ISCA Archive
http://www.isca-speech.org/archive

Speech Prosody 2002
Aix-en-Provence, France
April 11-13, 2002

# Pitch Contour Guides Spoken Word Recognition

*Claudia K. Friedrich, Sonja A. Kotz, Angela D. Friederici & Kai Alter*

Max-Planck-Institute of Cognitive Neuroscience
Stephanstraße 1A, 04301 Leipzig, Germany
fried@cns.mpg.de

10.21437/SpeechProsody.2002-63

## Abstract

Three experiments investigated the processing of pitch contour during spoken word recognition. In Experiment 1 event-related brain potentials (ERPs) were recorded while subjects made decisions to artificially pitch manipulated bisyllabic words. ERPs revealed that pitch contours are discriminated already within the first syllable of a word. In Experiment 2 subjects heard spoken word-initial syllables with re-synthesized pitch contours and were asked to name a word starting with that syllable. They answered more likely with an initially stressed word if the syllable carried a stressed pitch contour. However, if the same syllable carried an unstressed pitch contour, subjects more frequently responded with an initially unstressed word. In a cross-modal priming paradigm we presented pitch modulated initial syllables as auditory primes followed by visual targets (Experiment 3). Reaction times as well as ERPs were found to be sensitive to pitch information. Taken together the results indicate that pitch is automatically extracted during spoken word recognition and that this prosodic parameter guides lexical access.

## 1. Introduction

A spoken word with more than one syllable contains a specific stress pattern found to be processed during spoken word recognition (see [1] for a recent review). We conducted three experiments to investigated the role of a word's pitch contour during spoken word recognition, as pitch contour is one of the most important parameters that mark word stress in German.

In Experiment 1 we examined whether re-synthesized words that differ only in their pitch contour elicit different responses of the brain. Experiment 2 investigated if listeners exploit the pitch contour of word initial syllables for word completion. In the last experiment we presented pitch manipulated syllables from Experiment 2 as primes followed by visual target words. This paradigm is referred to as cross-modal word fragment priming and it is assumed to test lexical activation in spoken word recognition [2]. Experiment 3 was motivated by research that reported effects of the fragments' prosody on lexical decisions for targets with different stress patterns [3,4,5]. Thus, we wanted to explore with cross-modal word fragment priming, whether or not pitch contour as a single prosodic feature might modulate lexical activation.

## 2. Experiment 1

### 2.1. Method

#### 2.1.1. *Stimuli*

A set of 160 bisyllabic German words was used for all three experiments. 80 words were initially stressed, (e.g., TANgo [Engl.: tango]; Note that stress is marked with capital letters.), 80 were initially unstressed (e.g., faSAN [Engl.: pheasant]). The words were spoken by a female native speaker of German and digitized (44.1 kHz/16 bit sampling rate). For Experiment 1 an initially stressed pitch contour was applied to one version of each word and an initially unstressed pitch contour was applied to another version of the same word (see Figure 1 a,b for an example of both pitch manipulated versions of the word 'Tango').

12 subjects, that did not participate in Experiment 1, rated if the re-synthesized words sounded like a normal word or not. 40 initially stressed and 40 initially unstressed words that were rated as naturally sounding in both pitch versions were presented in Experiment 1.

#### 2.1.2. *Subjects and procedure*

Twenty-four right-handed native speakers of German evaluated the stress pattern (correct vs. incorrect) in one task and made animacy judgments (animate vs. inanimate) in another task. Words were presented via loudspeakers. Presentation order of correctly and incorrectly stressed words was counterbalanced within and across subjects. Task order and response hand were counterbalanced across subjects.

#### 2.1.3. *Electrophysiological response*

The EEG was continuously recorded (250 Hz/12 bit sampling rate) from 58 Ag-AgCl electrodes. All electrodes were referenced against the nose tip. Impedance was kept below 5 kΩ. Artifacts caused by facial and eye movements and by drifts were rejected offline.

### 2.2. Results and Discussion

An incorrect pitch contour (e.g., TANgo with initially unstressed pitch) delayed the reaction times (RTs) as compared to correct pitch (e.g, TANgo with initially stressed pitch contour). That is, across both tasks RTs were faster for words with correct pitch contour (mean ± s.d. 1228 ± 223 ms) than for the same words with incorrect pitch contour (1250 ± 226 ms; $t[1,23] = 5.39$, $p = .01$).

The ERPs for words with stressed pitch contour differed from ERPs for the same words unstressed pitch contour in a time window between 200-280 ms after word onset over posterior electrode positions (see Figure 1c). Words with initially unstressed pitch contour elicited a more positive ongoing P2 component in the ERP (-2.9 ± 3.2 µV) than the same words with initially stressed pitch contour (-3.6 ± 3.5 µV, $t[1,23] = 7.37$; $p = .01$).

The RTs reflected that the pitch contour of a spoken word modulates recognition processes of that word. The ERPs revealed that already the pitch contour of the first syllable is extracted by the brain. Consequently, Experiment 2 explored whether already the pitch information of the first syllable is used during spoken word recognition.

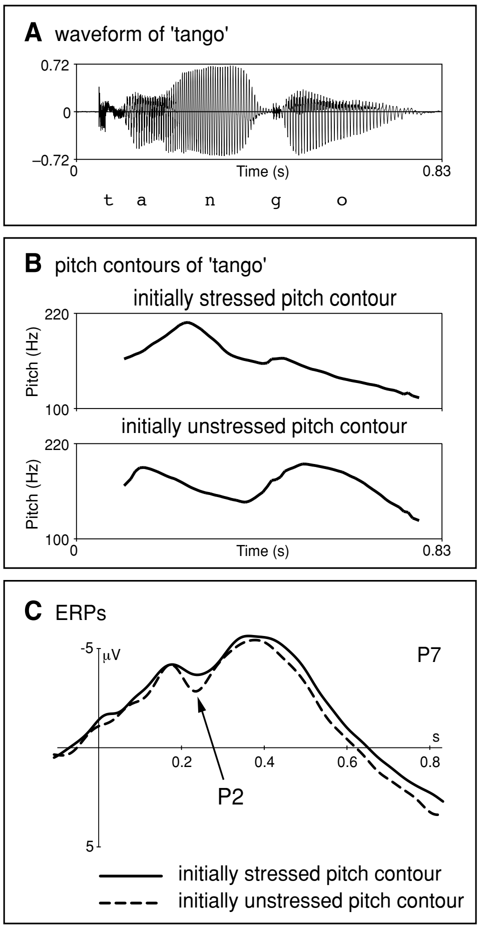


*Figure 1 a) Waveform of the German word 'Tango', b) initially stressed pitch contour (above) and initially unstressed pitch contour (below) of 'Tango', c) ERPs for a selected electrode (P7) for words with initially stressed pitch contour (solid line) and for the same words with initially unstressed pitch contour (dashed line) time locked to word onset. ( Negativity is plotted upwards.)*

# 3. Experiment 2

## 3.1. Method

### *3.1.1. Stimuli*

For Experiment 2 the first syllables of all 160 words were isolated. Again, a stressed pitch contour was applied to one version of each syllable and an unstressed pitch contour was applied to another version of the same syllable (see Figure 2a for an illustration of the pitch manipulation for the syllable 'tan' taken from the word 'TANgo').

### *3.1.2. Subjects and procedure*

Twenty native speakers of German, different from those of Experiment 1, listened to each syllable and were asked to write down the first word starting with that syllable that came to their mind. Presentation order of syllables with stressed and unstressed pitch was counterbalanced within and across subjects.

## 3.2. Results and Discussion

As expected, subjects responded more frequently with initially unstressed words after listening to syllables with unstressed pitch contour (23.8 ± 1.6% of the responses) than after listening to the same syllables with stressed pitch contour (19.6 ± 1.5%, $t[1,19] = 22.85$, $p < .001$). Vice versa, subjects were more likely to respond with an initially stressed word after hearing a syllable with initially stressed pitch contour (80.4 ± 1.5% of the responses) than after hearing the same syllables with unstressed pitch contour (76.2 ± 1.6%, $t[1,19] = 22.85$; $p < .001$).

These results, together with that of Experiment 1, provide clear evidence that already the word-initial pitch contour is exploited by the listener during spoken word recognition. However, the question remains whether or not pitch is used in initial spoken word processing. Thus, Experiment 3 investigated with cross-modal word fragment priming if pitch is mapped onto lexical representations during lexical activation. Next to behavioral responses, ERPs were recorded to investigate lexical access processes with an online method.

# 4. Experiment 3

## 4.1. Method

### *4.1.1. Stimuli*

80 syllables from Experiment 2 were used as auditory primes (e.g. 'tan'). These primes were combined with visual targets. 160 words produced by the subjects in Experiment 2 were used as segmentally matching targets, that is these targets began with the syllable that was used as prime (e.g., 'Tango' [Engl.: tango] and 'Tangente' [Engl.: tangent] for the prime 'tan'). 80 segmentally matching targets were initially stressed

(e.g., TANgo), 80 segmentally matching targets were initially unstressed (e.g., tanGENte). 160 segmentally mismatching targets were matched in stress pattern, word frequency and word length to the segmentally matching targets. Segmentally mismatching targets did not begin with the syllable that was used as prime (e.g., 'Route' [Engl.: route] or 'Meridian' [Engl.: meridian] for the prime 'tan'). Thus, 80 initially stressed segmentally mismatching targets (e.g., ROUte') and 80 initially unstressed segmentally mismatching targets (e.g. meriDIAN) resulted (See Figure 2b for a further illustration of the conditions realized for both pitch versions of 'TAN'). Additionally, 320 phonotactically legal Pseudowords were created by interchanging the last letters of each target (e.g. 'TANOG').

*4.1.2. Subjects and procedure*

Twenty-four right-handed native speakers of German, different from those of the two preceding experiments, made lexical decisions to targets and pseudowords which were presented immediately after the prime offset. Presentation order of primes with stressed and unstressed pitch, of segmentally matching and mismatching targets and of the stress pattern of targets was counterbalanced within and across subjects. Subjects were tested in two sessions. Response hand and session order was counterbalanced across subjects.

*4.1.3. Electrophysiological response*

see Experiment 1.

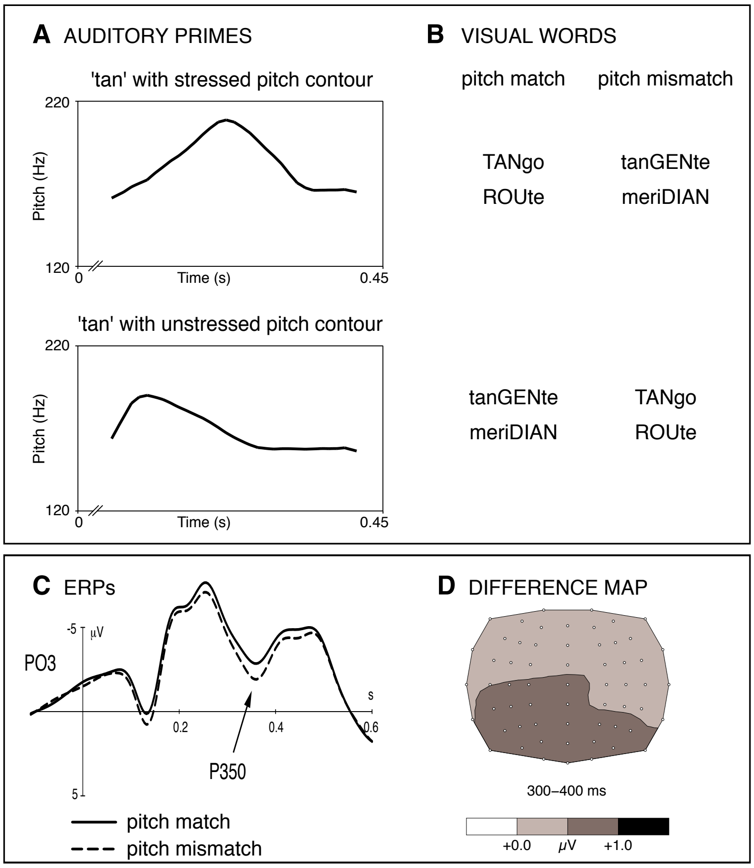


*Figure 2 a) Stressed pitch contour (above) and unstressed pitch contour (below) of the syllable 'tan' used in Experiment 2 and 3, b) Targets that were presented immediately following the auditory syllables 'tan' in Experiment 3, c) ERPs elicited by targets that did (solid line) or targets that did not match the pitch of the prime (dashed line) time locked to target onset in Experiment 3, d) Difference map for the ERPs for pitch matching targets subtracted from those for pitch mismatching targets in a time window between 300 and 400 ms.*

#### 4.2 Results and Discussion

RTs deviating more than 2 standard deviations from the individual mean of the subject were excluded. There was a slight behavioral facilitation for targets with a stress pattern that matched the prime's pitch contour. Targets with a stress pattern as indicated by the pitch contour of the prime were responded to faster (675 ± 61 ms) than targets with a different stress pattern (679 ± 61 ms; $t[1,23] = 4.74$, $p = .04$). Note, that this effect was found irrespective of a segmental match between prime and target. That is, 'TANgo' and 'rouTE' were facilitated as compared to 'tanGENte' and 'meriDIAN' if they were preceded by 'tan' with stressed pitch contour. However, RTs to 'tanGENte' and 'meriDIAN' were facilitated as compared to 'TANgo' and 'rouTE' if they were preceded by 'tan' with unstressed pitch contour.

Additionally, an ERP deflection, named P350, was found to differentiate pitch matching and mismatching targets (see Figure 2c). The P350 was analyzed in a time window of 300 to 400 ms. Mean amplitudes in this time window were for posterior electrode sites more positive for targets that did not match the prime's pitch contour (-3.5 µV ± 4.1 µV) as compared to pitch matching targets (-4.1 µV ± 4.4 µV, $t[1,23] = 4.93$; $p = .04$). Again, this effect was found irrespective of the target's segmental information. That is, the targets 'TANgo' and 'ROUte' elicited a smaller P350 if they followed 'tan' with stressed pitch contour than if they followed the same syllable with unstressed pitch contour.

As predicted, behavioral as well as ERP results of Experiment 3 indicated the mapping of pitch information, as a single prosodic feature, onto lexical representations during spoken word recognition. As cross-modal word fragment priming is thought to tap into lexical access [2,3,4,5], we interpreted the P350 as reflecting this process in the ERP.

## 5. Conclusions

Three experiments clarified the role of pitch during spoken word recognition. Auditory evoked ERP potentials in Experiment 1 indicated that the pitch contour of a spoken word is automatically extracted by the listener. In line with this, the results of a word completion task (Experiment 2) indicated a use of the word initial pitch contour during spoken word recognition. Finally, behavioral as well as ERP results of cross-modal word fragment priming (Experiment 3) confirmed a significant role of the initial pitch information during lexical access in spoken word recognition.

## 6. References


[1] Cutler, A.; Dahan, D.; van Donselaar, D., 1997. Prosody in the comprehension of spoken language: A literature review. *Language and Speech,* 40, 141-201.

[2] Marslen-Wilson, W. D., 1990. Activation, competition and frequency in lexical access. In: G. T. M. Altmann (Ed.). *Cognitive models of speech processing*. Cambridge: MIT Press, 148-72.

[3] Cooper, N., 2000. Native and non-native processing of lexical stress in English word processing. *Proceedings of the Workshop on Spoken Word Access Processes, May 2000, Nijmejen, The Netherlands,* 163-166.

[4] Cutler, A.; van Donselaar, W., 2001. Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech,* 44**,** 171-195.

[5] Soto-Faraco, S.; Sebastian-Galles, N.; Cutler, A., 2001. Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*, 45, 412-432.