

# ACOUSTIC CORRELATES OF REDUNDANCY AND INTELLIGIBILITY

SHERI HUNNICUTT

DEPARTMENT OF SPEECH COMMUNICATION AND MUSIC ACOUSTICS  
ROYAL INSTITUTE OF TECHNOLOGY, BOX 70014 S-10044 STOCKHOLM

## ABSTRACT

The relationship between context redundancy and key-word intelligibility was earlier examined in both high- and low-redundancy contexts. Word pairs were placed in similar positions in two sets of sentences: text-type sentence pairs and adages together with spoken-type sentences. With the text-type sentence pairs, there was an intelligibility advantage for the words in lower-redundancy contexts. For the adage and spoken-sentence pairs, there was no intelligibility advantage for words in either context.

An acoustic study has now been undertaken to determine whether differences can be measured in the production of the word pairs. The peak dB level of word pairs, their durations and the duration of various sub-word components, F0 maximum, range, excursion and contour complexity were considered as possible correlates. It was found that the correlation between any one of these factors and word intelligibility was quite low. However, it was also observed that in a pairwise comparison, differences in these factors occurred in the expected direction in a majority of cases.

## INTRODUCTION

The purpose of this study has been to determine to what extent context redundancy and key-word intelligibility are related, and to examine possible acoustic correlates of these factors. It was intended that the results should help to answer some general questions about information control by talkers. For example, Lindblom /1/ has hypothesized that talkers naturally control the amount of information they wish to give to a listener. An ideal talker might attempt to increase the intelligibility of words that he expects his listener to have fewer "higher level" cues about. How, then, is this higher intelligibility brought about? There are several possible avenues: grammatical inversions, phonetic precision, prosodic stress. And, further, do these adjustments actually improve intelligibility for the listener?

A study which has been cited for many years as support for an inverse relationship between context redundancy and key-word intelligibility is one by Philip Lieberman /2/. In his study, Lieberman investigated the intelligibility of word pairs and examined the correlation of percent identification in these pairs with word duration, VU meter reading and relative peak amplitude. Since Lieberman's

study was a small one, analyzing only seven words from eleven sentences, and since the redundancies in some of the word pairs were not well opposed, it was decided to replicate this work, considering more word pairs with more extreme redundancies.

## INTELLIGIBILITY-REDUNDANCY STUDY

### Test Materials

The initial study, in which subjects attempted to identify words extracted from sentence contexts, was described earlier (Hunnicut, /3/). The words were taken from similar high- and low-redundancy sentence contexts containing matched pairs of words. One set of high-redundancy sentences were adages. The corresponding set of low-redundancy contexts were from grammatically similar sentences that might be spoken. The two other corresponding sets of sentences were rather long, grammatically standardized sentences which one might find in a text. All sentences were read by one speaker.

There were 76 text-type sentences recorded. These were taken from a set of Swedish test sentences developed for speech perception tests by Rolf Lindgren /4/. The test words in these sentences are all common words beginning with a stop and are in sentence-object position, far enough along in the sentence for a specific context to be built up. Redundancies for the sentences were established by having subjects fill in the blanks left by removing the test words.

The 36 Swedish adages were collected with the criterion that they contain a non-initial noun, preferably also non-final and of two syllables. The companion spoken-type sentences were constructed to have similar grammatical and syllable structure to those of the adages, with the test word in the same position. These sentences were constructed to give as little information as possible about the key-word, i.e., they were constructed as low-redundancy contexts.

Half of each type of the sentences were used as actual test material and half as fillers.

### Speech Processing

The sentences were copied from tape onto a large computer disk where the test words were edited out using an interactive speech processing program /5/. A 2.1-second period of speech-like noise was added to cover each word, and the noise level chosen so that the average signal-to-noise ratio over the set of all words was 4 dB. The noise was increased from zero at the onset and attenuated at the offset over a period of 100

milliseconds to avoid an abrupt, and possibly distracting or confusing presentation. These words were then recorded on tape again with 4 seconds between words and 10 seconds after each set of 15 words. An open intelligibility test was then administered to 10 subjects.

#### Results

For the 19 text-type sentence pairs, there was a clear intelligibility advantage for the words in lower-redundancy contexts. There were 10 sentence pairs in which the low-redundancy words were more intelligible, 7 in which the low- and high-redundancy words were of the same intelligibility, and only 2 sentences in which the high-redundancy words were more intelligible. The mean number of correct answers for words in the low-redundancy contexts was found to be significantly greater using a paired-comparison test /6/.

There was no such advantage for words in lower-redundancy contexts for the 21 adage/spoken-sentence pairs, however. The number of sentence pairs in which the low-redundancy words were more intelligible was only 5, the number of sentences where the word pairs were of equal intelligibility was 9, and in 7 sentence pairs, the high-redundancy words were more intelligible. The mean number of correct answers were quite close, and the difference not statistically significant.

#### ACOUSTIC CORRELATE STUDY

An acoustic study has recently been undertaken to determine whether differences can be measured in the production of the word pairs, and whether observed differences could account for the previous results in word intelligibility. It was hypothesized that the more intelligible word of the pair should have some combination of the following attributes: higher intensity, longer duration, larger FO excursion or more complex FO pattern. A "perfect example" is shown for the word bollen below. This example, however, is not a norm, as will be seen in the text. ("Stress syll dur" is the heading for stressed syllable duration.)

BOLLEN		No.	dB	Stress level	Word dur (sec)	FO excursion (Hz)	FO complexity
Redundancy	Correct						
1.00	7	10.6	.31	.41	50	3	
.24	10	11.6	.34	.52	117	3	

#### Intensity

A comparison of peak dB levels for the word pairs in the text-type sentences revealed a tendency towards a lower dB level for words in the high-redundancy contexts, as expected. Of the 19 word pairs, 13 words in high-redundancy contexts had a lower dB value. One additional pair had equal dB-levels. A comparison of the adage/spoken-sentence pairs shows a lesser tendency in the same direction. Of the 21 word pairs, 12 words in high-redundancy contexts (57%) had a lower dB value. An additional three pairs had equal values.

A t-Test for related measures shows a significant difference in intensity of key-words in

high- and low-redundancy contexts in both types of sentence pairs. Correlation between words correctly identified and dB-level is low, however, being .07 for the words in more redundant contexts and .18 for the words in less redundant contexts.

#### Duration

Duration measurements of the word pairs were made on the whole word, the stressed syllable (and stressed open syllables plus following consonants), the stressed vowel and the consonant(s) preceding the stressed vowel. It was found that the duration of the stressed vowel and the duration of the preceding consonant were often complementary. The two measurements were therefore combined into a prestress consonant plus stressed vowel duration. This measurement is the same as that for the stressed syllable in open syllables, of course. There were three cases in which one word of a word pair was longer in its stressed open syllable and the other word of the pair longer when the duration of the following consonant(s) were added, so these two measurements were both retained. This division effectively gives two sub-word measurements: one for the initial consonant(s) and vowel of the stressed syllable, and the second for these plus the following consonants, regardless of syllable division.

For the text-type sentences, there was a slight tendency towards shorter durations for words in the high-redundancy contexts, as expected. Of the 19 word pairs, 12 words in high-redundancy contexts had shorter initial consonant plus vowel, and 2 were of equal length. Eleven words in high-redundancy contexts had shorter initial consonant plus vowel plus following consonant(s). Only 10 entire words were shorter, with 2 of equal length. All t-Tests of durational differences show small differences of rather low significance.

For the adage/spoken-sentence pairs, the results were the opposite: there was a slight tendency towards longer durations for words in the high-redundancy contexts, contrary to expectations. Of the 21 word pairs, only 8 words in high-redundancy contexts had shorter initial consonant plus vowel, and 1 was of equal length; 12 words had longer initial consonant plus vowel. Seven words in high-redundancy contexts had shorter initial consonant plus vowel plus following consonant(s), 3 were of equal length and 11 were longer. Only 9 entire words were shorter, 2 were of equal length, and 10 longer. Except for a nonsignificant difference in word duration, all t-Tests of duration differences show small differences of rather low significance. Correlations between the number of words correctly identified and each of the duration measures is fairly low for both types of sentence pairs. The highest correlations were with word duration in the adage/spoken-sentence pairs (.40) and with stressed syllable duration in the text-type sentence pairs (.33).

#### FO movement

Four measures of FO movement have been employed: maximum FO value, FO range and total FO excursion during a word, and FO contour complexity. Total excursion is taken to be the sum

of rises and falls (of at least 6 Hertz each), and contour complexity is expressed as the number of such rises and falls. Whereas intensity and duration measurements showed greater differences in the expected direction in the text-type sentences, FO measurements show greater expected differences in the adage/spoken-sentence pairs. All four FO measures show decided greater values for spoken-type low-redundancy context sentences compared to adages, as expected. Contrary to expectation, however, measurements were often smaller for words in low-redundancy context text-type sentences compared to high-redundancy contexts.

For text-type sentences, maximum FO value is larger for words in low-redundancy contexts in 11 of the 19 cases, and total FO excursion is larger in 10 of 19 cases. However, FO range is larger for words in low-redundancy contexts in only 7 cases, and the number of rises and falls is larger in only 3 cases, 10 words having an equal number.

In the adage/spoken-type sentence pairs, maximum FO value is larger for words in the low-redundancy spoken-type sentences in 13 of 21 cases, FO range is greater in 17 cases, total excursion greater in 18 cases and number of rises and falls greater in 8 cases, 9 words having an equal number.

#### GENERAL COMPARISON

In Table 1, shown below, an attempt has been made to condense the preceding information and to give a comparative overview. A check mark in the table indicates that the data supports the expected result of more correct identifications of a word or more marked prosody for a word in a low-redundancy context; an "x" indicates the opposite. An equals sign (=) indicates equivalent data and a percent sign (%), ambivalent data. A check mark appears in the first column if the word in the less redundant context was identified better (as expected), in the second column if it had a higher dB level, in the third column if it had a longer duration, and in the fourth column if it exhibited greater FO movement (indicating that at least 3 of the 4 FO measurements are greater).

Summing the check marks in the table for text-type sentences, we see that the word in the low-redundancy context was identified better in 53% of the cases. We also see that 53% of the possible prosody cues are present to support identification (i.e., of the 57 possible check marks, 30 are present). Summing x's, we also see that there were 26% contradictory cues, i.e., lower dB-level, shorter duration or less FO movement in the low-redundancy context. Only 11% of high-redundancy context words were better identified.

The number of prosody cues for the adage/spoken-sentence pairs is quite similar to the text-type sentences, even though identification was not, as previously reported. Words in low-redundancy contexts were better identified in only 24% of the cases, and words in high-redundancy contexts, 38%. However, we see that 54% of the possible prosody cues are present to support better identification of words in low-redundancy contexts. Contradictory cues were present in 24% of the cases.

	<Redund. >Correct	<Redund. >dB	<Redund. >Dur	<Redund. >FO Mvmt
TEXT-TYPE SENTENCES				
BOKEN	✓	✓	✓	✓
BOLLEN	✓	✓	✓	✓
BORDEN	✓	✓	✓	✓
DÖRREN	X	✓	✓	%
GOLVET	=	✓	✓	X
KAMMEN	✓	=	%	✓
KARTAN	✓	X	X	X
KLOCKAN	X	X	X	X
KNIVEN	✓	✓	%	✓
PENNAN	=	✓	%	%
TAKEN	✓	X	%	%
TAVLAN	=	X	✓	X
BREVEN	✓	✓	✓	✓
DUKEN	✓	X	%	X
HANDDUKEN	=	✓	X	%
SKÅPET	=	✓	✓	✓
TRAPPAN	✓	✓	%	X
ÄDELSTENEN	=	✓	✓	X
BILEN	=	✓	✓	%
ADAGE/SPOKEN-TYPE SENTENCES				
FÄGEL	✓	✓	✓	✓
HANDEN	✓	✓	X	✓
SKOGEN	X	✓	%	X
BÄCKEN	=	X	%	✓
SOMMAR	=	✓	%	✓
KORVEN	=	✓	✓	X
BÖRJAN-1	X	✓	✓	✓
BÖRJAN-2	=	X	%	✓
KVARNEN	X	X	X	✓
TJUREN	X	✓	✓	✓
GRYTOR	✓	✓	✓	✓
KVASTAR	=	✓	X	✓
HUNDEN	✓	X	X	%
VAGNEN	X	=	X	✓
RÄVEN	X	X	✓	%
KATTEN	X	✓	X	X
BORDET	=	✓	%	✓
SKINNET	✓	✓	%	✓
BJÖRNEN	=	=	%	✓
DÄRAR	=	=	%	✓
VATTEN	X	X	X	✓

TABLE 1. A GENERAL PROSODIC COMPARISON

Combining prosody cues for all 120 words, it can be seen that expected cues appear in 53% of the cases, contrary cues in about half so many, 26%. One observation supportive to the hypothesis is that at least one contrary cue appears in 8 of the 10 words in which identification was also contrary to expectation. An examination of the recordings and spectrograms of the two remaining word pairs reveals nothing of a segmental quality which could have caused these identification results. It should also be noted that there are 3 word pairs in which the word in the low-redundancy context was better identified even though 2 or 3 contrary prosodic cues were present. No particular segmental effects have been noted for these words either.

### CONCLUSIONS

An earlier study investigated the relationship of a word's intelligibility to the redundancy of its context. This study found an intelligibility advantage for words in lower-redundancy contexts in text-type sentences. For adage and spoken-type sentence pairs, however, there was no such advantage for words in either the low-redundancy spoken-type sentences or the high-redundancy adages. It was conjectured that the metaphorical nature of adages and the influence of the social conditions in which they are used may have prevented the realization of the intelligibility-redundancy relationship observed in the text-type sentences.

The current study examined possible prosodic correlates of redundancy and intelligibility. The peak dB level of word pairs, their durations and the duration of various sub-word components, F0 maximum, range, excursion and contour complexity were considered as likely correlates. It was found that the correlation between any one of these factors and word intelligibility was quite low. However, it was also observed that differences in word intensity, duration and F0 movement in a pairwise comparison occurred in the expected directions in a majority (53%) of cases. It was also observed, however, that in 26% of cases, differences occurred in the direction opposite to that expected. An examination of particular cases revealed that 8 of the 10 word pairs in which intelligibility was contrary to expectation also exhibited at least one contrary prosodic cue. In addition, there were 3 word pairs in which intelligibility was as expected even though 2 or 3 contrary prosodic cues were present. Segmental quality did not seem to be a factor.

We might say that if prosody indeed correlates with intelligibility, that it correlates as a whole, being expressed in various combinations of higher intensity, longer durations and more lively F0 contour. Correlation of these prosodic cues with lower context redundancy was somewhat more robust, especially in the case of F0 movement. This suggests that a talker is not always successful in improving word intelligibility for his listener in a low-redundancy context, but that subtle differences in performance intended to bring about this effect are often present.

### REFERENCES

- /1/ Lindblom, B., The Interdisciplinary challenge of speech motor control. In S. Grillner, B. Lindblom, J. Lubker and A. Persson (eds.), Speech Motor Control, Oxford, pp. 3-18, 1982.
- /2/ Lieberman, P., Some effects of semantic and grammatical context on the production and perception of speech. Language and Speech, 6, 172-187, 1963.

/3/ Hunnicutt, S., Intelligibility versus Redundancy - Conditions of Dependency. Language and Speech, 28, 47-56, 1985.

/4/ Lindgren, R., Informationsteoretiskt viktade testsatser för talperceptionstest. Joint publication of Karolinska Institutet and Stockholms Universitet, Stockholm, 1982.

/5/ Rolf Carlson, MIX, Dept. of Speech Communication and Music Acoustics, Stockholm, 1984.

/6/ Box, G., Hunter, W. and Hunter, J., Statistics for Experimenters, New York, 1978.