# SPEECH PERCEPTION IN NOISE: A COMPARISON BETWEEN SENTENCE AND PROSODY RECOGNITION

**Marianne van Zyl, Johan J. Hanekom**

Bioengineering research group, Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria, South Africa

## Abstract

The perception of speech in the presence of interfering noise remains an important issue in the field of audiology. Successful perception of speech under adverse listening conditions is facilitated to a large extent by the redundancy of the speech signal. An important cue that contributes to the redundancy of the speech signal is prosody, or suprasegmental speech features. The present study investigated the acoustic cues of a particular prosodic pattern, validated its recognition in quiet, and assessed its recognition in noise by normal-hearing listeners. The prosody under investigation was conditional permission, approval or agreement. A collection of sentences were recorded from two speakers (one male, one female). Two versions of each sentence were recorded, one giving unconditional permission or approval and the other adding a condition which was subsequently removed from the digital recording to eliminate differences in content between the two versions while retaining prosodic differences. Recorded materials were validated in a group of normal-hearing listeners (n=12) in a quiet listening condition. The recognition of the prosodic contrast was evaluated in a second group of listeners (n=9) in speech-weighted noise, at three different signal-to-noise ratio's (SNRs) and compared to recognition of words and sentences at the same SNRs. Findings indicated that the recognition of sentences and of words in sentences deteriorated significantly as the SNR deteriorated, while recognition of prosody did not, remaining significantly above chance, even at an SNR of -8 dB. These findings indicate the resilience of the prosodic pattern under investigation to the effects of noise.

**Key words:** speech recognition • prosody • signal-to-noise ratio

## Background

Speech recognition in background noise remains a great challenge to all listeners, especially those relying on amplification devices. For this reason, the perception of different speech cues in noise has received much attention in hearing research. One group of speech cues that has, however, not been investigated extensively is suprasegmental or prosodic speech features. Prosody includes such features as intonation, stress and juncture [1], and fulfil many important functions in communication, such as differentiating similar words with different stress patterns [2,3], distinguishing questions from statements [4–6] or conveying the emotion or attitude of a speaker [7–11]. Despite its important communicative functions, the success with which these cues are perceived in difficult listening situations have not been investigated. The present work explored this issue by comparing the recognition of a prosodic pattern that occurs on sentence level with the recognition of words in a sentence. The prosodic pattern selected for this experiment involved the speaker giving approval, permission, or agreement either unconditionally or conditionally (with reluctance).

## Material and Methods

### Speech material

All speech materials used in the study were digitally recorded in a sound-proof booth, using an M-Audio Fast Track Pro external sound card (sampled at 44.1 kHz with 24-bit resolution) and a Sennheiser ME62 microphone. Two sets of sentences were recorded, both in Afrikaans, a language native to South Africa. The first set of sentences was compiled to represent the selected prosody and was recorded using one male and one female speaker. Each sentence contained either permission for, approval of, or agreement with some statement. Two versions of each sentence were recorded. In the first version, the permission, approval, or agreement was unconditional and in the second version it was followed by a condition that was introduced with the word "but". All recordings were edited to leave equal amounts of silence (approximately 100 ms) before and after each utterance, and to equate the mean intensity (rms) value of each sentence to 70 dB SPL using "Praat" software [12]. The utterances that ended with a conditional phrase were edited to remove this part of the utterance (including the word "but"), so as to make the two versions of each utterance identical in content, with only the prosody differing. The recognition of the selected prosody was validated in a group of normal hearing listeners (n=12) in quiet. Average re-cognition scores of 97% were found for both the male and female speakers' material, with standard deviations of 5.6 and 6.6% respectively.

The second set of sentences, used for the sentence and word recognition task, were previously developed for a test of sentence recognition in noise and found to be of equivalent difficulty in noise [13]. For the present work, the material was re-recorded in order to have the same speakers as for the prosodic materials. Both sets of sentences were combined with speech-weighted noise (specific to each

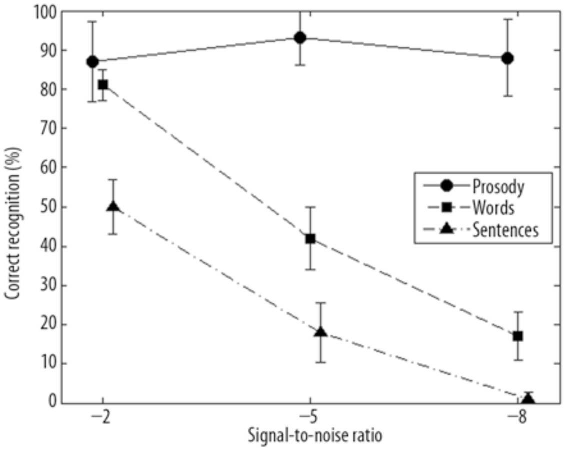

**Figure 1.** Recognition scores of normal-hearing listeners (n=9) of speech materials recorded from a female speaker. Error bars indicate 95% confidence intervals.

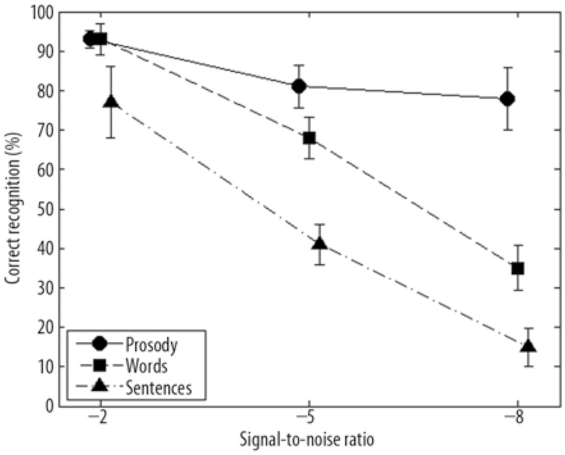

**Figure 2.** Recognition scores of normal-hearing listeners (n=9) of speech materials recorded from a male speaker. Error bars indicate 95% confidence intervals.

speaker), at three different signal-to-noise ratio's (SNRs) (–2, –5, and –8 dB SNR).

### Subjects

Nine listeners participated in this experiment (five male, four female). All participants were young adults (ages 19–25 years), native speakers of Afrikaans (the test language), and had normal hearing (pure tone thresholds ≤20 dB HL at 250, 500, 1000, 2000, 4000, and 8000 Hz). Informed consent was obtained from each listener prior to testing, and listeners were rewarded at the standard hourly fee of the research group. The research was approved by the relevant Ethics Committee at the institution where the experiments were conducted.

### Procedures

Subjects were seated in a sound-proof booth with the examiner for the duration of each experiment. Test materials were presented via an M-Audio Fast Track Pro external sound card connected to a personal computer, through an M-Audio EX66 Reference Monitor (–3 dB bandwidth from 37 Hz to 22 kHz, with flat frequency response in between that allows maximum variation of ±1 dB). Listeners were seated approximately one meter from the loudspeaker, facing it squarely. Materials were presented at 65 dB SPL as measured at the ear level of the test subject. The pre-sentation of the test items was controlled by the administrator and counterbalanced between subjects. Testing was preceded by a number of practice runs in order to reduce practice effects. Testing was conducted across three sessions, each time at a different SNR, and with two weeks waiting time in between sessions to minimise any residual learning effects.

Prosodic recognition performance was calculated as the percentage of sentences for which the prosodic version (conditional/unconditional) was identified correctly. Word recognition scores were calculated as the percentage of words repeated correctly from three phonemically matched lists, and sentence recognition was scored as the percentage of sentences that were repeated correctly in their entirety. The Wilcoxon signed-rank test was used to compare means of the different SNR conditions and different speech materials at each SNR. Confidence intervals were calculated using a student's *t*-distribution, owing to the small sample size.

## Results

Results are depicted in Figures 1 and 2, for recordings from the male and female speaker separately. Findings from both speakers indicate that the recognition of words and sentences deteriorated significantly more than the recognition of prosody as the SNR decreased. Differences were particularly noticeable at –8 dB, where recognition of words and sentences were significantly worse ($p<0.01$) than at –5 dB for both speakers, while prosody recognition did not deteriorate significantly at this level.

## Discussion

The findings of this study indicate that the recognition of the prosodic contrast investigated here is more resilient to the effects of background noise than the recognition of a whole sentence or words in a sentence. This is in agreement with the findings of Mattys [14], who demonstrated that syllable stress, a prosodic cue to word boundaries, was also resilient to background noise. The prosodic pattern under investigation was somewhat more complex than just syllable stress, containing cues related to stress or emphasis, intonation pattern, speech rate and even voice quality. The redundancy of the pattern may have contributed to its robustness, although the redundancy of the sentence materials used for word and sentence recognition did not prevent these materials from being severely affected by noise.

## Conclusions

The present study found that the recognition of a prosodic pattern in background noise was an easier task for the normal-hearing participants than the recognition of words in a sentence in the same level of noise. The findings warrant further investigation into the resilience of prosodic cues in noise, as these cues might contribute to successful speech perception in adverse listening conditions.

## References:

1. Raphael LJ, Borden GJ, Harris KS. Speech Science Primer: Physiology, Acoustics, and Perception of Speech. 5th ed. Philadelphia: Lippincott Williams & Wilkins, 2007
2. Fry DB: Duration and intensity as physical correlates of linguistic stress. Journal of the Acoustical Society of America, 1955; 27(4): 765–68
3. Lieberman P: Some acoustic correlates of word stress in American English. Journal of the Acoustical Society of America, 1960; 32(4): 451–54
4. Caspers J: Who's next? The melodic marking of question vs. continuation in Dutch. Language and Speech, 1998; 41(3–4): 375–98
5. Van Heuven VJ, Van Zanten E: Speech rate as a secondary prosodic characteristic of polarity questions in three languages. Speech Communication, 2005; 47: 87–99
6. Vion M, Colas A: Pitch cues for the recognition of yes-no questions in French. Journal of Psycholinguistic Research, 2006; 35(5): 427–45
7. Hammerschmidt K, Jürgens U: Acoustical Correlates of Affective Prosody. Journal of Voice, 2007; 21(5): 531–40
8. Murray IR, Arnott JL: Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. Journal of the Acoustical Society of America, 1993; 93(2): 1097–108
9. Williams CE, Stevens KN: Emotions and Speech: Some Acoustical Correlates. Journal of the Acoustical Society of America, 1972; 52(4): 1238–50
10. Pell MD: Reduced sensitivity to prosodic attitudes in adults with focal right hemisphere brain damage. Brain and Language, 2007; 101: 64–79
11. Fujie S, Ejiri Y, Kikuchi H, Kobayashi T: Recognition of positive/negative attitude and its application to a spoken dialogue system. Systems and Computers in Japan, 2006; 37(12): 45–55
12. Boersma P, Weenink D: Praat: doing phonetics by computer [computer program]. Version 5.1.32 *http://www.praat.org/*; 2010.Last accessed: 29 November 2010
13. Theunissen M, Swanepoel D, Hanekom JJ: The development of an Afrikaans test of sentence recognition thresholds in noise. International Journal of Audiology, 2011; 50(2): 77–85
14. Mattys SL: Stress Versus Coarticulation: Toward an Integrated Approach to Explicit Speech Segmentation. Journal of Experimental Psychology: Human Perception and Performance, 2004; 30(2): 397–408