

THE RELATION BETWEEN THE ERROR DISTRIBUTION AND THE ERROR RATE IN IDENTIFICATION EXPERIMENTS*

R.J.J.H. van Son

Abstract

The error dispersion is a perplexity related measure of the effective number of error categories used by listeners in an identification experiment. The error difference is calculated from the error dispersions of several confusion matrices and measures the difference between these confusion matrices. The error dispersion and the error difference are largely independent of the error rate. It is shown in this paper that correlations (and the absence thereof) between the error dispersion and the error rate that are found in identification experiments can be used to infer underlying regularities in the identification process. The use of these techniques is demonstrated with examples from the literature.

1. Introduction

For analyzing the results of identification experiments two different approaches are used. The simplest type of analysis uses the number of incorrect identifications, i.e., the error rate, to compare results from different experiments or for different conditions. The second, much more elaborate method uses Multi-Dimensional scaling to analyse the internal structure of the "response-space". Both methods have their strengths and weaknesses. The error rate is conceptually the simplest measure and can be used to quantify the differences between experiments and conditions, but it uses only the *correct* responses and ignores the *incorrect* ones. Multi-Dimensional scaling techniques can unravel intricate relations between the acoustic structure of stimuli, experimental conditions, and perceptual categories. However, these techniques can be used only on "full" matrices (not too low an error rate) whereas asymmetry is hard to handle which limits its use. Moreover, the level of detail given is often too high and it is difficult to quantitatively compare the results of different experiments. As a consequence, the result of a multi-dimensional scaling analysis needs a lot of "post-processing" to extract the relevant features. Beside these two broad approaches, there are other, more specialised methods of analysis, e.g., distance metrics (d' , see e.g., Van Wieringen, 1995), entropy related measures (Miller and Nicely, 1955; Allen, 1994), or articulation density (Allen, 1994). These approaches have in common that they are generally too specialised for the average identification experiment (e.g., d' and articulation density) or interpretation itself is non-trivial (entropy).

* A preliminary version of this paper was presented at Eurospeech 95 in Madrid (Van Son, 1995).

In many experiments, it is often the distribution of the errors, and not their number, that is the meaningful measure. However, at the same time there is no need for detailed structural information on the distribution of the responses. It can be seen from the previous description that there is a need for analysis tools of intermediate complexity. More detailed than the error rate, but more global than scaling techniques.

This leads us to the following two questions. First, how can we quantify in a meaningful way the distribution of errors in an identification experiment? Second, can we quantify the differences between individual error distributions?

2. The perplexity

The approach taken here to answer the questions posed in the previous section is based on standard Information Theory (Khinchin, 1957; Sveshnikov, 1968; Press et al., 1988; Allen, 1994). A very useful method to analyse stimulus-response relations is to use the perplexity. The perplexity is primarily used to describe syntax complexity in automatic speech recognition (Bahl and Jelinek, 1990). In the context of this paper the perplexity is the “effective” number of responses per stimulus or stimuli per response. The perplexity is calculated from the basic entropy values of a confusion matrix (see figure 1). Based on these basic entropies the conditional entropies and perplexities are defined as respectively equations 1 and 2 (Press et al., 1988; Bahl and Jelinek, 1990):

$$\begin{aligned} H(\text{Resp|Stim}) &= H(\text{Stim, Resp}) - H(\text{Stim}) \\ H(\text{Stim|Resp}) &= H(\text{Stim, Resp}) - H(\text{Resp}) \end{aligned} \quad (1)$$

$$\begin{aligned} \text{Perplexity}(\text{Stim}) &= 2^{H(\text{Resp|Stim})} \\ \text{Perplexity}(\text{Resp}) &= 2^{H(\text{Stim|Resp})} \end{aligned} \quad (2)$$

3. The error dispersion, d

The perplexity is dominated by the number of *correct* responses, i.e., $1 - \text{error rate}$. However, the aim is to describe the *distribution* of the *incorrect* responses. This can be done by “normalizing” the perplexity for the error rate. The result is called the error dispersion, d (equation 3, see Van Son, 1994, 1995 for a derivation).

$$\begin{aligned} d_s &= 2^{h_s}; h_s = \frac{H(\text{Resp|Stim}) - H_\epsilon}{\epsilon} \\ d_r &= 2^{h_r}; h_r = \frac{H(\text{Stim|Resp}) - H_\epsilon}{\epsilon} \end{aligned} \quad (3)$$

With the entropy of the error-rate defined as:

$$H_\epsilon = -\epsilon \cdot 2^{\log(\epsilon)} - (1-\epsilon) \cdot 2^{\log(1-\epsilon)}$$

The error dispersion, d, is the perplexity with the correct responses “removed”. It indicates the effective number of error categories per either stimulus (d_s) or response (d_r).

Mathematically speaking, the error dispersion is largely independent of the absolute error rate. There remains a small indirect dependence on the *distribution* of the correct responses. If the distribution of correct responses is unbalanced, i.e., errors

Confusion Matrix

Entropy

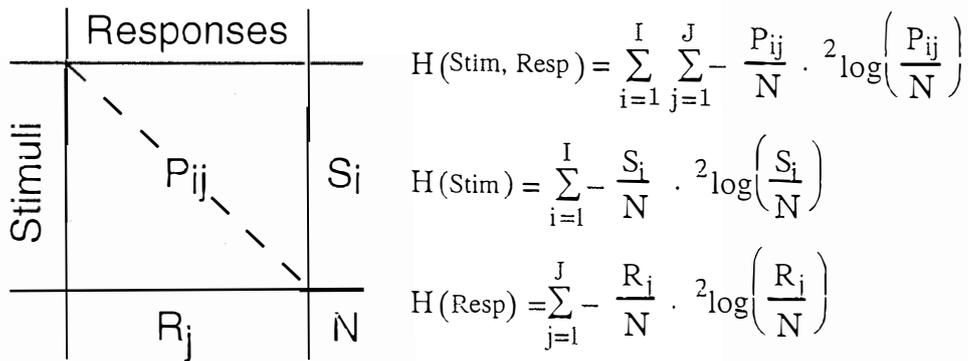


Figure 1. The entropy measures of a confusion matrix. $H(\text{Stim, Resp})$: the entropy of the stimulus response pairs. $H(\text{Stim})$: the entropy of the stimuli. $H(\text{Resp})$: the entropy of the responses. S_i : the number of occurrences of stimulus i . R_j : the number of occurrences of response j . P_{ij} : The number of occurrences of response j to stimulus i . N the total number of stimuli and responses.

are concentrated in a subset of the stimuli and responses, a change in the error rate will induce a change in the distribution of the correct responses. This latter change again will affect the error-dispersion. This residual dependency can be “normalized” out of the error dispersion by decomposing it into an error-dependent part and a part that depends on the distribution of the correct responses. However, this approach falls outside the scope of this paper.

4. The error difference, δ

When confusion matrices are combined, i.e., experimental data are pooled, the entropies ($H(\text{Stim, Resp})$, $H(\text{Stim})$, and $H(\text{Resp})$) of the resulting pooled matrix are larger than the corresponding mean entropies of the contributing matrices. The difference (i.e., $H_{\text{pooled}} - H_{\text{mean}}$) is a measure for the differences in the distribution of stimuli, responses, and stimulus-response pairs. The same holds for the derived measures of the perplexity and the error dispersion. This difference can be used to define a measure that quantifies the difference in the distribution of the errors between confusion matrices. The error-difference, δ , is such a measure. It is defined in equation 4. Its general definition takes into account the weighting of the individual matrices, i.e., not all matrices contribute an equal number of errors to the pooled matrix (see Van Son, 1994, 1995 for a derivation).

$$\delta_s = \frac{h_s^{\text{pooled}} - \bar{h}_s}{H_\alpha}; \quad \delta_r = \frac{h_r^{\text{pooled}} - \bar{h}_r}{H_\alpha} \quad (4)$$

In which:

δ is the fraction of the total (pooled) number of errors, ϵ^{pooled} , that is not “shared” between the matrices (i.e., errors that are “unique” to individual matrices)

h^{pooled} is calculated over the pooled confusion matrix

$\bar{h} = \sum_i \frac{\alpha_i \cdot \epsilon_i}{\epsilon^{\text{pooled}}} \cdot h_i$ is the mean of the individual matrices (weighted by their contribution to ϵ^{pooled})

H_α is the entropy of the weighting factors α_i of the matrices ($-\sum \alpha_i \log_2[\alpha_i]$)

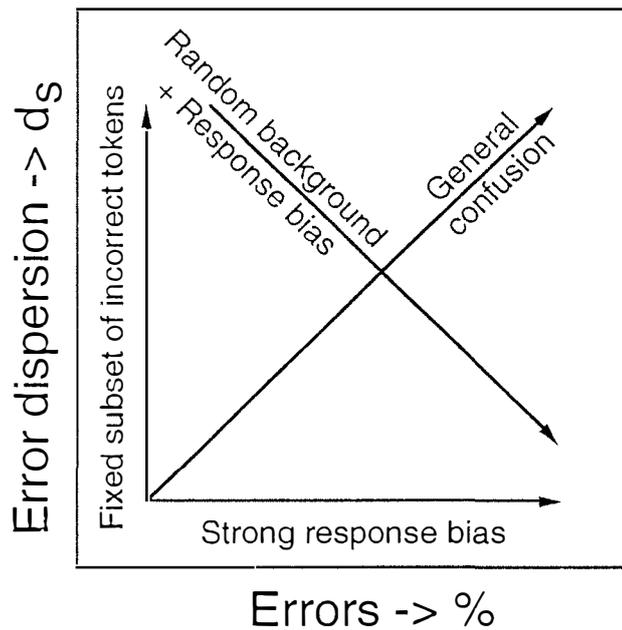


Figure 2. Prototypical relations between the Error dispersion and the Error rate (see text)

In words, the error difference, δ , is the fraction of the total (pooled) number of errors, ϵ^{pooled} , that is not “shared” between the matrices (i.e., errors that are “unique” to individual matrices).

The error difference is based on the logarithm of the error dispersion (i.e., on h) which is normalized with respect to the size of the confusion matrix, i.e., the error dispersion is the number of error categories *per* stimulus or response category. As a result of how h^{pooled} and \bar{h} are calculated, differences between individual matrices due to non-overlapping stimulus or response sets are discounted from the error difference itself. Furthermore, the error difference, δ , is normalized with respect to the combined stimulus or response sets. This means that the error difference is reduced when non-overlapping confusion matrices are used compared to when only the overlapping parts of these confusion matrices are used.

5. The relation between the Error Rate and the Error Dispersion

Although the error dispersion is mathematically (almost) independent of the error rate, in an actual experiment you will be looking for relations between these measures. The question is then to determine how the error dispersion changes with increasing (decreasing) error rates. The prototypical relations are given in figure 2.

In figure 2 several possible correlations between the error rate and the error dispersion are drawn as straight lines. One possibility is that both are not correlated because one of them is essentially fixed. In general, if the errors are dominated by a response bias (e.g., long/short vowel confusions) that varies in strength, the distribution of the errors does not change when the error rate itself changes. As a result, the error dispersion will be nearly fixed with error rate (the horizontal line in figure 2). At the other extreme, all errors can be concentrated in only a subset of the tokens which are always identified incorrectly (e.g., short realizations of intrinsically long vowels versus short vowels). As the specific *incorrect* responses vary with the experimental conditions, but the *number* of incorrect responses does not, the result will be a nearly fixed error rate with varying error dispersion (the vertical line in

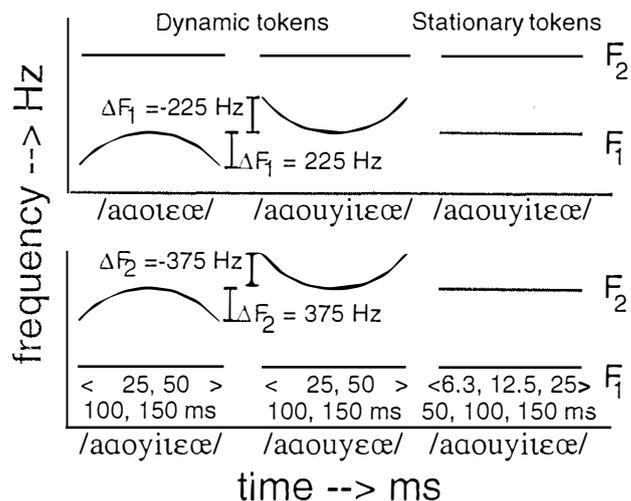


Figure 3. Token construction for the experiment of Pols and Van Son (1993). Tokens were synthesized with formant tracks as indicated in the two plots. Nine different F_1/F_2 midpoint values were used, corresponding to the Dutch phonemes shown below the plots. Each parabolic formant track shape was synthesized with 4 durations. The tokens with stationary formant tracks were synthesized with 6 durations. Because of limitations of formant synthesis, not all target points could be synthesized with all formant track shapes.

figure 2). In this case, it is an open question who made the incorrect identifications, the experimentator or the subjects.

An experiment in which either the error dispersion or the error rate is fixed will be quite exceptional. A more likely result is a positive correlation between the error dispersion and the error rate. Such a result indicates that new errors are distributed over new responses. That is, as a stimulus elicits more errors, these errors are distributed over more responses. In short, a growing error rate indicates a growing confusion. The actual slope of the linear regression line is a measure of the number of error categories that are added per "new" error.

A very interesting type of (cor-) relation is when the error dispersion *reduces* with increasing error rate, i.e., a negative correlation. This can happen when there is a fixed "background" of errors with a high error dispersion superimposed on a varying response bias with a low error dispersion. The error dispersion is then some kind of combination (not necessary linear) of the individual contributions. In this situation, the (combined) error dispersion decreases in size when the relative strength of the bias, and hence the error rate, increases. An example could be an experiment in which long/short vowel confusions are induced by varying token durations and indiscriminate errors by a fixed level of added noise (a fixed, low, Signal-to-Noise Ratio). In such a case, the error dispersion will reduce when the response bias becomes stronger, and as a consequence, the error rate becomes higher.

6. Examples from the literature*

The above theory is tested on a number of papers taken from the literature. When possible, the correlation coefficients are tested for statistical significance. However, statistically significant correlations are not used to infer a causal relation with

* These and other examples can be examined in detail on the WWW. URL: "<http://fonsg3.let.uva.nl/Services.html>". It is also possible to test new examples on-line

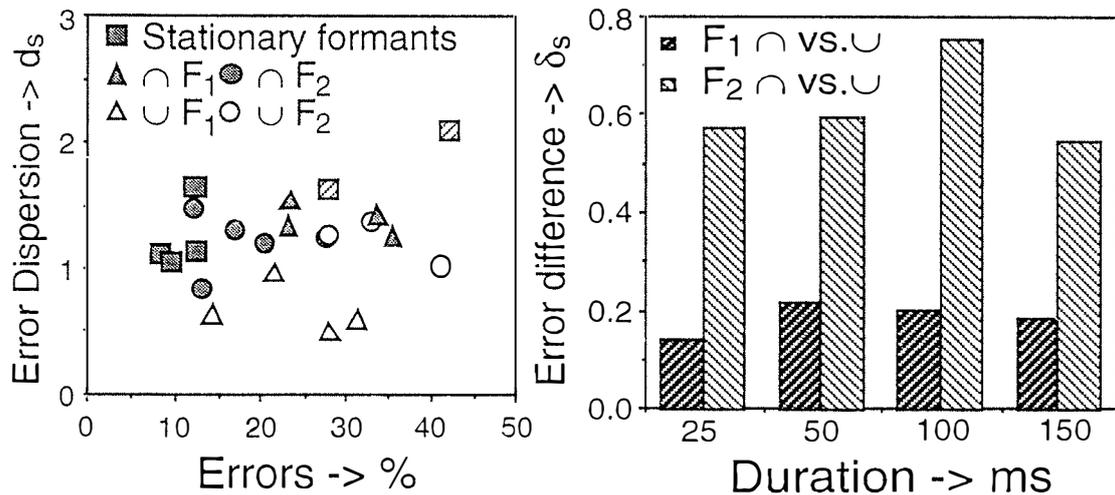


Figure 4. Results of the identification experiment of Pols and Van Son (1993). See figure 3 for token construction. Left panel: Error dispersion, d_s , versus error rate for each duration (ignoring long/short confusions). The light hatched squares are the results for the "stationary" tokens with durations of 12.5 and 6.3 ms. Right panel: Error difference, δ_s , between tokens with "opposite" formant track curvatures.

experimental conditions. Such an interpretation would require a detailed analysis of the underlying experiments which is outside the scope of this paper.

6.1. The influence of formant track shape on vowel identification

In an experiment to determine how vowel identity is influenced by the shape of formant tracks, Pols and Van Son (1993) performed a listening experiment with synthetic stimuli. For each of nine different F_1/F_2 "target" points distributed over the vowel triangle, five formant track shapes were constructed: one stationary and four with parabolic F_1 or F_2 tracks. All five track types had the same formant values in the center. These track shapes were synthesized with 4 or 6 different durations (figure 3).

Subjects were asked to identify the sounds as Dutch vowels. The results indicated a shift in the responses corresponding to an averaged formant value for the synthetic vowels rather than some form of "perceptual overshoot". The associated error dispersions, error rates, and error differences are calculated from the raw data (figure 4, Pols and Van Son, 1993, do not give error rates).

Figure 4 shows that there was a large spread in the error rates for the different conditions, from below 10% to over 40%. The individual conditions with respect to the formant track shape are nicely separated in the plot of error dispersion versus error rate. Up- and downward parabolic F_1 tracks have distinct error dispersions. Up- and downward parabolic F_2 tracks have distinct error rates. The short duration (12.5 and 6.3 ms) stationary tokens have high error rates and high error dispersion, which separate them from the non-stationary tokens which have comparable error rates but lower error dispersions. The fact that the error dispersions of the non-stationary tokens are all around 1 indicates that the errors are concentrated in a single category for each stimulus.

The separation between tokens with upward and downward pointing formant tracks is even more evident when the error difference is used (figure 4). Although the error dispersions of tokens with up- and downward pointing F_2 tracks are equal, the error difference is over 0.5, indicating that more than half the errors are different (figure 4). For the F_1 shapes, large differences are found too. However, here the error difference

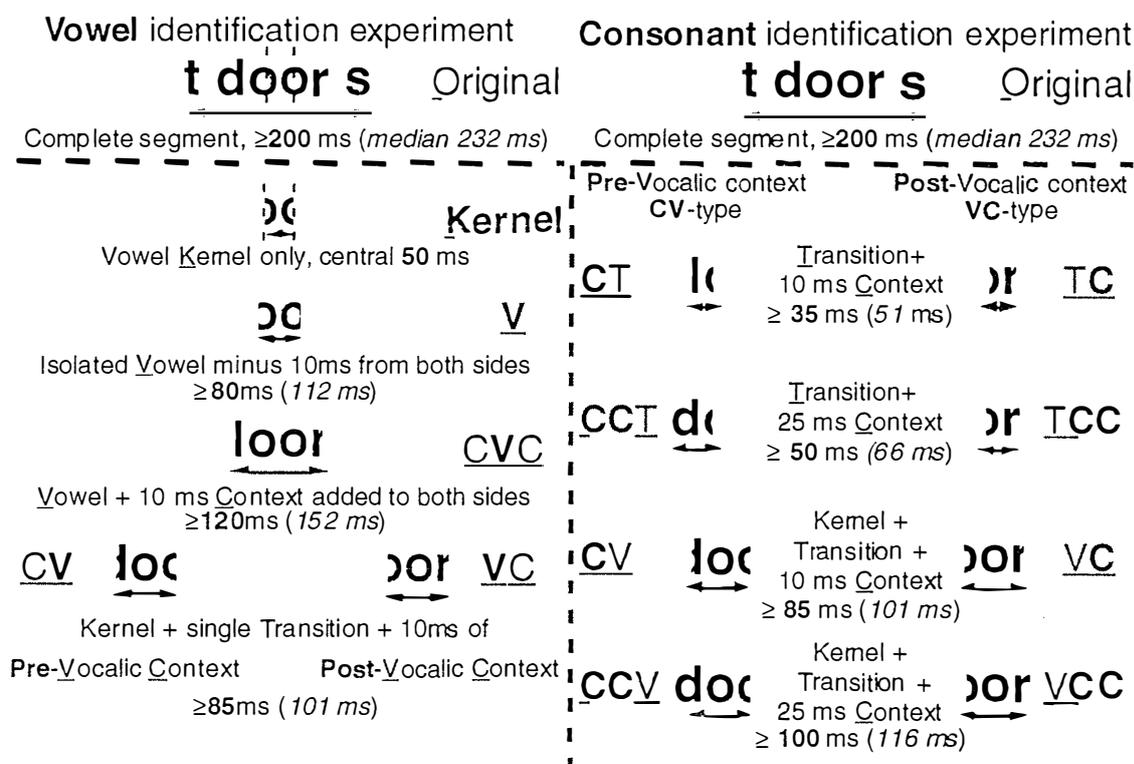


Figure 5. Token construction for the identification experiment of Van Son and Pols (1995). Starting from CVC fragments excised from read text, tokens were constructed by progressively removing more and more of the context and the transition region of the target phoneme, either the vowel or the pre- or post-vocal consonant. Either the vowels, or the pre- and post-vocal consonants were identified in different experiments.

is reduced due to the large difference between the sets of stimuli used (6 versus 9 formant “targets”, see figure 3). In total, the large error differences indicate that the errors are concentrated in different categories for different formant track shapes.

Using a different type of analysis, Pols and Van Son (1993) concluded that the parabolic formant tracks induced an averaging in the perception of the formant values. The result is that the responses are shifted towards vowels with a lower formant target value for tokens with downward pointing (\cap) formant tracks and towards vowels with a higher formant target value for upward pointing (\cup) formant tracks. The errors in both cases would exclude each other. This behaviour would result in a large error difference between tokens with upward pointing and tokens with downward pointing formant tracks, as is indeed manifest in figure 4.

6.2 The influence of context on identification

Using natural (read) speech, Van Son and Pols (1995a,b) investigated how the presence or absence of nearest neighbour context affected the identification of vowels and consonants from read CVC tokens from a long, meaningful text.

Although the error rates for CV- and VC-type tokens are comparable in size, figure 6 shows that the underlying cause of the errors could be very different. The VC-type tokens show a strong correlation between error dispersion and error rate, akin to the “general confusion” line in figure 2. The CV-type of tokens combine a virtually constant error dispersion with a widely varying error rate. Vowel identification (the CVC-type tokens) behaves more like consonant identification in CV-tokens than in

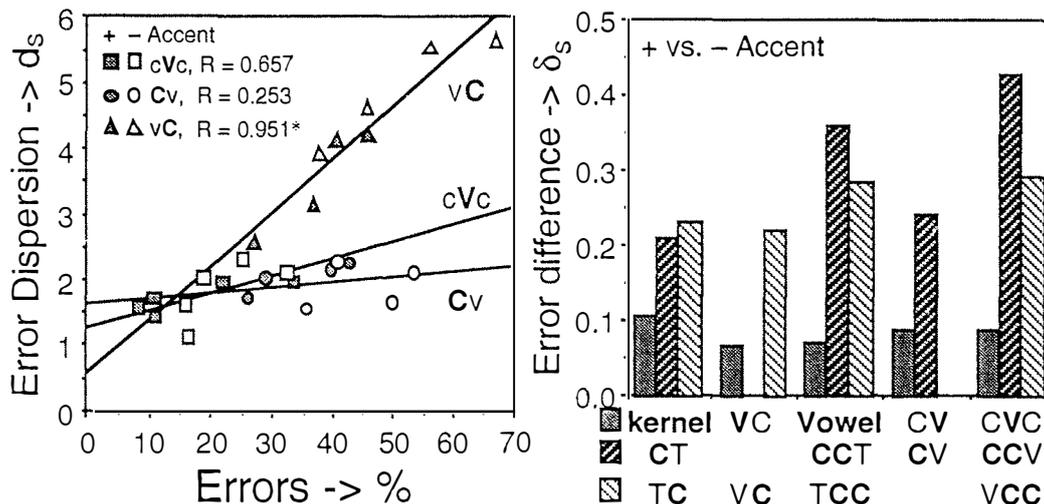


Figure 6. Results of the vowel and consonant identification experiments of Van Son and Pols (1995a,b). For a specification of the stimuli, see figure 5. Long/Short vowel, or consonant voicing, errors were ignored. Left panel: Error dispersion versus error rate, d_s , * $p \leq 0.01$, two tailed. Right panel: Error difference, δ_s , between tokens with accented and non-accented vowels.

VC-tokens. This suggests that the errors in the identification of pre-vocalic consonants and vowels (CV and CVC in figure 6) are concentrated in a few (approximately two) "response biases", whereas the errors in the identification of post-vocalic consonants (VC in figure 6) spread out over a large number of response categories.

Van Son and Pols (1995) found a consistent difference in error rate between tokens containing accented and tokens containing unaccented vowels. This difference can be seen, more or less, in figure 6, where the error rates for the unaccented tokens (open symbols) are generally higher than the error rates for the accented tokens (filled symbols). Van Son and Pols (1995) showed that this difference in error rate was consistent and statistically significant. The error differences between the accented and unaccented tokens show a difference between vowel identification and consonant identification. For the identification of the post-vocalic consonants (VC) this is to be expected, the error dispersion increases with the error rate, so if the unaccented tokens induce more errors, it is to be expected that these "extra" errors are different. For the identification of the pre-vocalic consonants (CV), the high error difference indicates that the fixed error dispersion for accented and unaccented tokens are the result of different sets of incorrect responses.

Combined, the results of this analysis indicates that there are differences between the identification of pre-vocalic and post-vocalic consonants. It seems that the patterns of identification errors of pre-vocalic consonants are more like those of vowels with respect to the number of error categories, and more like those of post-vocalic consonants with respect to the differences between accented and unaccented tokens. In their paper, Van Son and Pols (1995) do indeed conclude that there is a (cor-)relation between the identification errors in pre-vocalic consonants and vowels, but not between post-vocalic consonants and vowels. They also find that the differences between accented and unaccented tokens are different for vowel identification and consonant identification.

Several studies compared identification results for vowels presented with and without their original, syllabic, context. Both Huang (1991) and Kuwabara (1985) published confusion matrices that are relevant to this question. These are used here.

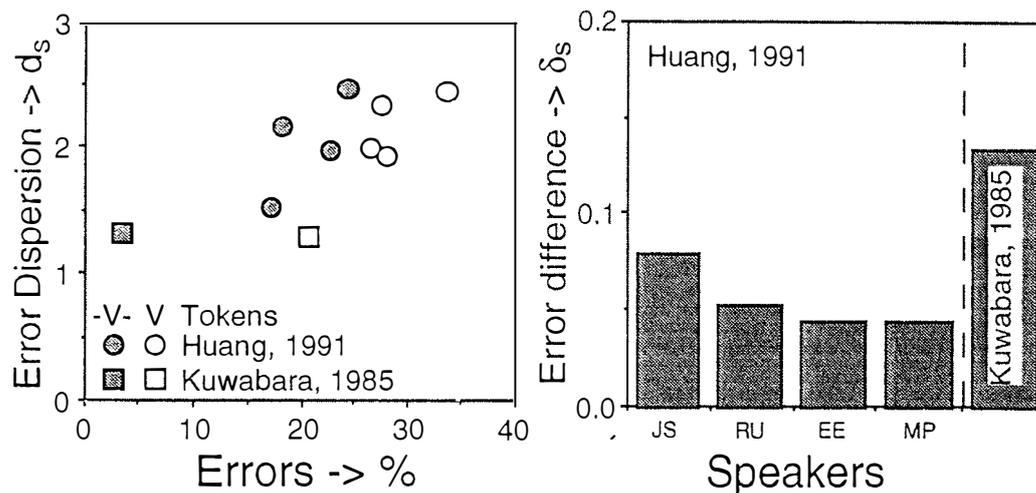


Figure 7. Vowel identification results for presentation with and without the original, syllabic, context. The results are taken from Huang (1991) and Kuwabara (1985). Left panel: Error dispersion versus error rate. Huang (1991): CVC (●) versus the excised (isolated) Vowel (○) for 4 speakers. Kuwabara (1985): Trisyllabic VVV sequences (■) versus the excised, medial Vowel in isolation (□), 4 speakers pooled. Right panel: Error difference between presentation with and without the original syllabic context.

From both studies it can be concluded that vowels presented in context (filled symbols) are identified better than those presented in isolation (open symbols). The range of error rates is large and clearly separates the two conditions in each experiment. At the same time, the error dispersions for the two conditions overlap. It seems that the increase in the error rate caused by the removal of the context only strengthens existing errors (or response biases), and does not introduce new errors. This idea is supported by the error differences which are quite small (figure 7, right-hand panel, compare with figures 6 and 4 and be aware of the differences in vertical scale). Inspection of the confusion matrices of Huang and Kuwabara confirms this picture. There is a strong overlap of identification errors between presentation of vowels in context and in isolation.

6.3 Vowel reduction

Vowel reduction is an obvious field of research for testing the use of this new analysis technique. In figure 8, some results of the study of Van Bergem (1993) are presented. Van Bergem (1993) studied the influence of stress on vowel reduction in Dutch words. Identical syllables were recorded when pronounced in isolation, as the stressed and unstressed syllables of content words, and pronounced as function words. An English example would be: *can, candy, canteen, can*. The words were part of carrier sentences and were pronounced both with and without sentence accent (except for the syllables pronounced in isolation). For the actual identification experiment, the excised vowel realizations were presented in isolation. These experimental conditions were chosen to induce varying levels of vowel reduction.

From figure 8, it can be seen that for all three speakers there is a strong correlation between the error dispersion and the error rate. Clearly, the different conditions in which the vowels were pronounced induced a varying amount of vowel reduction. This reduction translated into a large variation in error rates (~15-75%) and error dispersion (~1-4). The strong correlation between error dispersion and error rate suggests that many distinctive features of the vowel realizations are progressively

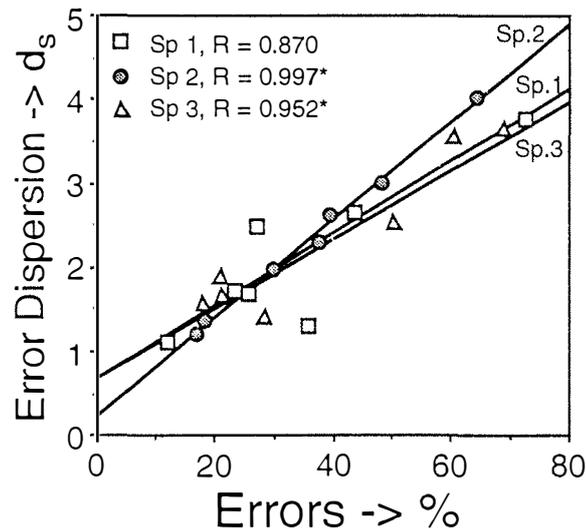


Figure 8. Results of Van Bergem, 1993. Vowels were pronounced under different conditions that were designed to vary the amount of vowel reduction. The results for all three speakers are plotted: Sp 1, Sp 2, and Sp 3. R: correlation coefficient, * $p \leq 0.01$, two tailed

affected (“reduced”) by vowel reduction, but in a consistent way. It also suggest that the error rate is the only relevant measure with respect to vowel reduction.

6.4 Mixed speaker vowel identification

As a final example, results are analysed from a series of experiments, aimed at determining whether listeners use context to normalize for speaker variation (Verbrugge et al., 1976; Strange et al., 1976). Tokens from several different speakers were presented in mixed order. In the two experimental “conditions”, the type of context varied: none, full CVC, and only /pVp/, as well as the type of precursor sentence: maximal vowel space (/hi/ /ha/ /hu/), minimal vowel space (/hʌ/ /hæ/ /hʌ/), and “natural” precursor sentences (figure 9).

Despite the small number of points available, some interesting differences between the conditions can be seen. The vowels and syllables presented in isolation (open circles, figure 9) give the familiar picture of increasing error dispersion with increasing error rates (the correlation is not statistically significant) that signifies “general confusion” (figure 2). However, the syllables presented in a sentence context show a completely different picture. One point, the /pVp/ syllable presented in a natural sentence context combines a markedly increased error rate with a lower error dispersion. From the discussion of figure 2, it is known that this is the hallmark of a strong response bias superimposed on a background of “random” errors. Verbrugge et al. (1976) do reach the same conclusion from a close inspection of the response data.

7. Discussion and conclusions

In the previous sections the results of several studies were re-analysed using the relation between the error dispersion and the error rate. These studies were not designed to be analysed this way so they serve mainly as an illustration of the potential use of this technique. I do not intent a re-interpretation of their results.

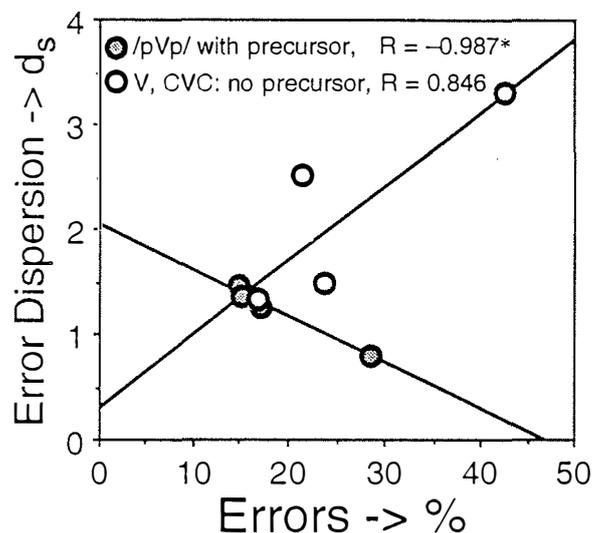


Figure 9. Speaker normalization results from Verbrugge et al. (1976), Strange et al. (1976). The error dispersion is plotted against the error rate. R: correlation coefficient, *: statistical significance, $p \leq 0.025$, two-tailed. ● Stressed and destressed pVp syllables presented with /hi/ /ha/ /hu/, /h/ /hæ/ /hʌ/, and “natural” precursor sentences (six points but they overlap) ○ Vowels, CVC syllables and pVp syllables (+/- stressed) presented in isolation (four points)

However, it is clear that the plots of error dispersion versus error rate can help a lot in determining whether experimental conditions do make a difference for identification (figure 4). Furthermore, the strength and direction of the correlation can indeed suggest what kind of process underlies the changes in the responses (figures 6, 7, 8, and 9). It was also demonstrated that conditions that elicit comparable error rates and error dispersion can sometimes be distinguished because of the large differences in the actual incorrect responses observed (figures 4 and 6). The reverse was also found, where large differences in error rates were based on the same type of errors (figure 7).

Therefore, it can be concluded that the error dispersion, d , is a meaningful measure of the number of distinct errors per stimulus or response (cf., perplexity). The error dispersion can separate cases for which the error rate alone is insufficient. Combined with the error rate, it also can suggest the kind of processes that underlie the differences in responses.

Furthermore, the error difference, δ , quantifies the (dis-)similarity of stimuli and conditions. It is useful to assess whether there are differences in the pattern of responses, independent of any differences in error rate or error dispersion.

8. Acknowledgements

I would like to thank dr. Dick van Bergem for making available the data of his identification experiments. This research was made possible by grant 300-173-029 of the Dutch Organization of Research (NWO).

9. References

Allen, J. (1994). “How do humans process and recognize speech?”, *IEEE Transactions on Speech and Audio Processing*, 2, 567-577.

- Bahl, L.R. & Jelinek, F.J. (1990). "A maximum likelihood approach to continuous speech recognition", in A. Waibel & K.F. Lee (editors) *Readings in speech recognition*, Morgan Kaufmann publishers, San Mateo, CA, USA, 308-319.
- Huang, C.B. (1991). "An acoustic and perceptual study of vowel formant trajectories in American English", Ph.D. Thesis, Massachusetts Institute of Technology, USA (Research Laboratories of Electronics, Technical report no. 563, Cambridge, MA), 203 pp..
- Khinchin, A.I. (1957). *Mathematical foundation of information theory*, translated by R.A. Silverman and M.D. Friedman, Dover Publications Inc., New York NY, 120 pp.
- Kuwabara, H. (1985). "An approach to normalization of coarticulation effects for vowels in connected speech", *Journal of the Acoustical Society of America* **77**, 686-694.
- Miller, G.A. & Nicely, P.E. (1955). "An analysis of perceptual confusions among some English consonants", *Journal of the Acoustical Society of America* **27**, 338-352.
- Pols, L.C.W. & Van Son, R.J.J.H. (1993). "Acoustics and perception of dynamic vowel segments". *Speech Communication* **13**, 135-147.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P. (1988). *Numerical recipes in C*, Cambridge University Press, Cambridge MA, second edition 1992, 632-635.
- Strange, W, Verbrugge, R.R., Shankweiler, D.P., & Edman, T.R. (1976). "Consonant environment specifies vowel identity", *Journal of the Acoustical Society of America* **60**, 213-224.
- Sveshnikov, A.A. (ed.) (1968). *Problems in probability theory, mathematical statistics and theory of random functions*, W.B. Saunders Company Philadelphia, 157-170.
- Van Bergem, D.R. (1993). "Acoustic vowel reduction as a function of sentence accent, word stress, and word class", *Speech Communication* **12**, 1-23.
- Van Son, R.J.J.H. (1994). "A method to quantify the error distribution in confusion matrices", *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* **18**, 41-63.
- Van Son, R.J.J.H. (1995). "A method to quantify the error distribution in confusion matrices", *Proceedings Eurospeech 95*, Madrid, Spain, 2277-2280.
- Van Son, R.J.J.H. (1995). "A method to quantify the error distribution in confusion matrices", *Proceedings of Eurospeech 95*, Madrid, Spain, 2277-2280.
- Van Son, R.J.J.H. & Pols, L.C.W. (1995a). "The influence of local context on the identification of vowels and consonants", *Proceedings Eurospeech 95*, Madrid, Spain, 967-970.
- Van Son, R.J.J.H. & Pols, L.C.W. (1995b). "How transitions and local context affect segment identification", *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* **19**, 51-69.
- Verbrugge, R.R., Strange, W & Shankweiler, D.P. & Edman, T.R. (1976). "What information enables a listener to map a talker's vowel space?", *Journal of the Acoustical Society of America* **60**, 198-212.