3-dimensional). We determined the optimal number of dimensions based on the stress and R^2 values.

2.3. Acoustic measurements

Before interpreting the results of MDS analysis in terms of the acoustic properties, it is necessary to see which acoustic properties of the original samples in fact remains in the LPC residual signals and possibly acts as a perceptual cue. Therefore, we analyzed the acoustic properties of the original samples as well as the residuals and compared them. Labeling and acoustic measurements were made with Praat (Version 4.0.5).

Labeling of the consonant segments in the original samples was carried out by visual inspection of waveforms and spectrograms with the help of aural judgement. The segment boundaries were determined so that the transitions to vowels were excluded, following [10] with necessary modifications. The actual labeling process of the original samples was as follows: Voiceless plosives were labeled from the beginning of release to the onset of voicing. Voiced plosives were labeled from the onset of prevoicing to the end of friction, i.e., they included both prevoicing and release. (We also tried segmenting voiced plosives into prevoicing and release, but the measurements obtained from this method of segmentation did not make much difference in the results.) Voiceless fricatives were labeled from the start of noise to the onset of voicing. Voiceless affricates were labeled from the beginning of release to the onset of voicing. Voiced affricates were treated with the criteria for the beginning of voiced plosives and the end of voiced fricatives. Nasals were labeled from the onset of voicing to the abrupt change of formants. /r/, a flap, was treated in the same manner as voiced plosives. /y/ was labeled from the onset of voicing to the dip of F3. /w/ was labeled from the onset of voicing to the start of F2 rise.

The residual signals were labeled with the same criteria as the original samples except for those that needed formant information, specifically, nasals and approximants. In determining the boundary from a nasal to a vowel, the boundary location in the original sample was copied, because no reliable cues were available. As for approximants, the change of intensity was adopted as the identifier of the boundary. (As a result, the residual signal of /y/ had the different boundary location from the original sample; and /w/ had the same boundary as the original sample.)

Nine acoustic measurements listed in Table 2 were obtained for each of the labeled segments. The table shows Pearson's correlation coefficients r between the residual signals and the original samples for each of the nine measurements. Six measurements of the residual signals that showed strong correlation ($r > .7$) with the original signals were used in the interpretation of the perceptual space produced by MDS: duration (sec), mean pitch (Hz), maximum of amplitude (Pascal), RMS of amplitude (Pascal), total energy (Pa^2 sec), and mean Harmonics-to-Noise Ratio (HNR, dB).

Table 2: Pearson's correlation of acoustic measurements for LPC residual and original samples. We used the measurements of r>.7 for further analysis (those with *).

	r
duration (sec)	0.986 *
pitch mean (Hz)	0.929 *
pitch max (Hz)	0.693
amplitude max (Pascal)	0.969 *
amplitude RMS (Pascal)	0.979 *
total energy (Pa^2 sec)	0.959 *
intensity (dB)	–0.006
HNR max (dB)	0.624
HNR mean (dB)	0.869 *

3. RESULTS OF THE ANALYSIS

3.1. Perceptual space

The MDS analysis of the confusion matrix produced results shown in Fig. 1. Although fitting of the data may not be satisfactory (Kruskal's stress 1: .267; R^2: .574), we can clearly see that consonants with the same sonority rank tend to cluster together (see Fig. 1a).

In the plane of Dimensions 1 and 2 (Fig. 1b), consonants are grouped into voiceless obstruents, voiced obstruents, and nasals/glides. /r/, a liquid, stands apart from the nasal/glide cluster and is located closer to voiced obstruents. This is plausible because Japanese flap /r/ is phonetically somewhat close to voiced plosives, different from English liquids /l/ and /r/.

In Dimensions 2 and 3 (Fig. 1c), nasals and glides are clearly apart from the other consonants.

In Dimensions 3 and 1 (Fig. 1d), all voiced consonants gather together, and voiceless consonants are separated into plosives/affricate and fricatives. This is in contrast with Dimensions 1 and 2 (Fig. 1b), where voiceless consonants cluster together and voiced ones are separated into two groups.

3.2. Correspondence of perceptual dimensions to acoustic measurements

We investigated whether each dimension of the perceptual space can be related to the six acoustic measurements obtained in 2.3. Table 3 shows the results of Spearman's correlation test by rank orders on pairs of consonants' coordinates in each dimension and acoustic measurements. "*" indicates that the correlation is significant at $p < .05$; and "**", $p < .01$. Note that not all measurements were available for all 17 consonants. "Pitch mean" misses the data from all voiceless consonants, and "HNR mean" misses the data from /ta/ and /pa/ because their durations were too short to estimate HNR.

Table 3: Spearman's rank-order correlations of acoustic measurements with each dimension of perceptual space. * indicates p<.05; **, p<.01.

	N	*ρ*		
		Dim 1	Dim 2	Dim 3
duration	17	−0.015	0.311	−0.691 **
pitch mean	10	0.200	−0.709 *	0.321
amplitude max	17	0.287	−0.419	−0.358
amplitude RMS	17	0.199	−0.527 *	−0.199
total energy	17	0.174	−0.380	−0.444
HNR mean	15	0.711 **	−0.046	0.104

Table 3 shows that Dimension 1 is correlated with mean HNR. This result is quite straightforward to interpret when we see Fig. 1. It is clear that Dimension 1 divides consonants into voiced (right side in Fig. 1b) and voiceless ones (left side in Fig. 1b). This division is plausible because HNR is considered to be a good indicator of voicing.

Table 3 also shows that Dimension 2 is correlated with RMS. Nasals and approximants have the largest degree of power, voiceless obstruents the next, and voiced obstruents the smallest (larger the power, more leftward in Fig. 1c). This dimension is also correlated with mean pitch, showing that nasals and approximants have higher pitch than voiced plosives (higher the pitch, more leftward in Fig. 1c). Consonants whose pitch is not

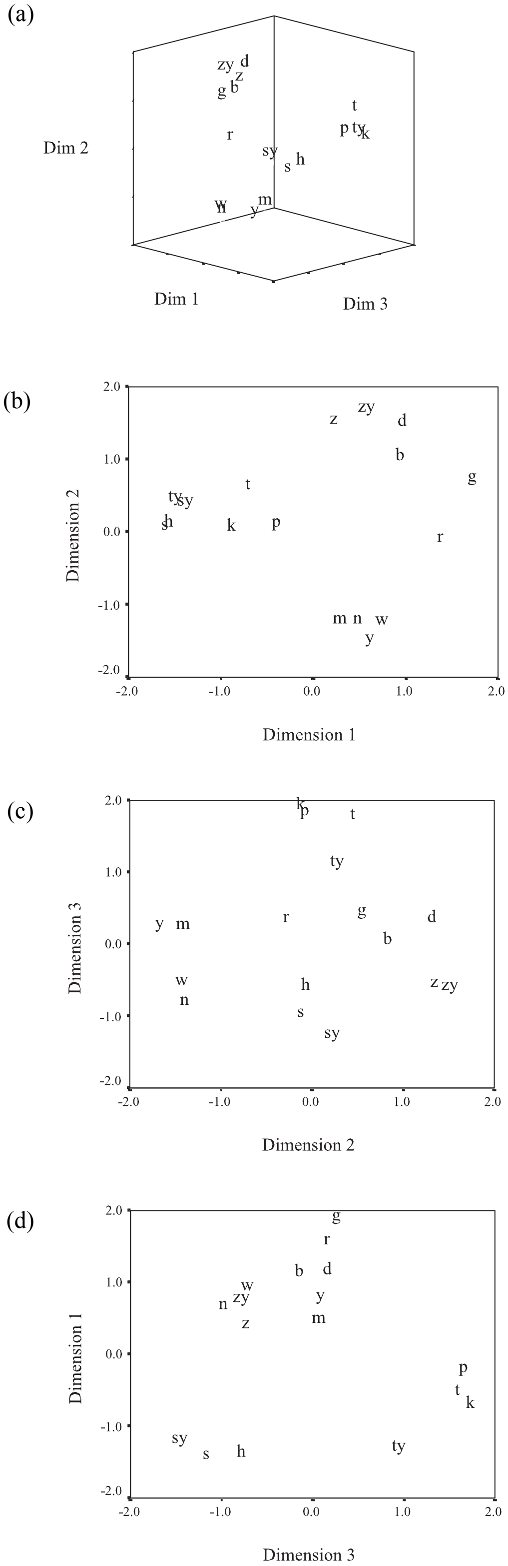

Figure 1: Three-dimensional analysis of consonant perception in LPC residual.

available are located in the region where Dimension 2 is around 0.

Dimension 3 is correlated with durations of segments. In this dimension, the consonants are located in the order of, from the longest to the shortest, voiceless fricatives, nasals/approximants, voiced obstruents, voiceless plosives, with some overlaps (longer the duration, more leftward in Fig. 1d).

4. CORRESPONDENCE TO PHONOLOGY

In Blevins' proposal [6], five distinctive features are relevant to fixing the sonority rank of consonants: [consonantal], [sonorant], [nasal], [continuant], and [voice]. Table 4 shows the feature specifications of each consonant used in the experiment. "*" indicates that the feature is not specified. The sonority ranks in the table are based on Blevins' proposal. The sonority rank of 4, missing in the table, corresponds to voiced fricatives, which do not appear at the word initial position in Japanese typically. "?" indicates that the rank cannot be defined in an uncontroversial manner. Affricates are usually analyzed to have both [−cont] and [+cont], and it is not clear how we can handle such phonemes in terms of Blevins' sonority definition. While English /r/ is categorized as an approximant (like /w y/) and should have the rank of 6, Japanese /r/ is a flap, phonetically somewhat close to voiced plosives (/b d g/) (e.g., see [11]) although it is phonologically a liquid.

Table 4. Phonological feature specifications and sonority according to Blevins' proposal [6].

	[cons]	[son]	[nas]	[cont]	[voice]	sonority rank
p t k	+	−	*	−	−	1
b d g	+	−	*	−	+	2
s sy h	+	−	*	+	−	3
ty	+	−	*	− +	−	?
z zy	+	−	*	− +	+	?
m n	+	+	+	*	*	5
r	+	+	−	*	*	?
w y	+	+	−	*	*	6

Ignoring these affricates and flap, we can see that all distinctions of consonant groups by sonority ranks are maintained in the perceptual space except for the nasal-approximant distinction (see Fig. 1a).

As shown in Table 4, the feature [son] separates consonants into two classes, namely the [−son] class (/p t k b d g s sy h ty z zy/) and the [+son] class (/m n r w y/). In Fig. 1c, the [−son] consonants are situated approximately in the region where Dimension 2 is 0 or positive (right half of Fig. 1c), and the [+son] consonants are in the opposite side in the figure. Thus, Dimension 2 approximately corresponds to the feature [son]; the positive region of the dimension corresponds to [−son], and the negative region to [+son].

Likewise, we can find the correspondence of [cont], which distinguishes /p t k b d g/ ([−cont]) from /s sy h/ ([+cont]). The positive region of Dimension 3 (right half of Fig. 1d) approximately corresponds to [−cont], and the negative to [+cont].

[voice] distinguishes /p t k s sy h ty/ from /b d g z zy/. The positive region of Dimension 1 (right half of Fig. 1b) corresponds to [+voice], and the negative to [−voice].

The feature [nasal] differentiating the nasals from approximants do not seem to appear in the perceptual space. This feature may correspond to some spectral property in acoustic domain, which is lost in the LPC residual signal.

5. CONCLUDING REMARKS

Our analysis has shown that sonority property can be modeled if we assume a multi-dimensional space. In phonetic terms, sonority is located in a multi-dimensional perceptual space, and its dimensions have correspondence to acoustic parameters. Phonologically, the dimensions can be associated with distinctive features. Sonority is not a mere notion that phonologists assume, but it has a phonetic basis, both auditory and acoustic.

LPC residual signal retains the suprasegmental properties of the original sample while it has lost its spectral properties; thus the confusion pattern of consonants in this signal indicates the consonants' similarities in suprasegmental domain. Because we can construct the perceptual space of sonority from this signal, we consider that most of the sonority information is contained in the suprasegmental property of speech. The nasal-approximant contrast, or [nasal], may be related to the spectral property in speech, which we could not confirm in our analysis.

* We are indebted to Noriko Yamane for offering information on several references cited in this paper.

6. REFERENCES

[1] P. Ladefoged, *A course in phonetics* (Harcourt Brace Jovanovich, New York, 1975).

[2] P. J. Price, "Sonority and syllabicity: Acoustic correlates of perception", *Phonetica*, **37**, 327-343 (1980).

[3] G. N. Clements, "The role of the sonority cycle in core syllabification", in *Papers in laboratory phonology I*, M. E. Beckman and J. Kingston, Eds. (Cambridge Univ. Press, Cambridge, 1990), pp. 283-333.

[4] J. J. Ohala and H. Kawasaki, "Prosodic phonology and phonetics", *Phonology Yearbook*, **1**, 113-127 (1984).

[5] G. Kawai, "Estimating sonority from F0 and speech energy measurements", in *Proc. Second Spontaneous Speech Science and Technology Workshop* (Tokyo, 2002), pp. 113-120.

[6] J. Blevins, "The syllable in phonological theory", in *The handbook of phonological theory*, J. A. Goldsmith, Ed. (Basil Blackwell, Cambridge, MA, 1995), pp. 206-244.

[7] E. S. Flemming, *Auditory representations in phonology* (PhD diss., Univ. of California, Los Angeles, 1995).

[8] M. Komatsu, W. Tokuma, S. Tokuma and T. Arai, "The effect of reduced spectral information on Japanese consonant perception: Comparison between L1 and L2 listeners", *Proc. ICSLP*, **2000**, Vol. 3, 750-753 (2000).

[9] K. Takeda, Y. Sagisaka, S. Katagiri, M. Abe and H. Kuwabara, *Speech Database User's Manual* (Advanced Telecommunications Research Institute International, Kyoto, 1988).

[10] G. E. Peterson and I. Lehiste, "Duration of syllable nuclei in English", *J. Acoust. Soc. Am.,* **32**, 693-703 (1960).

[11] T. Arai, "A case study of spontaneous speech in Japanese", *Proc. ICPhS*, **1999**, 615-618 (1999).