Dept. for Speech, Music and Hearing
Quarterly Progress and Status Report

Consonant confusions in English and Swedish. A pilot study

Fant, G. and Lindblom, B. and de Serpa-Leitao, A.

KTH Computer Science and Communication

journal: STL-QPSR
volume: 7
number: 4
year: 1966
pages: 031-034

http://www.speech.kth.se/qpsr

III. SPEECH PERCEPTION

A. CONSONANT CONFUSIONS IN ENGLISH AND SWEDISH, - A Pilot Study

G. Fant, B. Lindblom, and A. De Serpa-Leitão

Introduction

Our study was initiated with three questions in mind.

(1) What are the relative importance of various cues underlying a specific phonemic distinction?

(2) What are the rough perceptual distances between a set of consonants?

(3) Are English consonants more susceptible to distortion than Swedish consonants?

Confusion tests can provide some insight in all the three problems.

For evaluation of cues (1) we need a set of varied and highly selective distortions. The perceptual distances (2) should in general be defined for a situation without distortion but can also be defined with reference to some specific type of distortion. In the latter case it is of course necessary to state whether subjects have required a learning under the specific conditions or not. The question (3), whether any language is more difficult than others to perceive under specific conditions of distortion has no simple answer either. In a rhyme test the contextual redundandy is minimal. This is a suitable starting point.

In order to get some insight into this problem and some experience of methodology we designed a confusion test limited to one subject only who served both as speaker and listener. Our subject (HW), a teacher of English, is bilingual in the sense that he has, and has had, equal command of British English and Swedish since childhood. The purpose of this pilot study was to analyze the observed confusions in terms of the specific distortions involved and to interpret them in relation to spectrographic data.

Experiment

The experimental approach was that once introduced by Miller and Nicely [1]. We constructed nonsense syllable rhyme words of CV structure with V = [a:]. The set of words for the English test comprised all 22 possible consonant phonemes in initial position plus the voiced, flat, continuant fricative /ʒ/. The Miller and Nicely study was limited to 16 consonants. In the Swedish material we used all 17 possible initial single consonants.

For each language 10 randomized word lists were compiled. After recording in an anechoic chamber the entire speech material was submitted to the following processing:

(1) No distortion.

(2) Low-pass filtering at 2000 c/s with a high quality filter (60 dB per half octave).

(3) White noise added to give a S/N of 13 dB. The speech materials were played over a high-quality loudspeaker to the subject now serving as a listener.

Hi-Fi condition

It is of interest to note that the performance under condition (1) was not 100 per cent. The Swedish palatal continuant [ç] was once heard as the retroflex continuant [ʃ]. In the English material 3 out of 10 interdental unvoiced fricatives [θ] were heard as [f] and one such confusion occurred in the corresponding voiced pairs of [ð] and [v]. These distinctions are not very effectively maintained in speech.

LP filtering

Under condition (2) of low-pass filtering at 2000 c/s, see Fig. III-A-1, there is a clear tendency that dental stops and fricatives are almost never recognized as such. In the Swedish and the English material [t] is heard as [k] to almost 100 per cent, as can be seen in the confusion matrices of Fig. III-A-1 and III-A-2. In Swedish [d] is accordingly heard as [g] whilst the English [d] proves to be resistant. None of the Swedish [s] sounds nor the English [s] and [z] are heard as such, the Swedish [s] being confused with [ʃ], whereas in English [s] is mostly confused with [θ], and [z] with [ð]. The confusion of [ð] with [v] and [θ] with [f] is also typical.

The spectrograms of Figs. III-A-3, III-A-4, and III-A-5 provide a basis for discussion of confusion risks. Fig. III-A-3 illustrates the effect of filtering on the Swedish unvoiced stops. The removal of energy above 2000 c/s curtails the burst spectrum of [t] in [ta] so that it resembles that of the [k] in [ka] before filtering. The [k] loses very little of its energy by the filtering and the vowel [a] is also essentially intact since the loss of its third and higher formants has a very small perceptual effect. The F_2 transitional cues are almost identical for [ta] and [ka] and we infer that, normally, the burst provides a distinguishing cue.

Subject H.W., Swedish, LP 2000

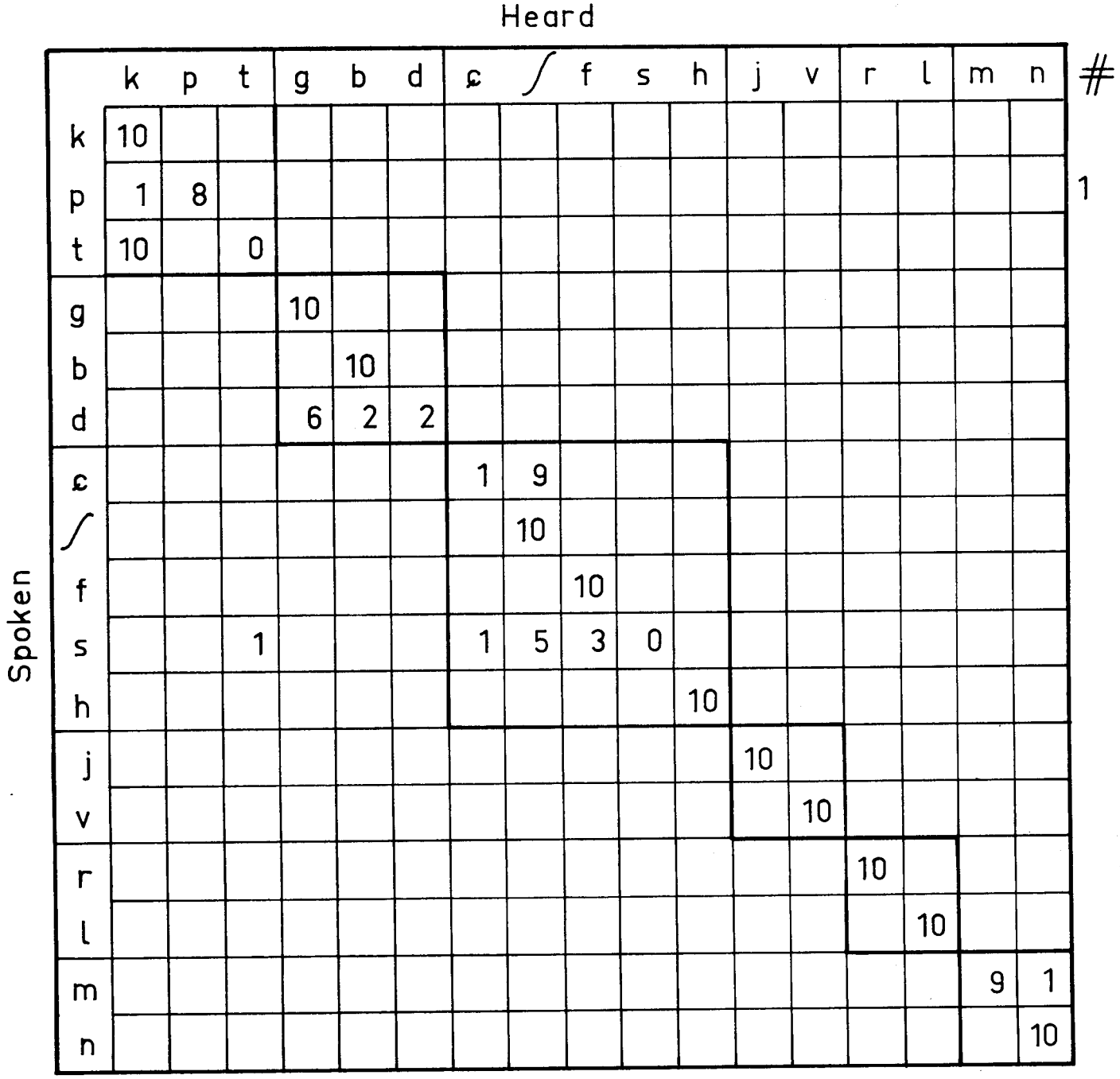

Fig. III-A-1. Confusion matrix of Swedish consonants in a Ca frame. One subject serving both as talker and listener. Low-pass filtering at 2000 c/s.

Subject H.W., English LP 2000

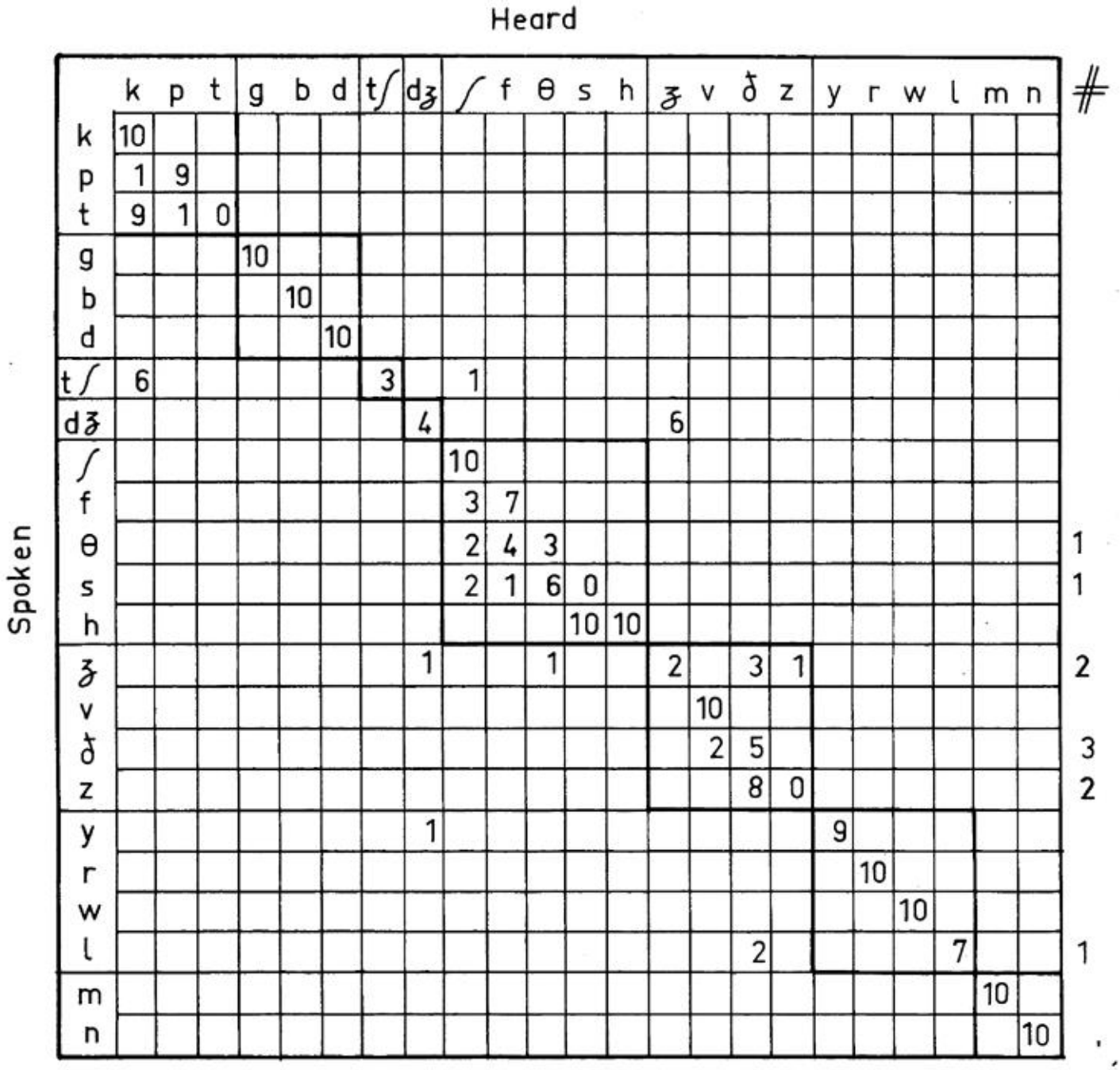

Fig. III-A-2. Confusion matrix of English consonants. Conditions same as those of Fig. III-A-1.

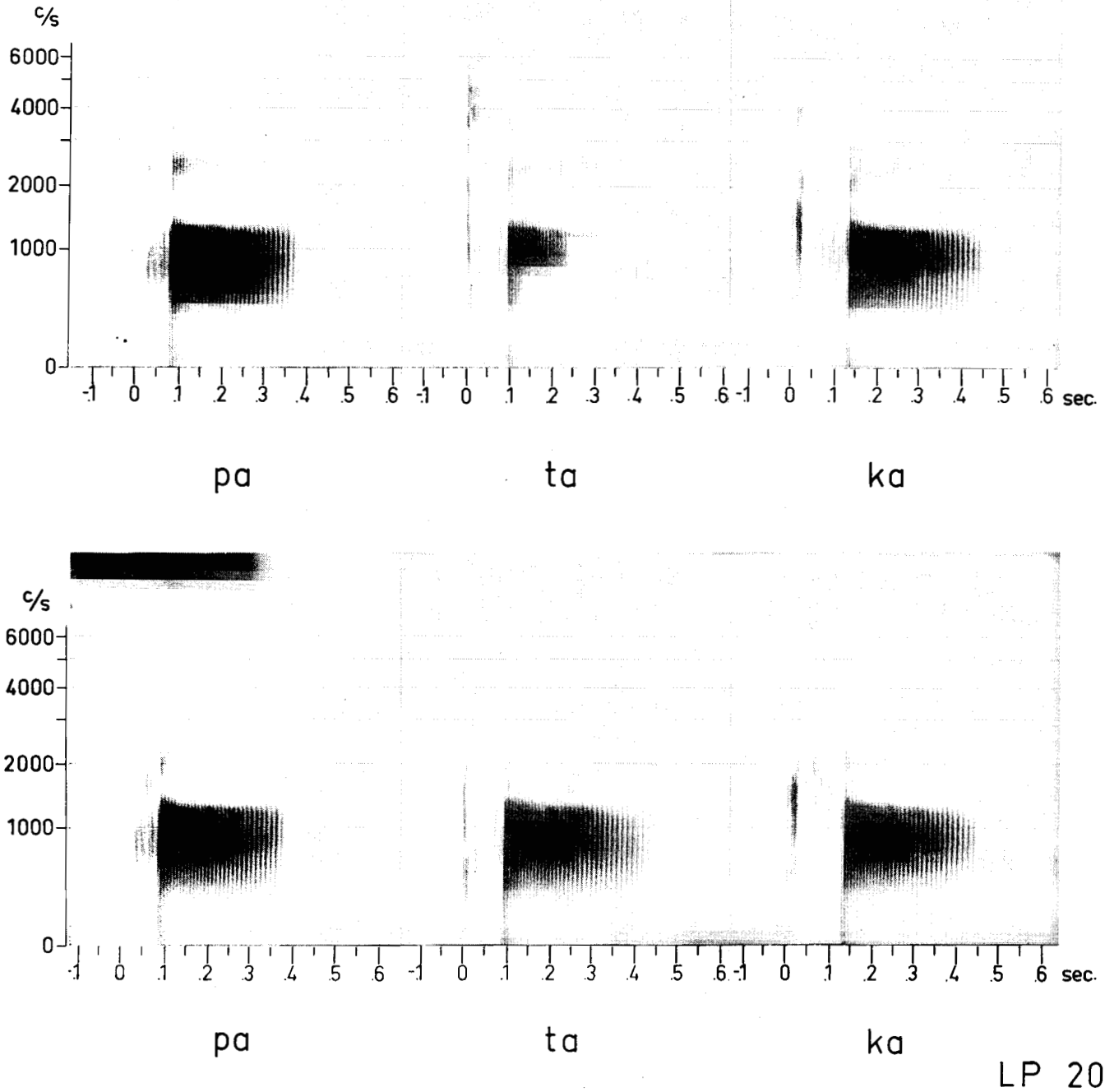

Fig. III-A-3. Spectrograms of Swedish stops before and after low-pass filtering 2000 c/s. Observe the technical mel scale along the ordinate.

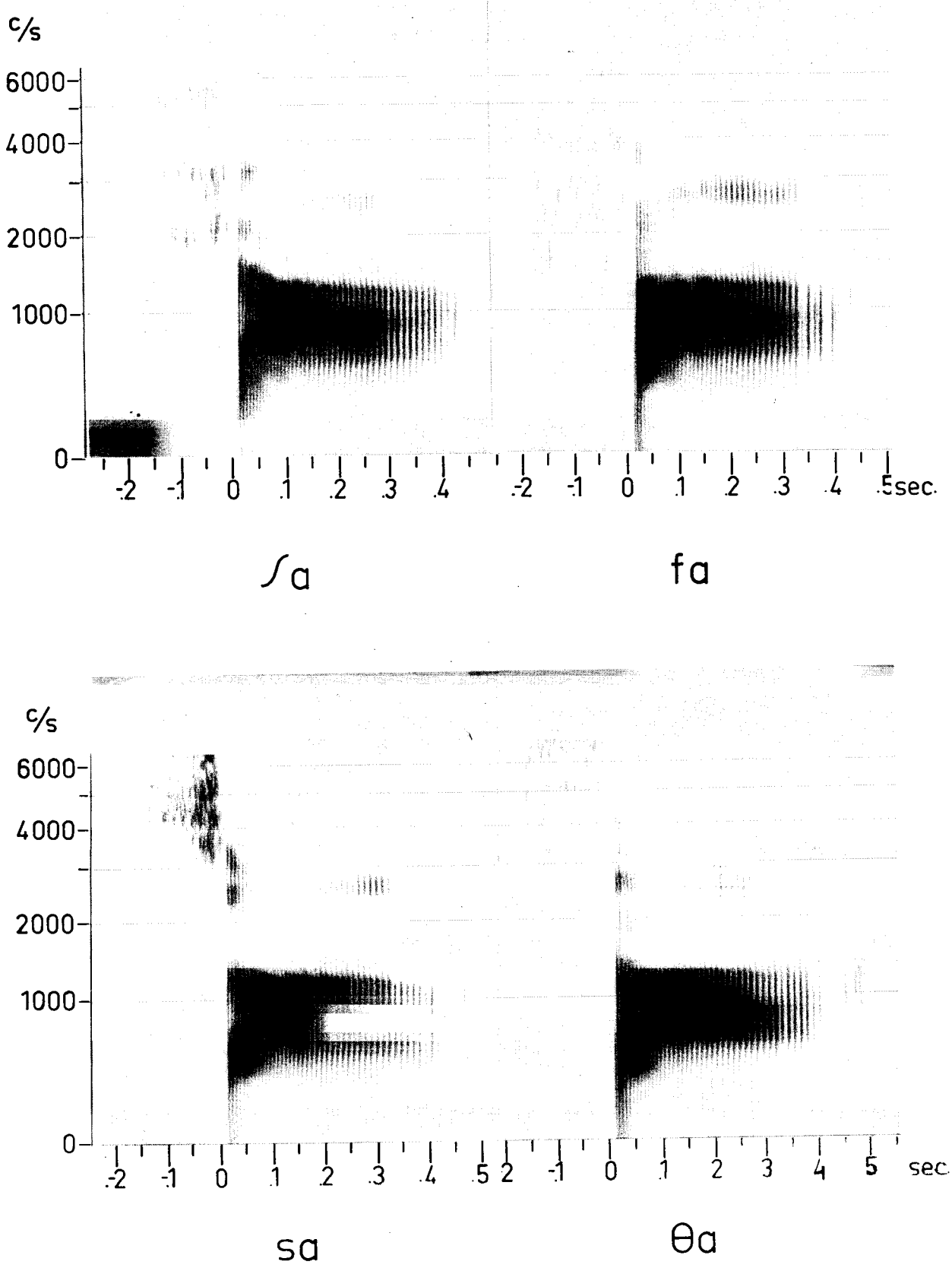

Fig. III-A-4. Spectrograms of English unvoiced fricatives.

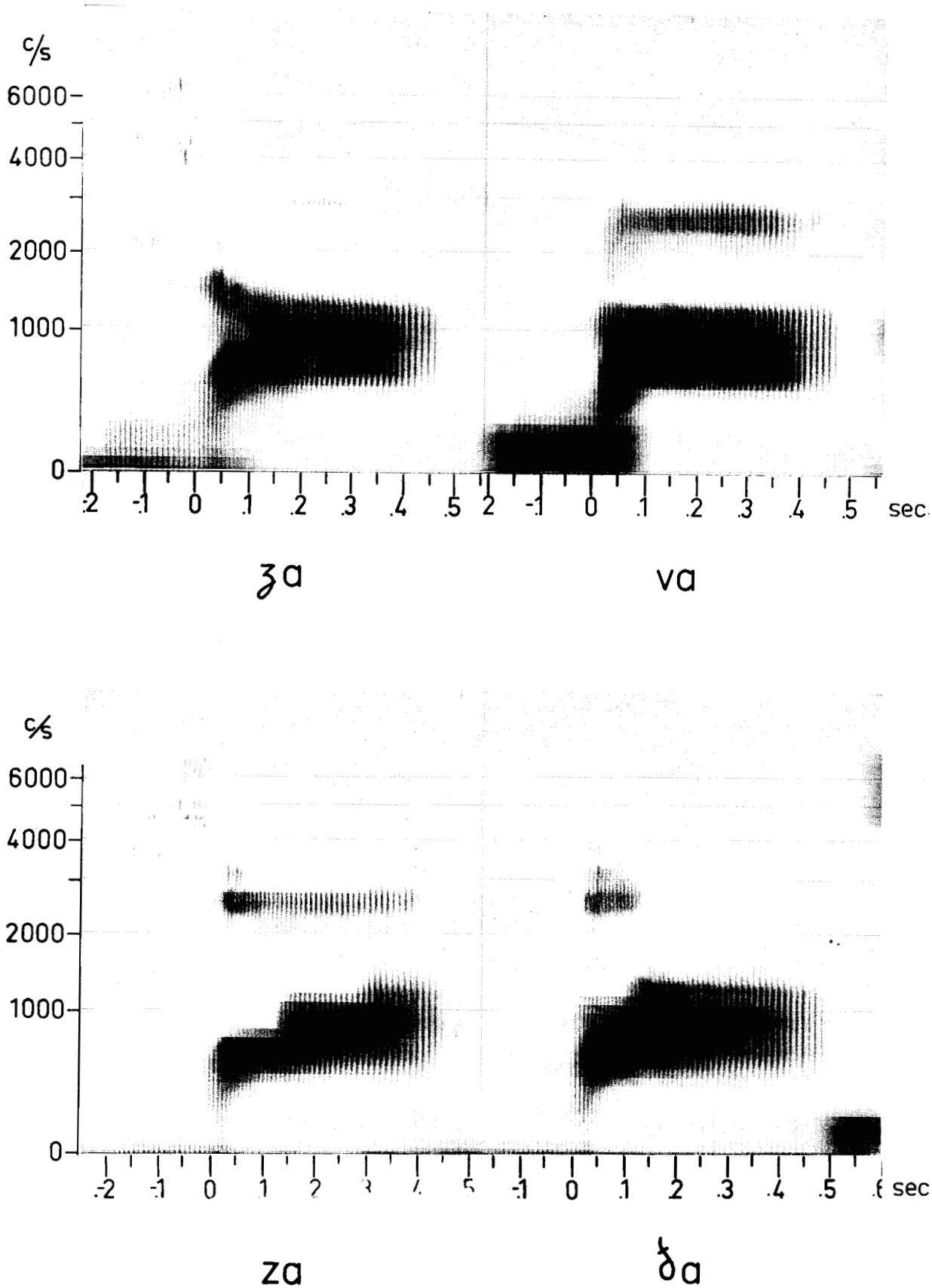

Fig. III-A-5. Spectrograms of English voiced fricatives.

This is a methodological example of how filtering may be applied as a means of separating transitional and segmental cues.

However, the spectrograms indicate that there is some small difference in the [ka] and [ta] after low-pass filtering which may serve as a means of discrimination after sufficient learning.

A study of the unvoiced English fricatives in Fig. III-A-4 reveals an approximate similarity of [ʃa] and [sa] in terms of transitional cues as well as in terms of intensity of the noise segment. As is well known they differ in the noise spectrum which has a lower bound of 2000 c/s in [ʃa] and 4000 c/s in [sa]. The [fa] and the [θa] both have very weak noise segments, the [θa] being characterized by a greater emphasis of the noise at higher frequencies and a higher F_2-locus. The F_2-locus of [θa] is quite similar to that of [sa] which in turn is slightly lower than the F_2-locus of [ʃa]. The [sa] after filtering at low-pass 2000 c/s loses most of its noise energy and the confusion with [θa] is therefore quite natural. [sa] was also heard as [ʃa] once and as [fa] once, all of which is quite plausible in view of the spectrographic patterns.

The voiced fricatives of Fig. III-A-5 display the same relations. In addition there is to be seen the superimposed voicing striations in all fricative segments. A typical feature of voiced weak fricatives as [v] and [ð] is that the noise energy is even weaker than in corresponding unvoiced fricatives [f] and [θ] and therefore below the threshold of spectrographic marking. However, all the 10 [va] syllables were correctly identified, whereas 3 of the 9 [fa] were heard as [ʃa]. The higher resistance of [va] to degrading is probably due to the fact that the transitional cue extends further into the consonant segment than in [fa].

Of the affricates the unvoiced [tʃ] tended to be heard as [k] when low-pass filtered and the voiced [dʒ] was received as [ʒ]. The glides [y] and [w] and the [r] were all correctly identified after filtering whereas the English [l] was heard as [ð] twice.

Noise condition

The effect of the noise in experiment (3) was not as drastic as that of the filtering. The reason is the relatively low noise level, 13 dB below the average speech level. As can be seen in the confusion matrices of Figs. III-A-6 and III-A-7 the noise affects the [p] which is heard as [h]. The interdental [θ] is heard as [f] and [ð] as [v] a few times.

Subject H.W Swedish S/N=13 dB

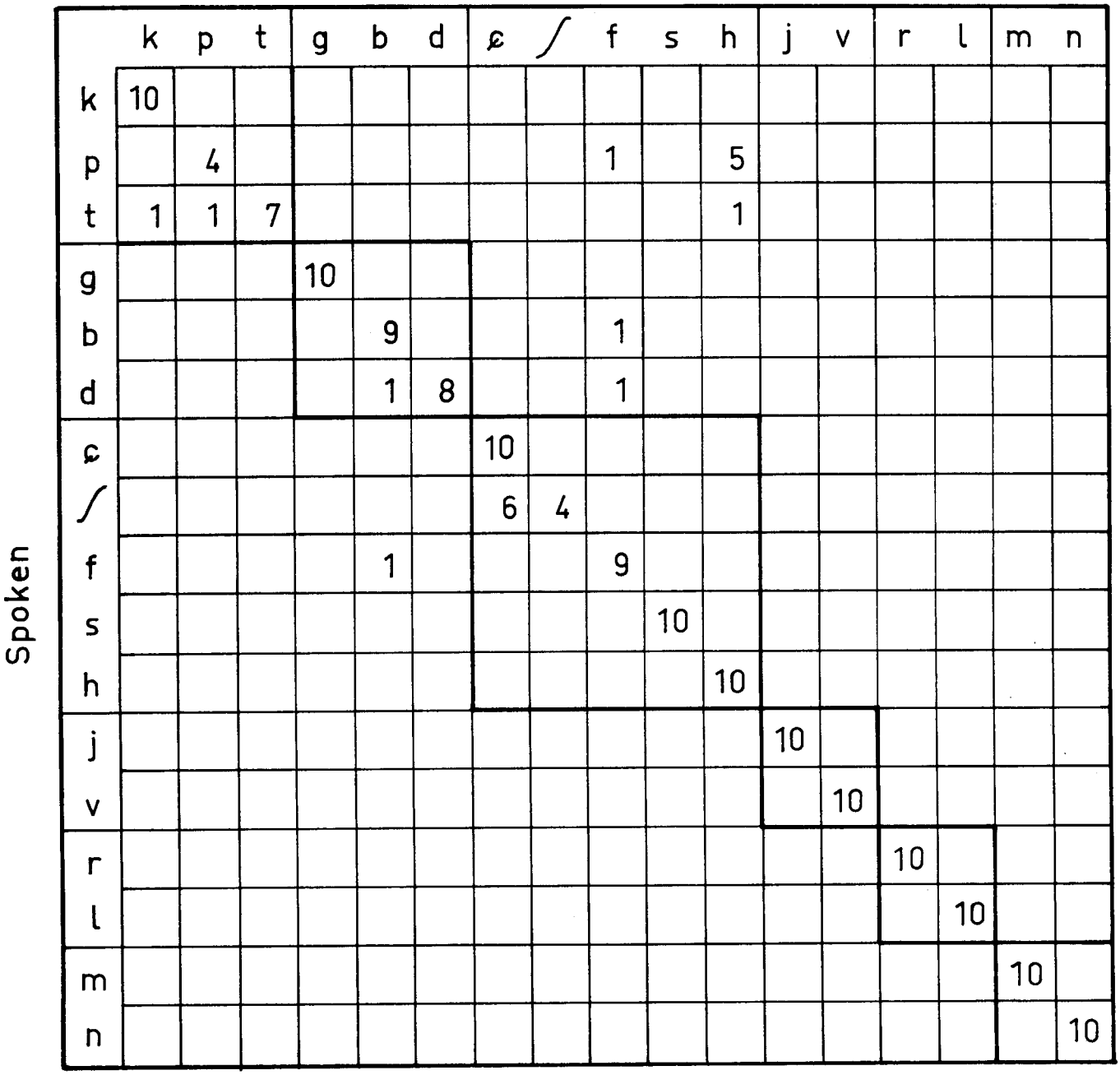

Fig. III-A-6. Confusion matrix of Swedish consonants. Noise of -13 dB relative signal level.

Subject H.W., English S/N = 13 dB

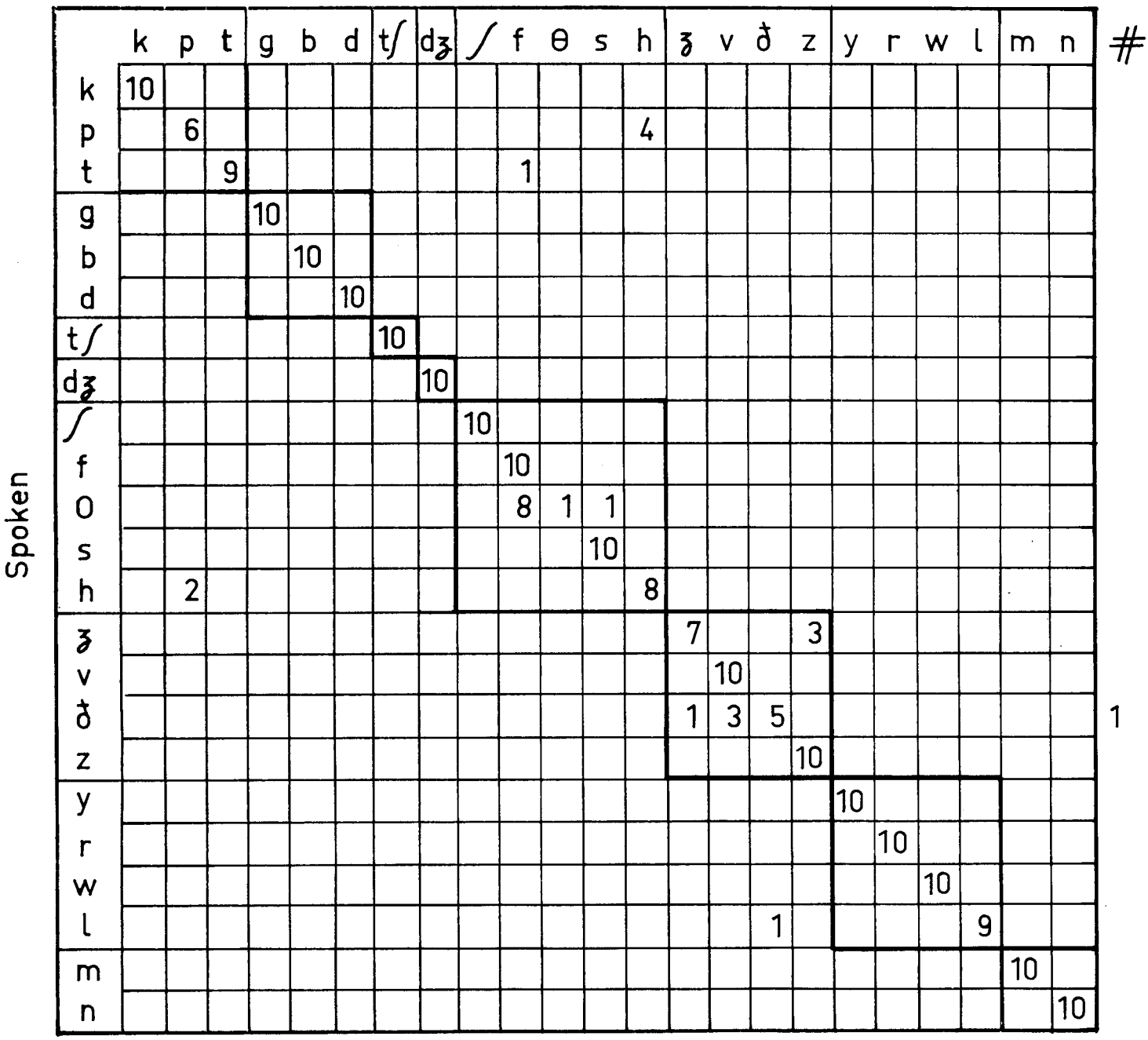

Fig. III-A-7. Confusion matrix of English consonants.
Noise of -13 dB relative signal level.

Discussion

There remain a few general statements to be made. The number of consonant confusions in the low-pass filter test was about twice as frequent in English as in Swedish. This is primarily the consequence of the larger ensemble of English consonants, 23 compared with 17 for Swedish. Within any subclass of consonants the number of confusions is about the same in Swedish and English. Consonants produced with different manner of production are more seldom confused. Nearly all the confusions are in terms of different places of articulation within a subclass of constant manner of articulation, as pointed out by Miller and Nicely [1] and indicated by squares along the diagonals of our confusion matrices. The effect of noise differs from that of filtering by affecting the identification of manner of articulation just as much as the place of articulation.

Attention is drawn to the very extensive confusion tests in auditoria reported by Ormestad [2]. His work is concerned with 4 Norwegian dialects. The most frequent confusion he found was between [b] and [v]. Echo effects and reverberation tend to destroy some of the manner cues in speech whereas filtering typically affects the place cues.

Finally it should be pointed out that the very specific confusions we have noted with our subject (HW) to some extent reflect his specific speech habits rather than an average speaker listener behavior. Thus his Swedish [ç] and [ʃ] were less contrastive than is generally found.

A confusion test undoubtedly has a specific diagnostic value and should be used more extensively in studies of communication systems as well as of speech and hearing defects. It is certainly a valuable technique for evaluating the perceptional significance of spectrographic data.

References:

(1) Miller, G.A. and Nicely, P.E.: "An Analysis of Perceptual Confusions Among Some English Consonants", J. Acoust. Soc. Am. 27 (1955), pp. 338-352.

(2) Ormestad, H.: Høreskarphet og taletydlighet og forståeligheten av norske språklyder (Akademisk Forlag, Oslo 1955).