Asymmetries in Consonant Confusion

Madelaine Plauché[1], Cristina Delogu[2], and John J. Ohala[1]

[1]University of California at Berkeley, U.S.A, mcp@socrates.berkeley.edu; ohala@cogsci.berkeley.edu
[2]Fondazione Ugo Bordoni, cristina@fub.it

ABSTRACT

Both historical sound change and laboratory confusion studies show strong asymmetries of consonant confusions. In particular, /ki/ commonly changes to /ti/, and /pi/ to /ti/, but not the reverse. It is hypothesized that such asymmetries arise when two sounds are acoustically similar except for one or more differentiating cues, which are subject to a highly directional perceptual error. This perceptual entropy can be explained as follows: if sound x possesses a cue that y lacks, listeners are more likely to miss this "all-or-none" cue than to introduce it spuriously. /k/ and /t/ before /i/ have similar formant transitions but differ in their burst spectra. /p/ and /t/ before /i/ also have similar formant transitions but differ in the intensity of their bursts. The importance of these differentiating features for listeners' perception were verified in a confusion study. The implications of the inversely related effects of perceptual and physical entropy for phonetic theory and speech technology is discussed.

1. INTRODUCTION

Asymmetrical confusions occur in many different perceptual modalities. A useful analogy to the acoustic asymmetries of consonants studied here is found in the realm of visual perception. In a visual perception study by Gilmore et al. [3] subjects confused certain Roman capital letters more often in one direction than in the other, as shown below (Fig. 1).

Q→O	E→F	R→P
B→P	P→F	J→I

Figure 1. Roman capital letter confusions. '→' indicates the direction of the confusion. Thus 'Q→O' indicates that 'Q' is confused for 'O' but 'O' is rarely confused for 'Q.'

In each of these cases, the letters on the left are identical to those on the right except that they contain an additional distinguishing feature. This agrees with Garner's claim [2] that this 'all-or-none' feature was more likely to be accidentally missed than to be added in.

An acoustic perception study run on Italian speakers by Delogu [1] showed a number of asymmetrical confusions in identifying CV sequences in both synthetic and natural speech. In particular, [ki] and [pi] were found to be confused for /ti/ but never the reverse. Her results for the confusions between the voiceless, non-aspirated stops /p/, /t/, and /k/ before the vowel, /i/, are shown below (Fig. 2).

ki→ti	50%	pi→ti	31%
ti→ki	0%	ti→pi	6%

Figure 2. Confusion rates between [pi], [ti], and [ki] for SNR at 0 dB.

These asymmetries parallel widespread historical sound changes, /ki/ > /tʃi/ and /p^j/ > /t/, and suggest there might be an all-or-none acoustic feature that distinguishes each pair of segments [4]. This study attempts to isolate the specific acoustic features that cause these asymmetrical confusions and to test their importance in perception.

2. ISOLATION OF THE 'ALL-OR-NONE' FEATURE

If the all-or-none features responsible for the asymmetries described above were removed, a perception study similar to that of Delogu should show a substantial increase in the rates of confusion between the pairs of segments.

2.1. The CV sequences ki and ti

A candidate for the all-or-none feature causing the asymmetric confusion between [ki] and [ti] has been found in the smoothed spectra of the voiced equivalents of these CV sequences. Stevens and Blumstein [5] found that the spectra for these two segments are similar except for a sharp peak in the 3-4 kHz of the [gi]. Ohala [4] proposed that this spectral burst might be missed by listeners, but would rarely be added in.

To determine the importance of this acoustic cue, we filtered the [ki] stimuli recorded from a Spanish speaker through a Hanning filter of order 10 between the frequencies of 2.5 and 5 kHz. We then spliced this filtered burst onto a full [ki] vowel to avoid distortions caused by the filtered vowel. The spectral burst of the filtered [ki] stimuli is comparable to that of a [ti] from the same speaker (Fig. 3).

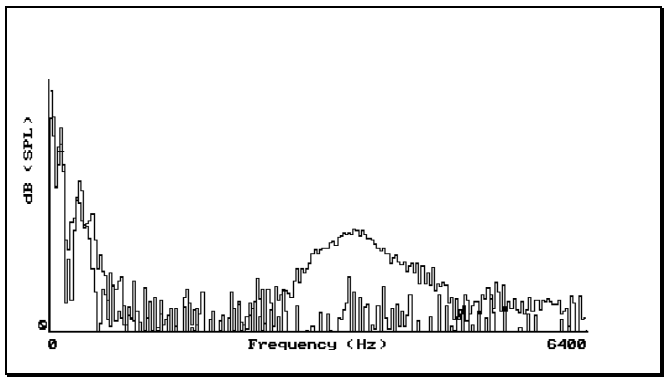

Figure 3. Spectral bursts of [ki] and [ti].

2.2. The CV sequences pi and ti

The syllables [pi] and [ti] are similar except for the intensity of the burst. Zue [7] found that "the spectra of p, b show no distinct burst frequency and the MS amplitudes of these stops are weak." In order to test the importance of amplitude in the perception of these segments, we doubled the amplitude of a [pi] syllable of the same Spanish speaker as above and spliced it onto an unmodified [pi] vowel. The burst of the modified [pi] is comparable to that of a [ti] (Fig. 4, 5, 6).

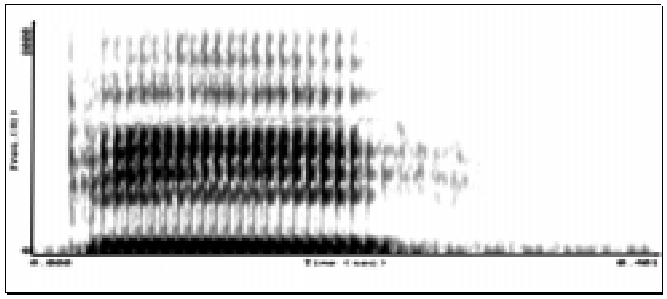

Figure 4. Spectrogram of [pi].

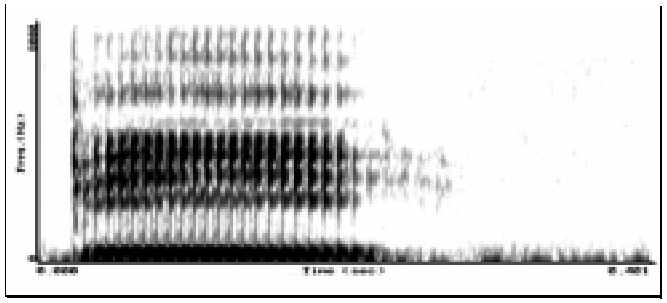

Figure 5. Spectrogram of the enhanced [pi].

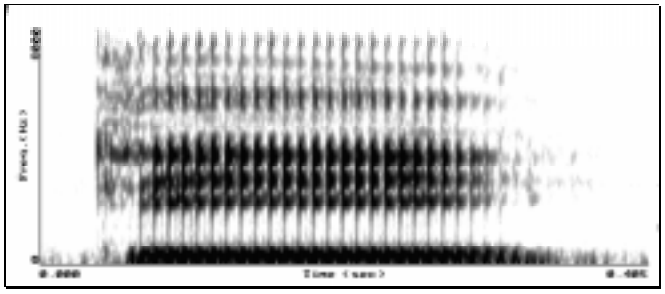

Figure 6. Spectrogram of [ti].

3. THE CONFUSION STUDY

3.1. Hypothesis

We hypothesize that [ki] and [pi] will be often misperceived as [ti], but that the reverse will not be true. We also expected to find that the segments missing the distinguishing features, filtered [ki] and enhanced [pi], are misperceived as [ti] at a higher rate than their unmodified counterparts.

3.2. Method

3.2.1. Subjects

Fifteen native Spanish speakers from various backgrounds, aged 20-30 years, were participated in this study. Spanish was used as the language for this study because its voiceless stops have low aspiration, similar to Italian.

3.2.2. Apparatus and Materials

Stimuli were recorded and prepared using Computerized Speech Lab (CSL). Data was presented and results were recorded with an Auditory Perception Program and Database.

3.2.3 Design

The test was composed of one training session and three test sessions with the number of trials indicated below (Fig. 7).

Training	A-Test	I-Test	U-Test
pa-4	pa-8	pi-6	pu-6
ta-4	ta-6	ti-4	tu-6
ka-4	ka-8	ki-6	ku-8
		filtered ki-2	
		enhanced pi-2	

Figure 7. Perception trials.

Each of the stimuli were randomized in preparation for the confusion study. The I-Test session was subsequently

adjusted to evenly distribute the modified speech segments, filtered [ki] and enhanced [pi], throughout the test session.

As the segments were played, the computer screen presented three buttons labeled 'ka,' 'ta,' 'pa,' in the first and training session, 'ki,' 'ti,' 'pi' in the second, and 'ku,' 'tu,' 'pu' in the third.

3.2.4. Procedure

The subjects were instructed to identify the consonant vowel sequence as quickly as possible, by clicking on one of three options presented on a computer screen. Although their reaction times were not recorded, they were told to answer as quickly as possible.

4. RESULTS AND DISCUSSION

4.1. Results

Below are the relevant results of the perception study described in the previous section (Fig. 4)[1]. The unfiltered [ki] is confused with [ti] 20% of the time, but once its spectral burst has been filtered, this confusion rate is increased to 100%. The enhanced [pi] is also confused for [ti] 100%.

Stimuli/ Response	pi	ti	ki
pi	**97%**	3%	0
ti	0	**100%**	0
ki	0	20%	**80%**
FLTRKI	0	**100%**	0
BIGPI	0	**100%**	0

Figure 8. I-test confusions.

4.2. Discussion

These results support two important claims. First, the asymmetries of confusion claimed by Winitiz et al. [6] and Delogu [1] are verified. Second, the CV sequences /ki/ and /pi/ are always confused for /ti/ once the distinguishing feature has been removed.

Clearly, ki→ti is an asymmetric confusion caused by the acoustic similarity of these two segments, which differ only in that [ki] includes a spectral burst in the 3-4 kHz region. /p/ and /t/ before /i/ also have similar formant transitions but differ in the intensity of their bursts: /p/ has a burst approximately half the intensity of /t/. The hypothesis above would suggest the asymmetry ti→pi, where the energy of the burst of /t/ is missed more often than introduced by listeners. Confusion studies of these consonants instead show the opposite: pi→ti. This asymmetry increases when the S/N ratio is increased.

In order to explain this apparent "exception" to our hypothesis consider how features differentiating certain phonemic contrasts may be placed on an entropy continuum (Figure 9).

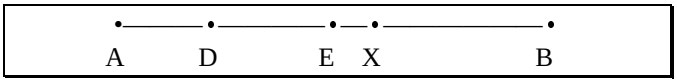

Fig. 9. Schematic representation of the entropy of a given feature differentiating two speech sounds. See text for details.

The entropy continuum of the feature is represented as a line AB, where A is the extreme of low entropy and B is the extreme of entropy, i.e., complete randomness or chaos. Somewhere in this continuum is placed a boundary, X, which is the dividing line between a given sound and another which the feature serves to differentiate it from. The entropy of a feature is inversely related to its acoustic-auditory salience and its susceptibility to variation. The natural direction of shift of this feature (insofar as it has a tendency to vary) is towards B. An acoustically and perceptually robust feature (D), would have an entropy value that is low, i.e., close to A. A feature that is not robust would have a value closer to B (although to the left of the boundary when it is clearly articulated), i.e., close to E.

The narrow band spectral peak in the burst of velar stops is posited to have relatively high entropy, that is, close to E on the line in Fig. 9. Thus it is subject to an increase in entropy which has the effect of shifting it past the boundary X which, in turn, means it is subject to confusion as the structurally similar sound which lacks this feature. This is the basis for the change /ki/ > /ti/. In the case of /ti/ the intense noise burst is posited to be very low in entropy, i.e. quite robust and stable, that is, close to D on the line in Fig. 9. Even though this feature is also subject to some entropy increase, it is unlikely to increase enough to have the feature pass the boundary X and thus make the given sound be confusable with /pi/. Thus it is unlikely that we would encounter the confusion /ti/ >/pi/. In the case of /pi/, what differentiates it from /ti/ is the lack of an intense burst but this is posited to be a highly variable feature, i.e., close to E in Fig. 9. Depending on the degree of closeness of the following /i/ vowel and the rate of airflow, the burst may increase, cross the boundary X and thus be confusable with /ti/.

[1] Confusion rates from the A-test and U-test were negligible.

5. CONCLUSION

This study offers insight into the acoustic structure of these particular segments as well as a method for explaining and predicting confusions found in synthetic and natural speech. The results of this confusion study also parallel important historical sound changes between velars and apicals as well as between labials and apicals as shown below (Fig. 10).

Old English	English	Gloss
[wæ**kki**ng(e)]	[wa**tʃi**ŋ]	"watching"
Standard Czech	East Bohemian	Gloss
[**pj**et]	[**t**et]	"five"

Figure 10. Historical sound changes that reflect the asymmetries ki→ti and pi→ti.

The parallels between these historical sound changes and this asymmetrical perception study support the claim that sound changes can result from speaker misperceptions and suggest that diachronic sound changes can be explored in the laboratory.

6. REFERENCES

[1] C. Delogu, A. Paolini, P. Ridolfi, and K. Vagges, *"Intelligibility of Speech Produced by Text-to-Speech Systems in Good and Telephonic Condidtions"*, Acta Acustica 3, pp. 89-96, 1995.

[2] W. R. Garner, *"Aspects of a Stimulus: Features, Dimensions and Configurations"*, in E. Rosch & B. B. Lloyd (eds.), Cognition and Categorization, Hillsdale, N. J., Erlbaum, pp. 99-133, 1978.

[3] G. C. Gilmore, H. Hersh, A. Caramazza, and J. Griffin, *"Multidimensional Letter Similarity Derived from Recognition Errors"*, Perception & Psychophysics, Vol. 25, pp. 425-431, 1979.

[4] J. J. Ohala, *"Linguistics and Automatic Processing of Speech"*, in R. De Mori & C. Y. Suen (eds.), New Systems and Architectures for Automatic Speech Recognition and Synthesis, [NATO ASI Series, Series F: Computer and System Sciences, Vol. 16], Berlin Heidelberg, Springer-Verlag, Vol. F16, p. 447-475, 1985.

[5] K. N. Stevens and S. E. Blumstein, *"Invariant cues for Place of Articulation in Stop Consonants"*, Journal of Acoustical Society of America, Vol. 40, pp. 123-132, 1966.

[6] H. Winitiz, M. E.Scheib and J. A. Reeds, *"Identification of Stops and Vowels for the Burst Portion of /p,t,k/ Isolated for Conversational Speech"*, Journal of the Acoustical Society of America, Vol. 51, pp. 1309-1317, 1972.

[7] V. W. Zue, "Acoustic Characteristics of Stop Consonants: A Controlled Study", Indiana University Linguistics Club, Indiana, 1976.