Reconsidering the method of finding phonemic category boundary in speech production

Shigeaki Amano1,*, Yukari Hirata2,† and Kimiko Yamakawa3,‡

1Faculty of Human Informatics, Aichi Shukutoku University, 2–9 Katahira, Nagakute, 480–1197 Japan
2Department of East Asian Languages and Literatures, Colgate University, 13 Oak Drive, Hamilton, NY 13346 USA
3Faculty of Contemporary Culture, Shokei University, 2–6–78 Kuhonji, Chuo-ku, Kumamoto, 862–8678 Japan

(Received 10 March 2022, Accepted for publication 15 June 2022)

Abstract: Several previous studies in speech production obtained a boundary of two phonemes by an exhaustive search for the “optimal boundary” that provides the best phoneme classification on an acoustic dimension such as voice onset time. However, they used this method without confirming its accuracy in estimating the true boundary. The present study examined this accuracy by conducting a Monte Carlo simulation under several conditions of the means and variances of two phoneme categories. The results revealed that, compared with the statistically estimated boundary, the “optimal boundary” tended to have bigger root mean square errors from the true category boundary. This finding indicates that the “optimal boundary” is less accurate than the statistically estimated boundary for estimating the true boundary. Our study considers the limitations of the previous method and recommends a statistically estimated boundary in future investigations.

Keywords: Optimal boundary, Phoneme category, Monte Carlo simulation, Discriminant analysis

1. INTRODUCTION

Phoneme categories and the boundary between them are important concepts in speech science. Specifying the precise boundary between two phonemes is crucial not only for distinguishing between meanings in very basic communication but also for cross-linguistic comparisons or describing challenges in second language acquisition. For example, when second language learners aim to accurately produce two distinct Japanese words, /saka/ ‘slope’ vs. /sakka/ ‘writer’, it is important to know the precise boundary in the duration of the contrasting underlined consonants with which the meaning changes. The category boundary of phonemes is often estimated in native speakers’ production using the major acoustic features of the phonemes (e.g., duration of contrasting consonants in the above example). Using the category boundary, the classification error of phonemes is obtained to show how distinctly the phonemes are separated from each other. Low classification error is used as one of the pieces of evidence that the acoustic feature is relevant to the phoneme category.

Some previous studies [e.g., 1–3] used a statistical method to estimate the category boundary, while other studies [e.g., 4,5] used a primitive but simple method to determine the “optimal boundary” of two categories. The optimal boundary has typically been obtained by an exhaustive search for the point providing the best classification of data in two phoneme categories. For example, the boundary of the voiced consonant /b/ and voiceless consonant /p/ was exhaustively searched from short to long voice onset time (VOT) with small steps. The VOT giving the minimum classification error for /b/ and /p/ was hence determined.

The “optimal boundary” method was used to classify voiced and voiceless stop consonants in English [4,5]; singleton and geminate consonants in Maltese [6]; voiced and voiceless stop consonants in Icelandic [7]; fortis, lenis, and aspirated voiceless stops in Korean [8]; short and long vowels in Japanese [9]; and singleton and geminate consonants in Japanese [10].

However, the optimal boundary was used without confirming its accuracy in estimating the true category boundary. The obtained optimal boundaries in previous

*e-mail: psy@asu.aasa.ac.jp
†e-mail: yhirata@colgate.edu
‡e-mail: jin@shokei-gakuen.ac.jp
[doi:10.1250/ast.43.322]
studies may be inaccurate or may lack generalization for applying to a wide range of other data samples. Therefore, the present study examined the accuracy of the optimal boundary and compared it with that of the statistically estimated boundary by conducting a Monte Carlo simulation [11] for classifying two categories with various means and variances.

2. FORMULATION

According to the previous studies [e.g., 4,5], the optimal boundary (OB) can be defined as Eq. (1):

\[ \text{OB} = (l + h)/2 \] (1)

where \( l \) and \( h \) are respectively the lower and upper limits of the range that gives the minimum classification error of two distributions by an exhaustive search. Note that the minimum classification error is not always obtained at a single point. It can be obtained in a range according to how the two distributions overlap (Fig. 1). Equation (1) indicates that OB is provided as the midpoint of the range defined by \( l \) and \( h \).

The statistically estimated boundary (SB) is a classifier for one-dimensional data in discriminant analysis [12]. Hence, it is defined as Eq. (2):

\[ \text{SB} = (m_1v_2 + m_2v_1)/(v_1 + v_2) \] (2)

where \( m_1 \) and \( m_2 \) are the means of each data distribution, and \( v_1 \) and \( v_2 \) are their unbiased variances, respectively. Note that \( m_1, m_2, v_1, \) and \( v_2 \) are obtained from the data.

The theoretical boundary (TB) is defined with the parameters of two normal distributions as Eq. (3):

\[ \text{TB} = (M_1V_2 + M_2V_1)/(V_1 + V_2) \] (3)

where \( M_1 \) and \( M_2 \) are the means of each population, and \( V_1 \) and \( V_2 \) are their variances, respectively (Fig. 2). TB provides a true category boundary whereas OB and SB provide a category boundary calculated from the data.

3. SIMULATION

3.1. Method

Simulation conditions were set up as follows. Two phoneme categories follow a normal distribution on the

![Fig. 1 Examples of two distributions, their classification error, and the optimal boundary (OB). According to the overlap of two distributions, the minimum classification error is obtained (a) at a single point or (b) in a range with the lower limit (l) and the higher limit (h). Note that l and h are equal in (a). OB is calculated using Eq. (1) in both (a) and (b).](image)

![Fig. 2 Conceptual diagram of the theoretical boundary of two normal distributions with the mean \( M_1 \) and \( M_2 \): (a) equal variance \( (V_1 = V_2) \) and (b) unequal variance \( (V_1 > V_2) \). The expected classification error is shown as a gray area.](image)
z-score having the means and variances shown in Table 1. Distance between the two categories (i.e., the difference between \( M_1 \) and \( M_2 \)) was altered by keeping distance from the origin (\( z = 0 \)) equal. Variance \( V_2 \) was altered from 0.5 to 1.0 by keeping \( V_1 \) constant. TB was obtained with Eq. (3) using these means and variances. The amount of expected classification error depends on the two means and variances (Table 1) because the error corresponds to the overlapping area of the two phoneme categories (Fig. 2).

Random numbers following the standard normal distribution were generated by the Box-Muller method [13], and they served as the sample dataset for two phoneme categories under specified conditions in Table 1. We set the sample size of each phoneme category to be 25, 50, 100, 250, 500, and 1,000. We choose these sample sizes and the expected classification errors in Table 1 so that they roughly covered those in the previous studies [e.g., 4, 5, 9].

Using the sample dataset, OB was calculated with Eq. (1). By seeking \( z \) from \( M_1 \) to \( M_2 \) in the 0.01 step, \( l \) and \( h \) were respectively determined as the lowest and highest \( z \) that provided the range of the minimum classification error of the two categories. SB was calculated with Eq. (2) using the mean and variance obtained from the sample dataset. OB and SB were calculated for every condition in Table 1 for every sample size. In each condition, OB and SB were obtained 100 times using the newly generated sample datasets each time. Note that OB and SB were calculated with the same pair of sample datasets for each time to enable an exact comparison of accuracy.

The root mean square errors (RMSE) of OB and SB from TB were calculated using 100 datasets for each OB and SB to reveal the accuracy in estimating the boundary. Because RMSE represents a deviation from TB, high RMSE indicates a low accuracy in estimating the true category boundary.

### 3.2. Results

Figure 3 shows the RMSE of OB and SB for each condition of the mean and \( V_2 \) as a function of sample size. The following four points were observed:

1. Overall, RMSE became bigger as the distance between \( M_1 \) and \( M_2 \) increased when Figs. 3(a), 3(b), and 3(c) were compared;

| Mean (\( z \)) | Variance (\( z \)) | Theoretical boundary (\( z \)) | Expected classification error (%) |
|---------------|-------------------|--------------------------------|----------------------------------|
| \( M_1 \)     | \( M_2 \)         | \( V_1 \)                      | \( V_2 \)                        |
| -1.28         | 1.28              | 1.00                           | 1.00                             |
| -1.64         | 1.64              | 1.00                           | 1.00                             |
| -1.96         | 1.96              | 1.00                           | 1.00                             |

\( z = 0 \) and \( V_1 = 1.00 \), \( V_2 = 0.75, 0.50, 1.00 \).

Fig. 3 Root mean square error (RMSE) of optimal boundary (OB) and statistically estimated boundary (SB) from the theoretical boundary obtained from the means (\( M_1 \) and \( M_2 \)) and variances (\( V_1 \) and \( V_2 \)) of two distributions where \( V_1 \) is always 1.00. (a) \( M_1 = -1.28 \) and \( M_2 = 1.28 \), (b) \( M_1 = -1.64 \) and \( M_2 = 1.64 \), (c) \( M_1 = -1.96 \) and \( M_2 = 1.96 \).
(2) Both RMSE of OB and SB decreased as the sample size increased;
(3) As $V_2$ decreased, RMSE of OB increased whereas RMSE of SB tended to be constant;
(4) RMSE tended to be bigger in OB than SB.
To precisely examine the tendency described in above (4), ratios of the RMSE of OB to the RMSE of SB were calculated (Fig. 4). In all cases in Figs. 4(a)–4(c), the ratio increases as the sample size increases and $V_2$ decreases. In addition, the ratio tended to be more than 1.0, indicating that OB is less accurate than SB.

4. DISCUSSION

Figures 3 and 4 show that OB tends to have a bigger RMSE than SB. In the worst case, OB was 7.51 times bigger than SB. Average RMSE ratio was 2.34 ($n = 54$, $SD = 1.65$). These results indicate that OB is less accurate than SB for estimating TB.

Both RMSEs of OB and SB decreased as a function of sample size, which is reasonable because more samples provide a more accurate estimation according to the law of large numbers [cf. 14]. On the other hand, the RMSE ratio increases as a function of sample size (Fig. 4), indicating that OB becomes less accurate than SB when sample size increases. The ratio increase is caused by the fact that the RMSE of SB decreases more than that of OB as the sample size increases. In other words, SB can estimate TB more accurately than OB when the sample size becomes large. Moreover, OB is less stable than SB in terms of the variance inequality of two distributions because the RMSE of OB became worse as $V_2$ decreased to be more different from $V_1$ in Fig. 3. Meanwhile, SB is stable even if the variances are unequal.

OB is disadvantageous because it is obtained without assuming and estimating a population (Eq. (1)), whereas SB is obtained by estimating the population using means and unbiased variances (Eq. (2)). In other words, OB is optimal for one particular sample dataset but not for the population. Because the sample dataset usually fluctuates owing to random factors, OB is strongly affected by the fluctuation. In contrast, SB is less affected by the fluctuation because it uses the mean and unbiased variance that are statistically stable measures as long as a sufficient amount of data is provided. This stability of SB can partly be observed in Fig. 3 where RMSE of SB is not affected by $V_2$ conditions.

The simulation of this study assumed that phoneme data follow a normal distribution. The results of this study cannot be applied to the data not having a normal distribution. Therefore, it is vital to confirm whether the phoneme data normally distributed before applying the results of this study to actual phoneme data.

Finding categorical boundaries for phonemic contrasts is one of the key issues in speech science. What we want to know is the true boundary of two phoneme categories in a population, instead of one particular category boundary in a sample dataset that can be affected by various random factors. The present study provides empirical evidence that SB is more appropriate than OB that has conventionally been used in previous studies. We suggest that future studies should consider the limitations of the previous method, and we encourage the use of SB in future investigations.

ACKNOWLEDGMENT

This study was supported by JSPS KAKENHI Grant Number JP21H00533.

REFERENCES

[1] S. Amano and Y. Hirata, “Perception and production boundaries between single and geminate stops in Japanese,” J. Acoust. Soc. Am., 128, 2049–2058 (2010).
[2] S. Amano and Y. Hirata, “Perception and production of singleton and geminate stops in Japanese: Implications for the theory of acoustic invariance,” Phonetica, 72, 43–60 (2015).
[3] K. Nagao and K. de Jong, “Perceptual rate normalization in naturally produced rate-varied speech,” J. Acoust. Soc. Am.,
[4] J. L. Miller, K. P. Green and A. Reeves, “Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast,” *Phonetica*, 43(1–3), 106–115 (1986).

[5] S. Nakai and J. M. Scobbie, “The VOT category boundary in word-initial stops: Counter-evidence against rate normalization in English spontaneous speech,” *Lab. Phonol.*, 7(1):13, 1–31 (2016).

[6] H. Mitterer, “The singleton-geminate distinction can be rate dependent: Evidence from Maltese,” *Lab. Phonol.*, 9(1):6, 1–16 (2018).

[7] J. Pind, “Speech segment durations and quantity in Icelandic,” *J. Acoust. Soc. Am.*, 106, 1045–1053 (1999).

[8] E. J. Oh, “Voice onset time of Korean stops as a function of speaking rate,” *Phon. Speech Sci.*, 1(3), pp. 39–48 (2009).

[9] Y. Hirata, “Effects of speaking rate on the vowel length distinction in Japanese,” *J. Phon.*, 32, 565–589 (2004).

[10] Y. Hirata and J. Whiton, “Effects of speaking rate on the single/geminate stop distinction in Japanese,” *J. Acoust. Soc. Am.*, 118, 1647–1660 (2005).

[11] S. S. Sawilowsky, “You think you’ve got trivials?” *J. Mod. Appl. Stat. Methods*, 2, 218–225 (2003).

[12] Y. Tanaka and K. Wakimoto, *Methods of Multivariate Statistical Analysis* (Gendai-Sugakusha, Kyoto, 1983) (in Japanese).

[13] G. E. P. Box and M. E. Muller, “A note on the generation of random normal deviates,” *Ann. Math. Stat.*, 29, 610–611 (1958).

[14] H. Tijms, “The law of large numbers and simulation,” in *Understanding Probability: Chance Rules in Everyday Life*, 2nd ed. (Cambridge University Press, Cambridge, 2007), Chap. 2, pp. 17–72.