# Perception of the Place of Articulation Feature for Plosives and Fricatives in Noise

**Willa Chen*** and **Abeer Alwan**†

* TRW, CA

† Depart. of Electrical Engineering, UCLA
alwan@icsl.ucla.edu

## ABSTRACT

This study aims at uncovering perceptually-relevant cues for place of articulation in noise. Stimuli consisted of /CV/ syllables where /C/ is one of /b,d,p,t,f,s,v,z/ and /V/, one of /a,i,u/, in different levels of white noise (SNR range: -15 to 10 dB.) Four subjects participated in the study. Results show that the perception of the labial and alveolar place distinction in noise is dependent on the manner, voicing, and vowel. Perception of place for fricatives is more robust than that for plosives except for voiced /Ca/ syllables. For voiceless consonants, differences in thresholds between fricatives and plosives are, on average, 11.8 dB, 3.8 dB, and 5 dB for the /a/, /i/ and /u/ contexts, respectively. Voiceless consonants are typically more robust than their voiced counterparts; this is especially true for fricatives. As for context, the /a/ context is more robust than /u/ which is more robust than /i/ except for the /p,t/ case.

## 1 INTRODUCTION

This paper is part of a series of papers [1, 2, 3] which attempt to uncover perceptually-relevant acoustic correlates of linguistic features in noise. The focus here is on the perception of the labial and alveolar place of articulation distinction for plosives and fricatives in noise. One of the earliest studies on perceptual confusions in noise was done by Miller and Nicely in 1955 [4]. They used 16 consonants across five articulatory features (voicing, nasality, affrication, duration and place of articulation), with initial or final consonant positions in monosyllabic nonsense syllables with the vowel /a/. Each of the 16 phonemes can be confused with any other phoneme. The study showed that voicing is less affected by masking noise than place of articulation (voicing was still discriminable at SNRs as poor as -12 dB while place confusions become significant at about 6dB). They also found that plosives are less robust than fricatives in noise. Other researchers used an information/theoretic approach to try to model confusion matrices of speech in noise [5], [6]. The studies attempt to find out which acoustic and phonetic cues account for perceptual results in noise by analyzing confusion matrices statistically. Results are highly dependent upon the acoustic and phonetic dimensions chosen.

In this study, we focus on the labial/alveolar place of articulation distinction for plosives and fricatives in three vowel contexts.

## 2 EXPERIMENTS

### 2.1 SUBJECTS:

Two males and two females with normal hearing participated in the perceptual experiments. Subjects were native talkers of American English and their age ranged between 18 - 36 years old. Each subject was paid $10 an hour. None of the subjects was experienced with perceptual experiments; subjects did go through one-hour long training sessions before the experiments started.

### 2.2 STIMULI

**2.2.1 Syllables:** The /CV/ syllables are part of the UCLA Speech Processing and Auditory Perception Laboratory speech database. The sound samples were recorded from four native American English talkers, two males and two females. Each token was normalized so that the maximum energy of the syllable is the same for all tokens. The consonant was one of /b,d,p,t,f,s,v,z/ and the vowel, one of /a,i,u/. The signals were played at 60 dB SPL for 160 ms. There are four tokens per speaker per syllable. Each token was played twice, thus each /CV/ was played 32 times.

**2.2.2 Masker:** The additive masker was white Gaussian noise. The signal level in the experiments was fixed while the level of the white noise was adjusted to result in different SNRs.

### 2.3 EXPERIMENTAL PROTOCOL

The experiments were of the 2AFC (Two Alternate Forced Choice) type. Each experiment focused on one pair of /CV/s that only differed by place of articulation, i.e. /ba,da/, /fi,si/ or /vu,zu/, and was divided into seven sections. Each section involved stimuli at a certain SNR. Noise levels were set at -15, -10, -5, 0, 5, 10 dB SNR and one section had no additive noise. Subjects were played one /CV/ at a time and prompted to



ISBN 1-876346-48-5 © 2003 UAB

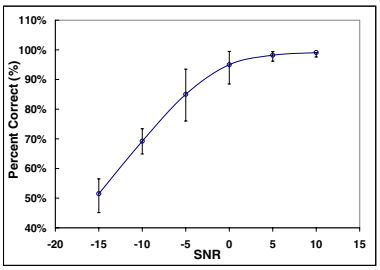


**Figure 1:** Percent correct identification data for /ba,da/ and the sigmoidal fit curve

choose between one of the two consonants. Confusion matrices were then generated.

### 2.4 THRESHOLD DEFINITION

A perceptual threshold is defined by the SNR level (in dB) at which the number of correct responses is at 79 %. A bias in the responses, however, occurred in some experiments. For example, in the /pa,ta/ case, subjects were biased toward /pa/ whenever a confusion occurred. To address this issue, the threshold for one set of /CV/s is determined from the average percent correct of the two /CV/s in each set. For example, the percent correct for the set /ba,da/ will be the average percent correct from /ba/ and /da/. A sigmoidal fit is then performed on the percent correct data to obtain a more accurate threshold value. The sigmoid function had the form:

$$y = max + (max - min)(\frac{1}{2} - \frac{1}{2}(\frac{1 - e^{(\frac{x-b}{a})}}{1 + e^{(\frac{x-b}{a})}})) \quad (1)$$

where $max$ and $min$ are the maximum and minimum values of percent correct responses. In our case, we set these values to 30 and 100, respectively. $a$ and $b$ are parameters that adjust the slope and position of the transition of the sigmoid function between the top and bottom flat areas. These values are determined through an automatic iterative process. The 79% threshold is calculated from the sigmoidal fitting function. From the data in the confusion matrices, we plot percent correct vs. SNR along with the sigmoidal fit. An example of such fitting for /ba,da/ is plotted in Fig 1. The solid line is the sigmoidal fit to the actual percent identification (in circles). The error bars represent the minimum and maximum numbers among the four listening subjects.

### 2.5 EXPERIMENTAL RESULTS

As expected, experimental data showed a decrease in percent correct with decreasing SNR. Percent correct ranged from 100% for clean tokens to about 30-40% at the lowest SNR (-15dB). Results varied from subject to subject and were largely dependent upon the /CV/ pair and vowel context.

Tables 1- 2 summarized the threshold SNR (in dB) for each syllable and the threshold value from the average correct response of the corresponding /CV/ pair (sigmoidal fit was performed on the ratio between the sum of correct responses to both syllables in each /CV/ pair to the total number of tokens played for both syllables). The thresholds in these tables were taken from the average correct responses from all four listeners.

| | /b/ | /d/ | **avg** | /p/ | /t/ | **avg** |
|---|---|---|---|---|---|---|
| /a/ | -7.3 | -7.3 | **-7.3** | -0.0 | 6.9 | **6.7** |
| /i/ | 3.5 | 3.9 | **4.2** | 1.4 | -2.2 | **0.1** |
| /u/ | -1.2 | -0.5 | **-1.6** | 0.1 | -0.2 | **-0.0** |

**Table 1:** Thresholds for each plosive syllable and the thresholds from averaged correct responses (in dB SPL)

| | /f/ | /s/ | **avg** | /v/ | /z/ | **avg** |
|---|---|---|---|---|---|---|
| /a/ | -7.7 | -3.8 | **-5.1** | -7.0 | -1.8 | **-4.5** |
| /i/ | -4.2 | -4.3 | **-3.8** | -1.5 | -1.3 | **-1.2** |
| /u/ | -7.2 | -4.2 | **-5.0** | -1.9 | -3.8 | **-3.4** |

**Table 2:** Thresholds for each fricative syllable and the threshold from averaged correct responses (in dB SPL)

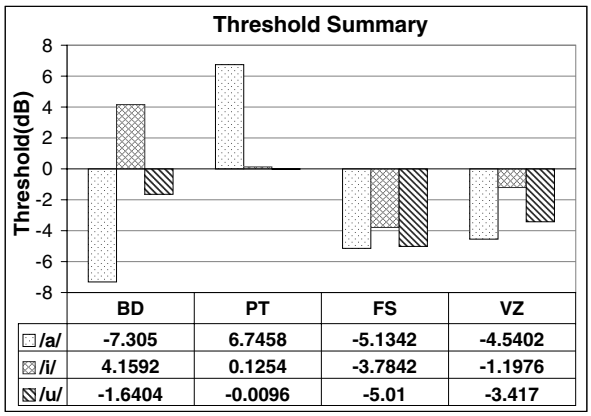


**Figure 2:** Summary of the perceptual thresholds

As mentioned earlier, biasing is a factor for any two forced-choice experiment, thus the average correct responses from each /CV/ pair are used and analyzed to further examine perceptual responses. Threshold values are summarized in Tables 3, 4 and 5 (S1-S4 refer to the different subjects.) Perceptual thresholds are summarized in Figure 2.

**2.5.1 Vowel Context Comparison:** Threshold values across vowel context are summarized in Figure 3. With the exception of /p,t/ the vowel context /a/ was the most robust context (lowest threshold) for all consonant pairs, while the context /i/ was the least robust for distinguishing the labial from alveolar place of articulation.

| | **Avg** | S1 | S2 | S3 | S4 |
|---|---|---|---|---|---|
| /ba,da/ | **-7.3** | -8.9 | -6.4 | -7.8 | -4.9 |
| /pa,ta/ | **6.7** | 3.8 | 5.9 | 4.0 | 8.1 |
| /fa,sa/ | **-5.1** | -4.7 | -5.8 | -5.6 | -4.3 |
| /va,za/ | **-4.5** | -4.0 | -6.7 | -6.3 | -2.8 |

**Table 3:** Thresholds for /Ca/ syllables in dB SPL

| | **Avg** | S1 | S2 | S3 | S4 |
|---|---|---|---|---|---|
| /bi,di/ | **4.2** | 2.8 | 4.8 | 3.5 | 4.0 |
| /pi,ti/ | **0.1** | 2.8 | 1.0 | -2.9 | -1.3 |
| /fi,si/ | **-3.8** | -3.1 | -4.4 | -5.8 | -2.5 |
| /vi,zi/ | **-1.2** | -1.7 | -6.6 | -2.0 | -1.0 |

**Table 4:** Thresholds for /Ci/ syllables in dB SPL

| | **Avg** | S1 | S2 | S3 | S4 |
|---|---|---|---|---|---|
| /bu,du/ | **-1.6** | -3.7 | -2.2 | -0.0 | -2.3 |
| /pu,tu/ | **-0.0** | 0.0 | -2.6 | 1.3 | 1.1 |
| /fu,su/ | **-5.0** | -6.5 | -7.8 | -5.1 | -4.1 |
| /vu,zu/ | **-3.4** | -4.6 | -2.2 | -3.5 | -1.8 |

**Table 5:** Thresholds for /Cu/ syllables in dB SPL

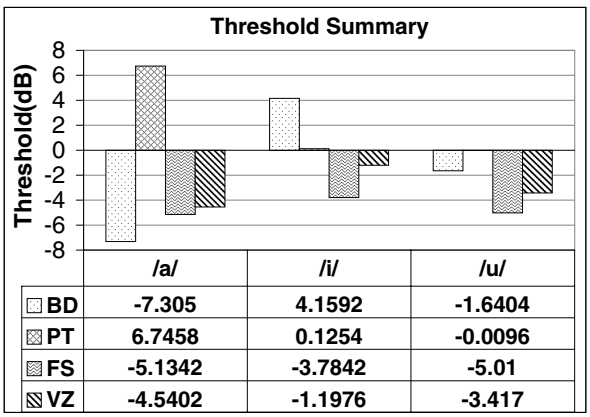


**Figure 3:** Thresholds grouped by vowel

**2.5.2 Effect of the Feature 'Manner':** Perceptual thresholds grouped by manner are shown in Figures 4 and 5 for voiced and voiceless /CV/s, respectively. Plosives are characterized by their bursts and aspiration noise while fricatives are characterized by their longer-duration frication noise.

From the results, we can see that the perception of place for plosives and fricatives depends on whether the consonant is voiced or voiceless. For voiced /CV/s, we observe that plosives are harder to distinguish in the /i/ and /u/ contexts but the opposite is true for /a/. The voiceless pairs, however, have a clear trend. Fricatives consistently have lower thresholds than plosives. Differences in threshold are approximately 11.8 dB, 3.8 dB, and 5 dB for /a/, /i/ and /u/, respectively.

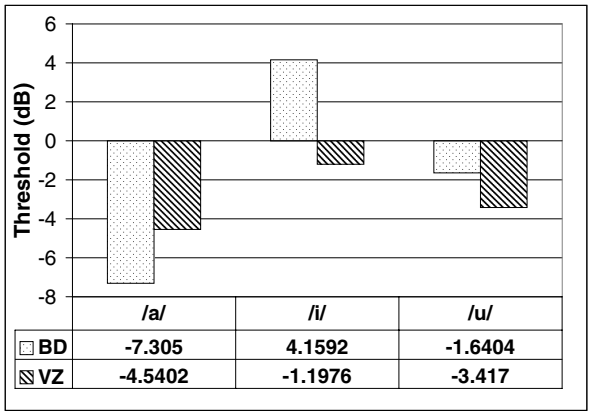


**Figure 4:** Voiced /CV/s grouped by manner

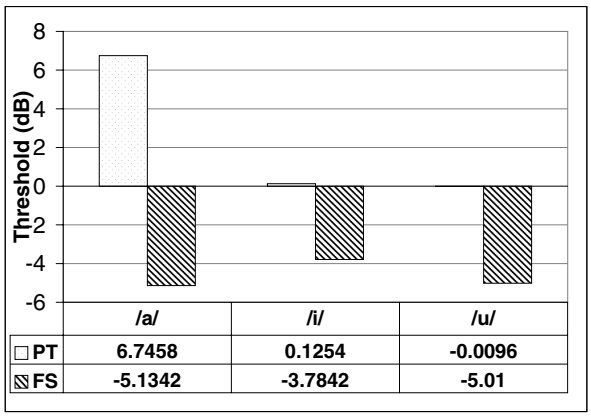


**Figure 5:** Voiceless /CV/s grouped by manner

**2.5.3 Voicing Comparison:** Perceptual thresholds grouped by voicing are shown in Figures 6 and 7.

For plosive /CV/s, voiced pairs have lower thresholds than the voiceless ones except for the /i/ case; this difference was greatest for the /a/ context. A 14 dB difference is observed between /ba,da/ and /pa,ta/. However, /bi,di/ seems to be harder to distinguish than /pi,ti/ since the threshold for /bi,di/ is higher. As for the fricatives, there is a clear trend across all three vowel contexts: voiceless fricatives, /f,s/ perform better than their voiced counterparts /v,z/ by about 1.5 dB on average.

## 3 SUMMARY AND CONCLUSIONS

In this paper, we report on perceptual experiments that attempt to uncover perceptually-relevant cues for the labial and alveolar place of articulation features in noise. Three vowel contexts were examined (/a, i, u/) using both plosives and fricatives.

Results show that the perception of place of articulation in noise is dependent upon the manner, voicing, and vowel context. In general, fricatives are more robust than plosives except for voiced /Ca/ syllables


ISBN 1-876346-48-5 © 2003 UAB

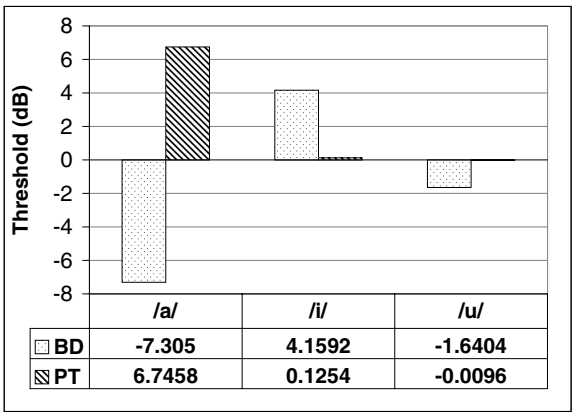


**Figure 6:** Plosive /CV/s grouped by voicing

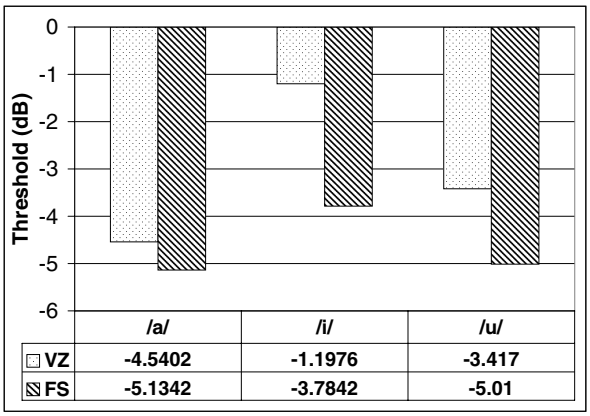


**Figure 7:** Fricative /CV/s grouped by voicing

where the opposite is true. For voiceless consonants, differences in thresholds between fricatives and plosives are, on average, 11.8 dB, 3.8 dB, and 5 dB for /a/, /i/ and /u/ contexts, respectively. This result agrees the Miller and Nicely study [4].

Voiceless consonants are typically more robust than their voiced counterparts, especially for the fricatives (the opposite was found true in [4]). Voicing in plosives does not show a distinct pattern: /ba,da/ are more robust than /pa,ta/ (14 dB difference in thresholds) but /bi,di/ are less robust than /pi,ti/ (by about 4 dB). For /Cu/ syllables, voiced plosives are slightly more robust than the voiceless ones (by 1 dB).

As for context, the /a/ context is more robust than /u/, which is more robust than /i/ except for the /p,t/ case.

Overall, the labial/alveolar distinction was most robust for the syllable pair /ba,da/. The reason may be that the distinction between the formant information in /ba/ and /da/ are most prominent: /ba/ has a rising F2 while /da/ has a falling one. In addition, the onset frequencies for /da/ are higher than /ba/ for F1, F2 and F3. Another observation is that /Ci/ syllables have the lowest perceptual thresholds. One possible reason may be that the frequency extent (transitions from the consonant to the vowel) for /Ci/ syllables are the shortest.

To explore these results further, we conducted a pilot correlation analysis study between acoustical measurements and perceptual results. Since perceptual thresholds depend on manner, voicing, and context, no single acoustic feature could account for the thresholds. For example, for the voiced plosives, the onset frequency of the second formant (F2) was modestly correlated with the perceptual results in only the /Ca/ context. For the /i/ and /u/ contexts, the amplitude of the burst relative to the vowel, and the burst or fricative noise duration showed some correlation with the thresholds. We will report on these studies in a subsequent paper.

## ACKNOWLEDGMENTS

This work was supported in part by NIH-NIDCD grant 1R29-DC02033-01A1. We thank Marcia Chen for her help in data analysis.

## REFERENCES

[1] J. J. Hant, B. Strope, and A. Alwan, "A psychoacoustic model for the noise masking of plosive bursts," *The Journal of The Acoustical Society of America*, vol. 101, no. 5, pp. 2789–2802, 1997.

[2] J. J. Hant and A. Alwan, "A psychoacoustic-masking model to predict the perception of speech-like stimuli in noise," *Speech Communication*, to appear, 2003.

[3] M. Chen and A. Alwan, "On the perception of voicing for plosives in noise," *Proc. EUROSPEECH*, vol. 1, pp. 175–178, 2001.

[4] G.A. Miller and P. E. Nicely, "An analysis of perceptual confusions among some english consonants," *The Journal of The Acoustical Society of America*, vol. 27, no. 2, pp. 338–352, 1955.

[5] S. D. Soli and P. Arabie, "Auditory versus phonetic accounts of observed confusions between consonant phonemes," *Journal of the Acoustical Society of America*, vol. 66, pp. 46–59, 1979.

[6] M. D. Wang and R. C. Bilger, "Consonant confusion in noise: a study of perceptual features," *Journal of the Acoustical Society of America*, vol. 54, pp. 1248–1265, 1973.