Chapter seven

Prosodic markers of the statement – question contrast in Kutai Malay

Sugiyono
Pusat Pembinaan dan Pengembangan Bahasa, Jakarta

7.1 Introduction

Kutai Malay is spoken along the Mahakam River in the Kutai regency, East Kalimantan Province. Its speakers call it a language, but it is labeled a dialect of Malay by Wurm & Hattori (1981).

According to Collins (1992) Kutai is used in daily life by people along the Mahakam River. However, other Malay dialects are spoken in Kutai as well, viz. Banjar Malay, Berau Malay and Pasir Malay, and also the languages of newcomers, like Javanese and Buginese. Pernyata (1992: 2) states that the majority of inhabitants of Kutai (around 784,860 people) use the Kutai language. Kutai is spoken in at least 15 out of the 32 districts of Kutai regency; Mursalim & Gazali (1995) claim that it is spoken in the 18 districts that they researched. In these 18 districts they attested the following five varieties: Kutai Tenggarong, Kutai Kotabangun, Kutai Muara Muntai, Kutai Muara Ancalong and Kutai Melak. Amongst these, Kutai Tenggarong is seen as the principal/original variety as it has more speakers and is more wide-spread than the others. Moreover, the area where it is spoken includes Tenggarong, which is the cultural centre of Kutai. Tenggarong is therefore called the centre of the Kutai language (Mursalim & Gazali 1995: 329).

Like other Malay languages, Kutai Malay has six vowels, i.e. two front vowels /i, e/, two mid vowels /ə, a/ and two back vowels /u, o/ (Suryadikara, Dursid, Kawi & Ismail 1979: 8; Collins 1992: 7). In comparison with the consonant systems of Indonesian and other modern Malay languages, Kutai Malay does not have many fricatives. Fricatives like labiodental /f/, palatal /ç/ and alveolar /z/ do not occur. As regards prosody Suryadikara et al. (1979: 10) mention that there are no supra-segmental phonemes, although there are indications that certain speech sounds are lengthened. The resemblance in the sound system is paralleled by a similarity in lexicon, in which Kutai Malay resembles Indonesian (and other varieties of Malay) very much.

7.2 Prosodic characteristics of statement and question; speech production data

A production experiment was set up to study the melodic and temporal structure of statements and questions in Kutai Malay. The melodic structure will be described by measuring melodic features such as onset and terminal pitch, peak height and pitch range. For each of these characteristics we will not only determine the mean value within and across speakers but also their variability. These gross characteristics serve as a frame of reference against which the individual pitch movements can be expressed in a speaker-independent way, so that, ultimately, profiles for statement and question intonation can be obtained.

7.2.1 Methods

For our materials we chose statement and question pairs that have exactly the same lexico-syntactic structure, so that the functional difference can only be signaled through prosody. Such pairs cannot be easily obtained from spontaneous or quasi-spontaneous recordings. Therefore, play-acting seems the only reasonable way of eliciting the materials. Subjects were asked to read out sentences differing in length, as follows:

a. *Sida busu mancing.* [SV]
[sida busu]$_{NP}$ [maɲciŋ]$_{VP}$
'Uncle is fishing'.

b. *Sida busu mancing jukut patin.* [SVO]
[sida busu]$_{S}$ [maɲciŋ]$_{V}$ [ɟukut patin]$_{O}$
'Uncle is fishing patin fish.'

c. *Sida busu mancing di Mahakam.* [SVAdv]
[sida busu]$_{S}$ [maɲciŋ]$_{V}$ [di mahakam]$_{Adv}$
'Uncle is fishing in the Mahakam.'

d. *Sida busu mancing jukut patin di Mahakam.* [SVOAdv]
[sida busu]$_{S}$ [maɲciŋ]$_{V}$ [ɟukut patin]$_{O}$ [di mahakam]$_{Adv}$
'Uncle is fishing patin fish in the Mahakam.'

These target sentences were realized as statements, with various sub-modes, like (A) answer statement, (B1) confirmation statement, and (C) contrastive statement, and as questions with various sub-modes like (B2) echoic-agreement question and (D) echoic-contrastive question, (E, F, G) confirmatory tag question with the particles *-kah*, *-kan* and *-yo*, and (H) informative question, for which in this research those with the adverb *bilakah* were chosen. The number of target sentence types to be realized by each speaker was 48, as summarized in Table 1. The full set of stimulus sentences is listed in the appendix to this chapter.

Table 1: Stimulus types used in experiments

Statements	N	Questions	N
(A1) Answer: broad focus on VP	6	(B2) Echo Q	4
(A2) Answer: narrow focus within VP	6	(D) Declarative Q	4
(B1) Agreement statement	4	(E) Tag Q (*-kah*)	4
(C) Contrastive statement	8	(F) Tag Q (*-kan*)	4
		(G) Tag Q (*-yo*)	4
		(H) Wh Q (information Q)	4
Total	24		24

Each stimulus type was presented three times so that three tokens of each type were obtained per speaker. Stimuli were presented in printed form on sheets of paper. Recordings were made on a Sony WM-D6C cassette recorder using a head-mounted Shure SM10A close-talking microphone.

7.2.2 Subjects and corpus

At first I hoped to be able to include monolingual native speakers of Kutai Malay (KM) as subjects, but no such speakers can be found any longer. Any speaker of KM is also a speaker of Indonesian. Subjects were between 20 and 55 years of age, had lived in Tenggarong during their entire life, had no higher education, and did not have any speech or hearing defects.

Altogether there were 14 speakers (eight male and six female). From these the primary data were collected, i.e. 2,016 utterances (14 speakers × 48 targets × 3 repetitions). These primary data were subsequently subjected to a perceptual screening test which involved four native KM listeners, who had not taken part in the recordings. They selected the 672 best utterances from the data set such that the best token of each triplet of repetitions was selected, reducing the dataset to one third. The listeners were instructed to determine, first of all, whether the utterance was a statement or a question and, secondly, which of the three tokens had the most natural intonation.

7.2.3 Analysis and results

Raw F0 curves were produced for the 672 best target utterances by the autocorrelation pitch extraction algorithm as implemented in the Praat speech processing software (Boersma & Weenink 1996). The raw curves were manually corrected. The following characteristics of the F0 curves were then measured and collected in a database for off-line statistical processing:

(1) Onset: the first reliable pitch point in the contour
(2) P1_pre: the low (valley) preceding the pitch peak at the end of the first prosodic constituent *sida busu* (syntactically the subject)
(3) P1_peak: the highest pitch associated with this first phrase
(4) P1_trail: the lowest pitch after (3)
(5) P2_pre: the lowest pitch preceding the peak on the predicate *mancing*
(6) P2_peak: the highest pitch on the predicate
(7) P2_trail: the lowest pitch point following (6). This low always had the same frequency as the terminal pitch of the utterance.

From these seven measurements several more complex measures can be derived, such as the pitch range associated with the utterance (difference, in semitones, between the highest peak – whether P1_peak or P2_peak – and the lowest low), and the pitch interval of specific rises and falls (also expressed in semitones)[54]. Longer and more complex utterances were first divided into prosodic phrases, and then stylized with three pitch points per phrase, i.e., a pitch peak (associated with the last word in the phrase) preceded and followed by a low. These longer utterances are not discussed in this paper.

The temporal features were described in a simpler way. I started with the segmentation of the individual segments forming the utterance. Although the segmentation was done on the level of individual phonemes, duration was not measured on this level but on the level of the syllable. The aim of the measurements was only to differentiate the temporal features of statements from those of questions in pairwise comparisons; more fine-grained temporal organization at the segmental level will be disregarded for the time being.

The result of the acoustical analysis indicates that, on the whole, acoustic features like onset pitch, final pitch, pitch range, pitch peak, pitch movements and duration can be markers of the contrast between declarative and interrogative utterances. But comparison of each of these acoustic features indicates that some are better discriminators of clause type than others.

The *pitch range* of KM statements and questions taken together ranges from 50 Hz (5 st) to 367 Hz (22 st) with an average of 133 Hz (10 st). Apart from the variable *gender* – the pitch range of the women is larger than that of the men ($p < .001$) – on the whole the clause type also has a significant influence ($p = .004$); I found evidence that statements (9.9 st) have a smaller pitch range than questions (10.6 st). Thus, from these data we can conclude that KM utterances have a pitch range of roughly almost one octave (12 st) and that there is evidence that the range discriminates the clause types, even though the difference in range between statement and question is only .7 st.

Trivially, female voices have higher frequencies than male voices. The difference in *onset pitch* between male and female pitch is very significant ($p < .001$). In both statements and questions the male onset pitch (182 Hz) is approximately 8 st lower than that of the female speakers (283 Hz).

[54] Pitches are, rather arbitrarily, expressed in semitones above the C_1 on the piano keyboard, i.e. 65.5 Hz.

Less obviously, the statement contour starts at 233 Hz on average whereas the question contour starts at 218 Hz, i.e., statements start about 1 st higher than questions. Although, in general, statements start somewhat higher than questions, when the types are compared separately, the onset frequency of the statement only differs from that of the confirmatory and informative questions. The onset pitch of statements (9.1 st) is approximately the same as that of echo questions (9.7 st), but both are higher than the onset of confirmatory-question contours (7.9 st), and all three of them are higher than the onset of the informative questions (5.6 st; cf. Figure 1). The onset pitch is relatively stable regardless of the different durations and number of constituents of the utterances.

Apart from the predictable effect of gender ($p < .001$), the *final pitch (P2_trail)* also clearly signals the difference between statement and question ($p < .001$). The F0 of the final pitch of the statement amounts to 211 Hz against 247 Hz in questions (2.3 st higher).

Neither onset pitch nor final pitch is influenced significantly by the length of the utterance (expressed in number of constituents). The final pitch in SV as well as SVO utterances is at 8.7 st, in SVAdv 8.5 st, and in SVOAdv 8.0 st. The trend suggested here, however, does not reach significance.

The higher the onset pitch the higher also the final pitch; this holds for both statement and question contours. In statements, however, the final pitch is lower than the onset pitch, with the tendency that the higher the onset pitch, the larger the difference between onset and final pitches in the contour. Conversely, in the question contour the final pitch is higher than the onset pitch, with again the tendency that the higher the onset, the larger the difference between onset and final pitches. The difference between onset and final pitch is very significant ($p < .01$).

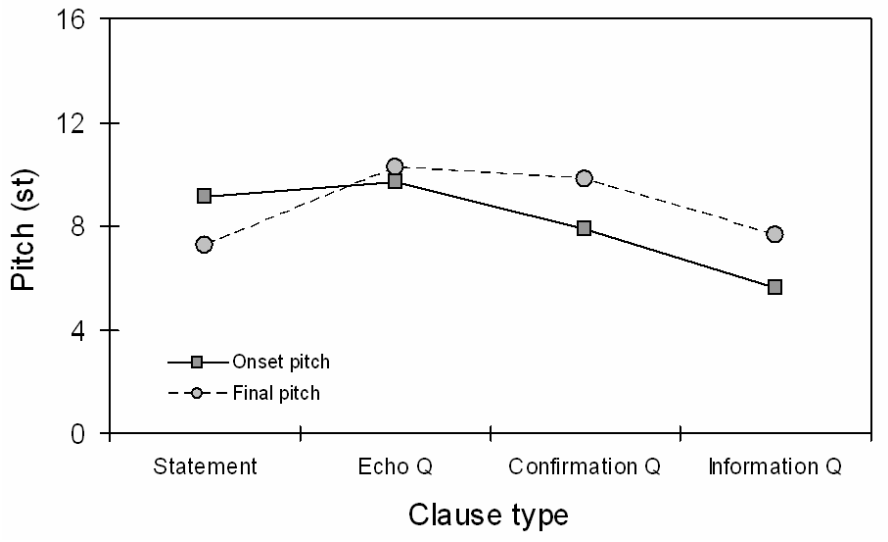

Figure 1: Onset and final pitch (semitones) of utterances broken down by clause type.

As already mentioned, in questions the final pitch is always higher than the onset pitch, whereas the final pitch of the statements is lower than the onset. The difference between onset and final pitches of echo questions is only around .6 st, yet it is significant. The differences between onset and final pitches in the other utterance types amount to 1.8 st to 2.1 st. Amongst the four types in Figure 1, the difference between onset and final pitches is largest in the information questions.

After this general comparison of statements and questions I will now focus on the differences between statement and echo question contours with two constituents. This contrast is visualized in Figure 2.

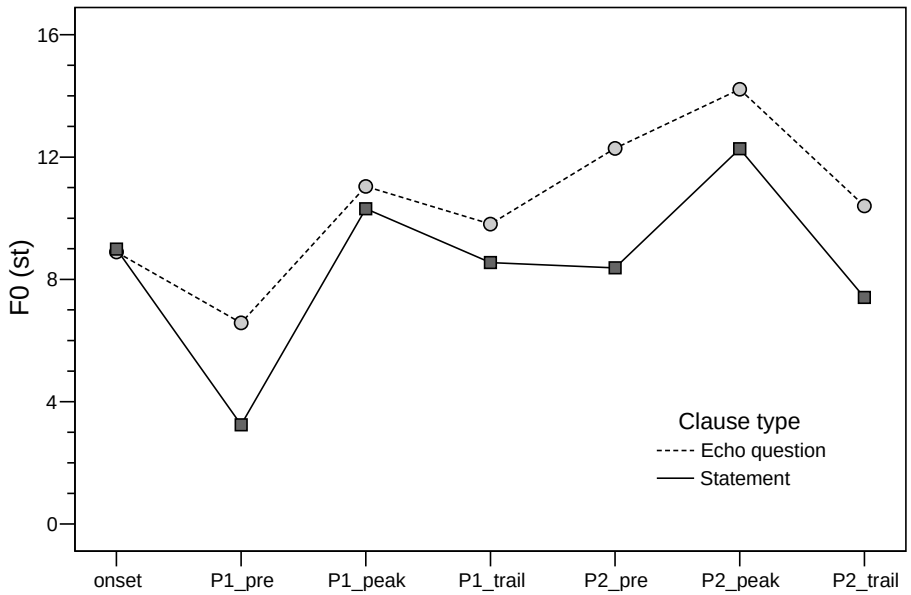

Figure 2: Statement and echo question contours in simple SV sentences.

The pivot pitches differ between statement and echo question at P1_pre, P2_pre, P2_peak, and P2_trail (all $p < .1$). The P1_pre of the statement (3.2 st) is lower than that of the (echo) question (6.6 st). The P1_peak of statements (10.1 st) is roughly as high as that of the questions, but the interval between onset and P1_peak of the statement (1.0 st) is slightly smaller than that of the question (2.3 st). The P2_pre of the statement (8.5 st) is lower than that of the question. Notice that the rise towards P2_peak of the echo question already starts at the P1_trail pivot point (at 9.8 st). The statements' P2_pre tends to be at the same pitch as the contour onset, whereas the pitch of the echo question at the P2_pre position is approximately 3.4 st higher than the contour onset. The second pitch peak (P2_peak) of the statement contour (12.3 st) is lower than the P2_peak of the echo question (14.2 st). The difference between onset pitch and P2_peak amounts to 5.5 st in the question, and 3.1 st in the statement contour, pitch peaks being higher than onsets.

The contrast between statement and question is clearly signaled by their final pitch movements. Statements show Level-Rise-Fall, and questions Rise-Fall final pitch movements. In utterance-initial and medial positions of both statement and question, pitch movements usually consist of Level-Rise or Fall-Rise pitch movements with varying degrees of steepness. Finally, the P2_trail of the statement (7.4 st) is lower than that of the echo question (10.4 st). The largest difference between statement and echo question is in the P2_pre. This seems due to the fact that the rise towards the second peak starts at the end of the fall of the first peak in the question version, but is separated from the latter by a stretch of low pitch in the statement.

A more general comparison, then, of the F0 in the statement and echo question contours shows that there is a significant difference for each of the seven pivot points, with the exception of the utterance onset. Moreover, there seems to be a tendency for the difference between the two clause types to increase towards the end of the utterance. The correlation between the difference in F0 at each of the seven pivot points (question minus statement) and the ordinal position of the point from the beginning of the utterance is r = .508 (p = .130, one-tailed). The relevant data have been plotted in Figure 3.

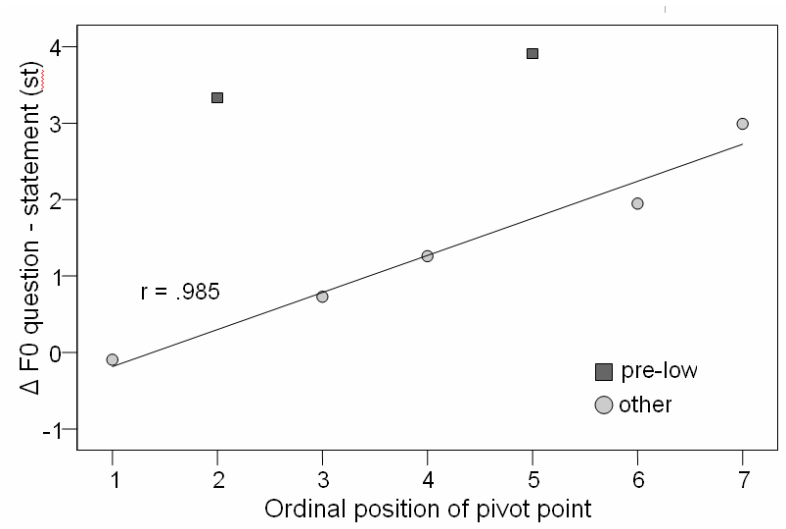

Figure 3: Pitch difference between echo question and statement utterance at each of seven pivot positions in utterance. Pre-low: P1_pre and P2_pre. The correlation coefficient relates to the five 'other' points only.

Closer inspection of the data, however, reveals that the trend towards more strongly diverging pitches as the utterance proceeds is very strong indeed if we separate between the lows preceding peaks on the one hand (two points, no meaningful correlation can be computed), and all other pivot points on the other (*r* = .985, p = .001, one-tailed). This clearly suggests, then, a global difference between the

statement and (echo) question intonation such that the pitch of the utterance generally falls (or remains level) from beginning to end of the sentence in the former, but gently rises in the latter clause type. Such global trend differences have been noted in other languages as well, e.g. Copenhagen Danish (Thorsen 1978) and Dutch (van Heuven & Haan 2000, Haan 2001). A second, more local effect, seems to be that the lows preceding the peaks in the questions are less pronounced than in the statements, leading to flatter F0 curves in questions.

Finally, *syllable duration* turns out to be a very significant discriminator. Overall, statements are longer than questions ($p < .001$). Since statements and questions in our materials differ in length and complexity, I have limited the comparison to short SV statements and the corresponding declarative question versions, which are lexico-syntactically identical. Figure 4 plots the mean syllable durations of these two clause types.

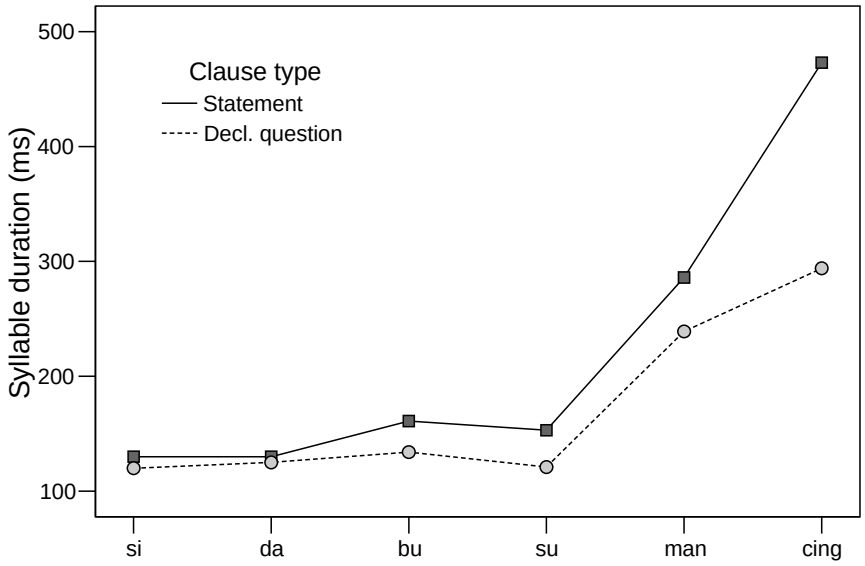

Figure 4: Mean syllable duration in statements and corresponding declarative questions.

Figure 4 shows a lengthening of the last four syllables in the statement relative to the corresponding question. It is not clear how this effect should be interpreted. My proposal would be to consider the lengthening on *busu* as the marking of an utterance-internal phrase boundary, and the much stronger lengthening on *mancing* as utterance-final lengthening in statements, which is considerably suppressed in questions. Recent research by van Heuven & van Zanten (2005) reveals that there is

a tendency, cross-linguistically, for (declarative) questions to be spoken faster than the corresponding statements. KM conforms to this cross-linguistic finding.[55]

7.3 Perceptual thresholds as markers of the declarative-interrogative contrast

As was shown by Figures 1 to 4 above, KM statements differ from the corresponding questions along several dimensions simultaneously. The most obvious differences between statement and (echo) question are (i) flatter F0 curve for questions, specifically implemented by raising the lows preceding pitch peaks on content words, (ii) a global uptrend (inclination) of F0 in questions as opposed to a flat or downward trend in statements, starting from the same pitch at the beginning of the utterance, (iii) a qualitative difference in the rise-fall pattern on the last content word in the utterance, and (iv) slower speaking rate in the statements, especially in the last two syllables in the utterance. In the perception studies to be reported below I tried to determine the relevance of each of these dimensions.

Four perception experiments were done to determine the threshold of the F0 excursion (experiment I), the durational threshold (experiment II), the contrast threshold (experiment III), and the acceptability of the basic sentence contour (experiment IV). With experiments I to III this research endeavors to determine the lower and upper threshold values of the prosodic markers that signal the statement versus (echo) question contrast. Experiment IV will test the acceptability of the basic sentence intonation of several utterances as statement or question, both with and without lexical markers of clause type. Experiment IV will also give an indication of the applicability of the results of experiments I-III, which only tested statement and echo question, to other sentences types.

7.3.1 Technical procedures

In each perception experiment the stimuli were constructed within a paradigm known as *constant stimuli* in the psychoacoustic literature (Small 1973: 254). In this paradigm prosodic features are systematically manipulated with fixed step sizes. For the four experiments altogether 625 stimuli were presented.

The stimuli were based on the statement and echo-question utterances with an SV structure. The utterances were chosen from a male speaker with a high score; the contours closely resembled the typical profile contours found in the acoustic analysis. In this way, the basic stimuli were truly ideal utterances, which, however, differed from the mean values of production data. Pivot point positions and pitches of the basic stimulus contours are specified in Table 2.

---

[55] The faster rate and stronger pre-pausal lengthening may be (partly) artefactual. The question tokens were recorded in the middle of paragraphs whilst the statements typically occurred in final position. Paragraph boundaries are deeper than utterance boundaries and are known to be more strongly affected by domain-final lengthening (Lehiste 1970, Sluijter & Terken 1993). However, paragraph-final position does not have effects as large as the difference found here.

Table 2: Position (ms from start) and pitch (in st re. 65.5 Hz) at seven pivot points in the basic statement and question stimulus contours, and difference between the two (Δ).

	Onset		P1_pre		P1_peak		P1_trail		P2_pre		P2_peak		P2_trail	
	ms	st	ms	st	ms	st	ms	st	ms	st	ms	st	ms	st
Statement	0	10.0	430	3.9	530	11.2	693	9.5	852	9.4	1000	11.3	1298	8.5
Question	92	9.8	352	7.2	412	11.6	474	10.5	−*	−*	762	15.2	932	11.8
Δ		−0.2		3.3		0.4		1.0				3.9		3.3

* Question has no P2_pre pivot point, as P1_trail coincides with start of rise towards P2.

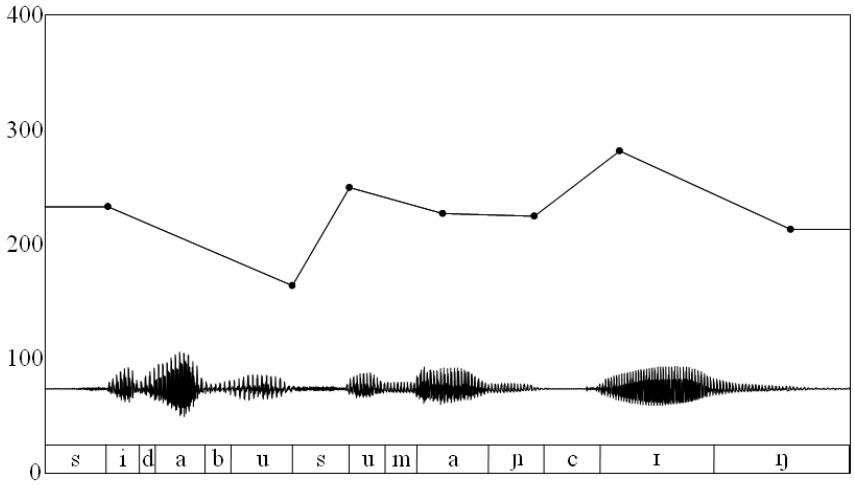

Figure 5: Stylized pitch contour on basic statement stimulus.

In the basic statement contour (cf. Figure 5) P1_pre coincided with the end of the /u/ of the pre-final syllable of the noun phrase *sida busu* ‘uncle’, whereas P1_peak coincided with the start of the final vowel /u/) of this noun phrase. In the second configuration, P1_trail was positioned at the beginning of the /a/ of the first syllable of the verb *mancing*. In this syllable P2_pre was also situated, whereas P2_peak and P2_trail fell on the final syllable of this verb. P2_peak fell at the beginning of the vowel /ɪ/ whereas the terminal pitch P2_trail was reached at the temporal midpoint of the consonant /ŋ/.

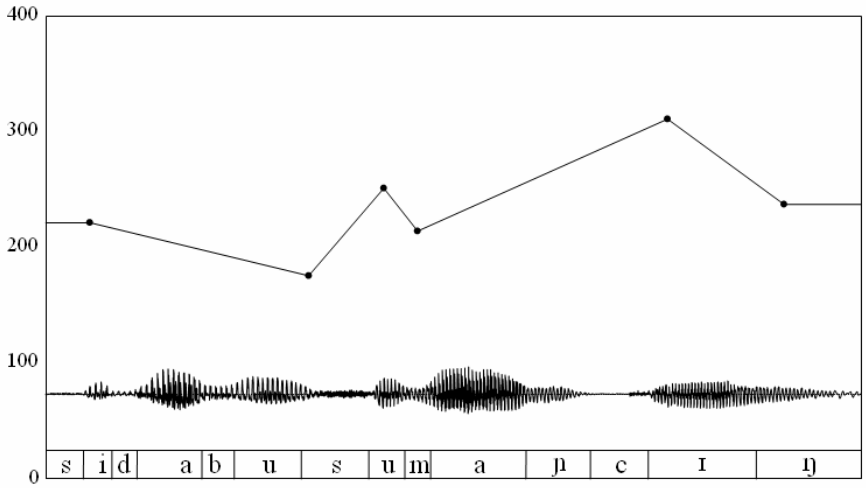

Figure 6: Stylized pitch contour on basic (echo) question stimulus.

In the basic question stimulus contour (cf. Figure 6) again, the P1_pre coincided with the end of syllable *bu-* of the noun phrase *sida busu* and P1_peak coincided with the beginning of the final vowel /u/ of this noun phrase. The P1_trail (which was also the start of the rise towards P2_peak) fell slightly earlier than in the statement contour, i.e. at the beginning of the syllable *mang-* of the verb; P2_peak fell again on the vowel of the syllable *-cing*. The terminal pitch P2_trail was again located at (or even before) the temporal midpoint of the final segment. A comparison of the pitches of the two basic stimuli is shown in Table 2.

The differences between the basic statement and question contours are very much like what was seen in the acoustical profiles exemplified in Figure 2. As far as duration is concerned, the general feature is that each syllable is longer in the statement than in the question. The differences range from 5 ms to 133 ms with a tendency of the differences to increase in the final syllables, as indicated in Table 3.

Table 3: Syllable and utterance durations (ms) of the basic statement and question stimuli.

Clause type	Syllable						Total
	si	da	bu	su	maɲ	cɪŋ	
Statement	164	114	153	161	277	532	1401
Question	112	109	134	104	217	419	1095
Δ	52	5	19	57	60	113	306

Except for experiment III, stimuli were made audible to the subjects through a Sharp Simba tape recorder equipped with a five-band equalizer so that the sound quality could be adjusted according to the subjects' wishes. The stimuli in experiment III were presented on-line from a computer such that the subjects had full flexibility in determining the timing and number of repetitions of the stimuli.

Before the subjects listened to the stimuli, they were asked to produce the utterances themselves, in the statement mode as well as in the echo question mode, with the aim that they would have a model utterance to determine whether a stimulus should be classified as a statement or as a question. Next, the subjects were asked to classify the basic stimuli to make sure that their reference was more or less correct. Only when they could correctly identify the original utterances as statement versus question was the experiment done.

In each experiment the subjects were asked to classify the stimulus they had just heard as either statement or question, and to evaluate the adequacy of the melody of that stimulus given the classification the subject had just made (typicality judgment). After that, the qualifications of the subjects were weighed by the typicality judgments (between −2 to +2). Consequently, zero (0) indicated that the clause type of the stimulus was indeterminate; a positive score index indicated a tendency to statement perception, whereas a negative index indicated a question perception.

7.3.2 Listeners

The subjects that were selected had not only to be native speakers between 20 and 55 years of age using KM daily, they also had to be educated. It is my experience that educated subjects are more accurate (and also more patient) when listening to large sets of stimuli.

The total number of subjects in the perception experiments was nineteen (eight male and eleven female), but only a few took part in all experiments. Only four of them took part in experiment III. After quantification, 787 data were collected from the subjects; these were stored in an SPSS database for statistical testing. In the database the acoustic features of each stimulus were specified as well as the perception index (between −2 and +2) attributed to it by the subjects.

7.3.3 The experiments: Variables and results

In this report, I only give a short summary of the main results of the perception experiments. For a full report see Sugiyono 2003:216 ff.

7.3.3.1 Experiment I: Effect of F0 excursion

In this experiment 324 stimulus types were presented to the listeners. One subset of 162 was generated by modifying four pivot points in the base statement contour (Figure 7), a second subset of equal magnitude introduced the same manipulations to the base question type (Figure 8). Within each subset the target values of P1_pre,

P1_peak, P2_peak and P2_trail were changed in three steps each for both minimum and maximum pitch ranges.

More specifically, the statement stimuli with maximum pitch range were created by changing one to four pivot point values of the basic statement contour (cf. Figure 5), i.e. by lowering P1_pre and P2_trail and/or increasing P1_peak and P2_peak by 5 st. In this way, twenty-seven stimuli were created. Similarly, twenty-seven stimuli were created by raising/lowering the pivot point values by 10 st, and the same number again by a 15-st change. Thus 81 maximum statement stimuli were created. The statement stimuli with minimum pitch range were based on a flat contour at 10 st. For this subset one to four pivot points were lowered/raised by 1 or 2 st. Thus 81 statement stimuli with minimum pitch range were created (cf. Figure 7). In the same way 81 stimuli with maximum pitch range and 81 with minimum pitch range were created based on the basic question contour. Altogether 4 × 81 = 324 stimuli were generated (cf. Figure 8).

Generally speaking, stimuli with a flat contour tended to be perceived as statements, be it with low typicality scores. Apparently, when an utterance has a flat contour (and no lexical question marker), it is understood as a statement even though the intonation is considered not to be good.

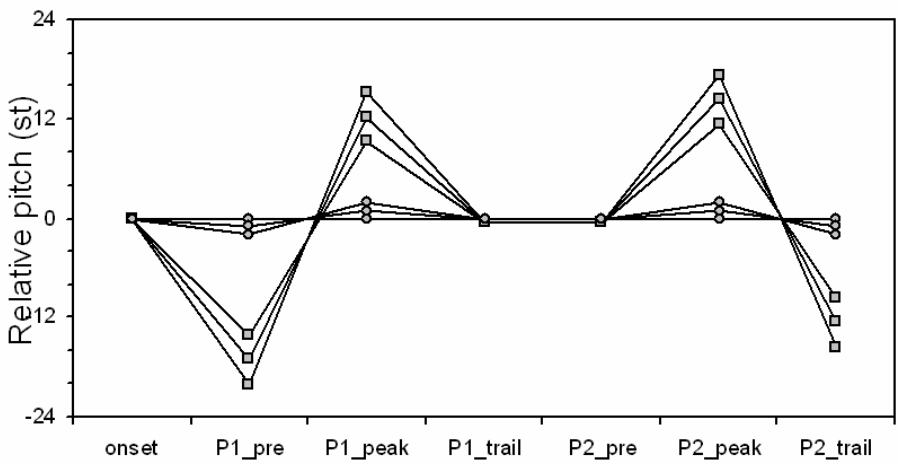

Figure 7: Statement stimulus with minimum and maximum excursions

The statement-based stimulus contour with the 3-st pitch excursion was the minimum excursion to trigger ‘perfect statement’ perception. This minimal contour had a P1_pre at the same pitch as the contour onset and P1 and P2_peaks of 1 st. It ended in a P2_trail at −2 st.

The maximum contour which triggered perfect statement perception had a 35.3-st pitch range. It had a P1_pre at −20.1 st followed by a P1_peak at 15.2 st, a P2_peak at 14.3 st and a P2_trail at −15.5 st. In both the minimum and maximum

contour the pitch peaks (P1 and P2) had to be positive, and the final pitch negative to trigger 'perfect statement' perception.

As regards the question-based stimuli this research has not been able to determine a minimal contour that triggered 'perfect question' perception; all interrogative lower threshold stimuli triggered non-perfect question responses. The minimum contour that was perceived as (non-perfect) question, was the question contour with P1_pre of −1 st followed by a P1_peak of 1 st, P2_peak of 2 st and final pitch at onset level. This contour had a pitch range of 3 st with the highest pitch on the P2_peak and the lowest on P1_pre and P2_trail. It resembled the original contour except for the height of the peaks and the final pitch, which was only as high as the onset pitch.

The maximum contour that triggered a perfect question perception had a P1_pre at −16.6 st, followed by P1_peak at 15.8 st, P2_peak at 19.4 st and P2_trail at 16.0 st. It thus had a 36-st pitch range with the lowest pitch at the P1_pre and the highest at the P2_peak.

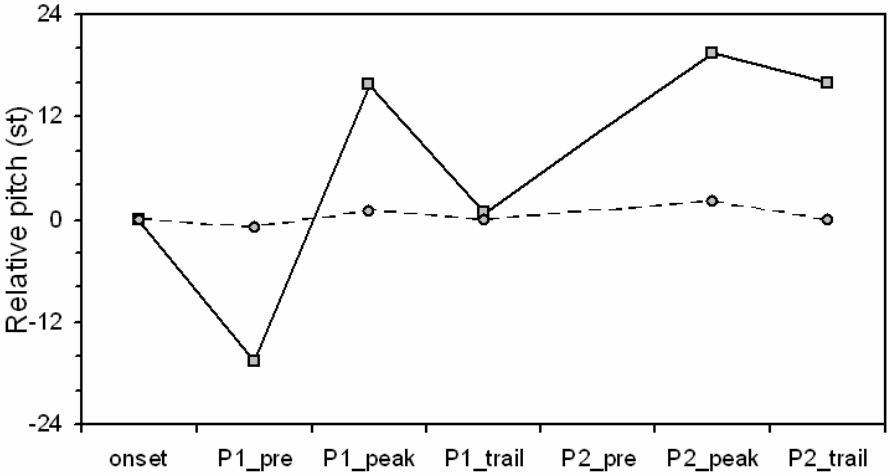

Figure 8: Question stimulus with minimum and maximum excursions. Only minimum and maximum contours indicated.

7.3.3.2 Experiment II: Durational thresholds

The durational configurations of first and second syllables of the verb *mancing* in the two types of stimuli that are able to trigger 'perfect' statement or question responses are shown in Figure 9.

The manipulation of the first syllable of the VP of both the statement and the question-based stimuli has a large influence on the subjects' perception. The difference in perception is significant when the duration of these syllables is manipulated with a change of over 10%. Here, the duration of the syllable *man* always has to be smaller than the duration the syllable *cing* in the verb *mancing*. In

the statement-based stimuli, the difference in duration of these two syllables in the temporal configuration of the stimuli changes from 90 to 450 ms, whereas in the question-based stimuli a difference of 120 ms to 416 ms is reached.

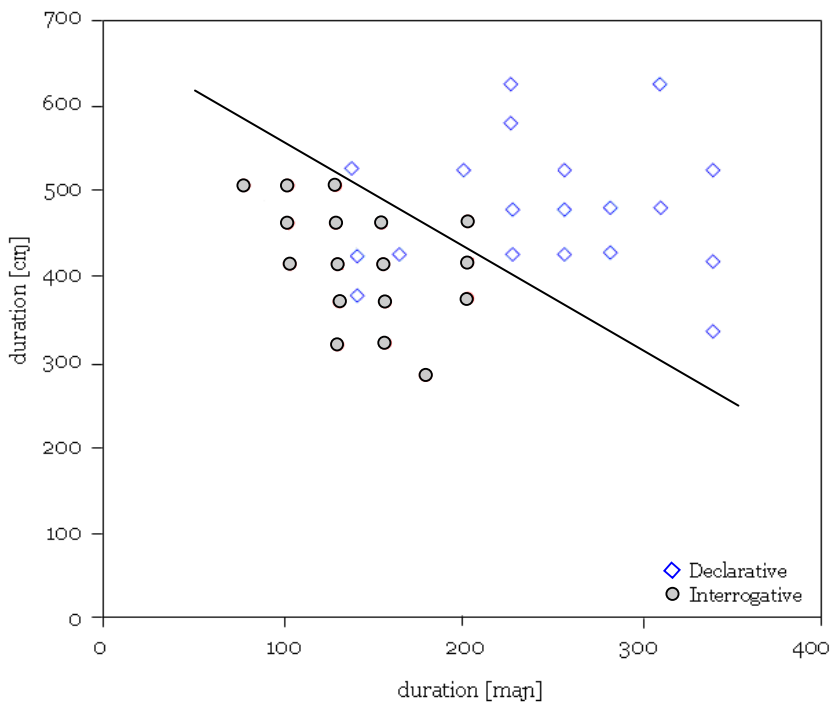

Figure 9: Syllable durations (ms) in stimuli that trigger a perfect perception as either statement (diamonds) or as question (circles). The boundary separating the statements from the question responses has been drawn by hand.

Quite clearly, then, the duration of the word *mancing* influences the choice between statement and question. Both longer duration for the first syllable *man* and for the second syllable *cing* trigger statement responses, whilst short syllables lead to interrogative percepts. The optimal boundary between statement and question responses runs at an angle, indicating that the durations of first and second syllables contribute about equally to the perception, i.e. it would seem to be total word duration that counts. When total word duration is less than 500 ms, questions are perceived, when it is longer than 700 ms statements are heard. In the uncertainty margin between 500 and 700 ms total word duration classification is less predictable.

It has been shown before that longer duration (slower speaking rate) is a correlate of the statement mode and short duration (faster speech) of the question mode (see Van Heuven & van Zanten 2005), but this experiment is the first to show that manipulating the duration of a word may effect a cross-over from statement to question in a perception study.

7.3.3.3 Experiment III: Thresholds of the statement-question contrast

In this experiment duration as well as pitch of the VP *mancing* was manipulated. Results are of limited importance, as only four subjects took part in the experiment.

Results indicate that the question contour is more sensitive to changes in the VP than the statement contour. Of the question-based stimuli around 76% was perceived as perfect statement, and 19% as non-perfect statement. Thus, virtually all (95%) question-based stimuli were perceived as statements.

In contrast, only about 8% of all the statement-based stimuli triggered perfect question perception and around 34% a non-perfect question perception. So only less than half (42%) of the statement-based stimuli were perceived as questions.

Statement-based stimuli were perceived as questions if they had a rise in pitch between the P1_trail and P2_pre, even though this difference in pitch was quite small. In addition, the duration of the syllable *man* had to be shortened by 30% (to 194 ms) and the syllable *cing* by 60% (to 213 ms). I conclude that a statement is perceived as question when its entire VP is raised in pitch and shortened in duration. The minimum pitch values to trigger question perception with this durational structure of the final constituent are visualized in Figure 10. In section 7.2.3 I found that questions are spoken more quickly than statements in KM. Again, it now seems that this finding is corroborated by perceptual evidence.

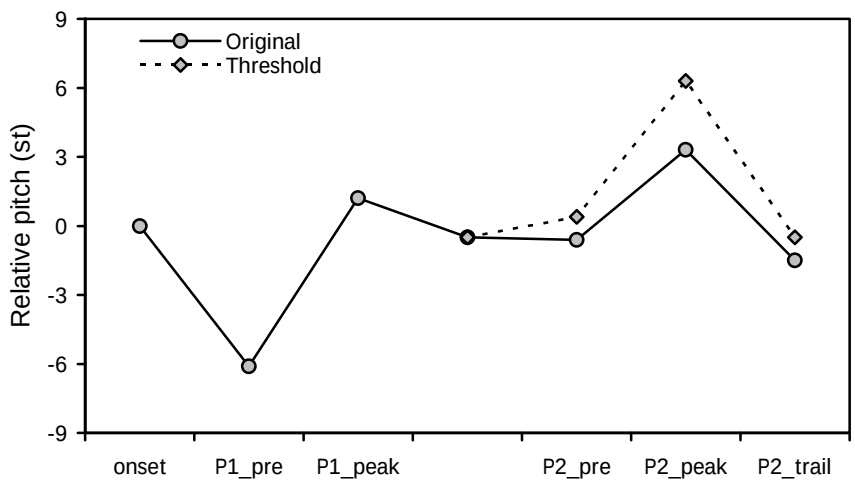

Figure10: Contrast threshold of the statement contour.

In the question-based stimuli the perception of the subjects was also very much influenced by durational changes ($p<.001$). Question-based stimuli triggered a statement perception with indexes between 1.44 and 1.87. When the duration of the syllable *man* is lengthened to 282 ms (130% of the original) and the syllable *cing* is

lengthened to 670 ms (160% of the original), question-based stimuli were perceived as perfect statements.

The minimum question-based contour to trigger a perfect-statement perception ended in a final pitch movement with P2_pre at .3 st followed by a P2_peak at 1.9 and P2_trail at –1.6 st, or a P2_peak at.9 st and P2_trail at –.6 st. When *man* and *cing* are lengthened P2_peak has to be minimally at –.2 st, with P2_trail below the onset pitch.

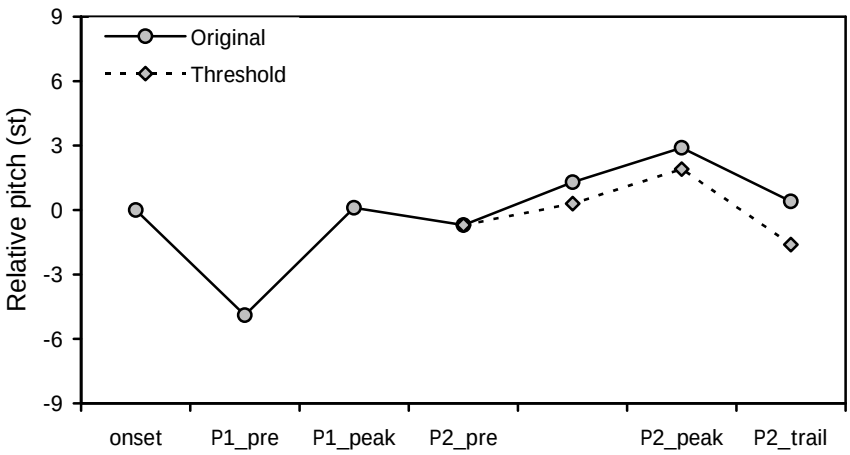

Figure 11: Contrast threshold of the question contour.

7.3.3.4 Experiment IV: Acceptability of basic sentence contours

In this experiment I tried to establish to what extent the lexical question markers (the particles *kah*, *kan*, *yo* and the question word *bila* 'when') change the mode perception. The results show, not surprisingly, that the presence or absence of lexical markers changes the perceived sentence mode significantly. Stimuli without question markers tend to trigger statement perception, whereas stimuli with question markers trigger question perception – although the stimuli were often judged to be atypical. I conclude that the question markers have a strong influence on the subjects' perception of clause type.

When question markers are added, statement-based sentences only form interrogative utterances with a low perception index (–.5 to –.8), whereas the (echo) question-based sentences form other types of interrogative utterances with a higher interrogative perception index (–.5 to –1.3).

The basic confirmative question with particle *kah* tends to be perceived as an echo question when the marker is deleted, whereas the other two types of confirmative questions, i.e. with particles *kan* and *yo*, tend to be perceived as statements when these particles are deleted. The question-based sentence with question word *bilakah* causes confusion as regards its mode when *bilakah* is

removed. Thus removing the markers does not have the same result for all the question-based utterances.

In contrast, statements generally tend to be perceived as questions if a lexical marker is added. However, this does not hold for the particle *yo*; adding this particle does not cause statements to be perceived as questions. The fact that questions are perceived as statements when *yo* is deleted, and that statements with added *yo* are still perceived as statements seems to indicate that *yo*, as opposed to *kan* and *kah*, has an affirmative meaning. More research into the specific meanings of these particles is required.

7.4 Conclusions

The results of the acoustical analysis (section 2) can be summarized as follows.

(1) MK utterances have a pitch range of almost one octave, with an average of 10.3 st. The statement pitch range (9.9 st) is about 1 st smaller than the question pitch range (10.6 st).
(2) Overall, the onset pitch of statements is about 1 st higher than the onset pitch of questions. The final pitch of statements is around 2 st lower than the final pitch of questions. This means that statements have negative excursions, whereas questions have positive excursions.
(3) The difference between onset and final pitches is not conspicuous, but the higher the onset, the larger the difference between onset and final pitches.
(4) The P1_peaks of statements and questions are of approximately the same height, but statements have a much lower P1_pre than questions. Consequently, statements have a larger rise (5.5 st) from P1_pre to P1_peak than questions (3.1 st) do.
(5) The final pitch movement also marks the statement-question contrast. The final pitch movement is Level-Rise-Fall for statements and Rise-Fall for questions. Beginning and mid positions usually contain Fall-Rise or Level-Rise movements with varying degrees of slope.
(6) Statements are approximately 10% longer than questions. The duration of simple SV statements is around 2.2 s and that of the questions around 2.0 s.
(7) Echo questions and statements start at the same pitch. After that the echo question contour is significantly higher than the statement contour. This difference increases towards the end of the utterance.

Preliminary results of the perception experiments are as follows.

(1) Questions are more sensitive to changes in the acoustic features than statements. Ninety-five percent of question-based stimuli were perceived as statements and 42% of statement stimuli were perceived as questions.
(2) Flat contours – with 0 st pitch range – tend to be perceived as statements, be it with imperfect intonation. To be rated as perfect, statements had to have a pitch range of between 3 and 3.5 st.

(3) The finding in the production experiment that statements are longer than questions is corroborated: statements were perceived as questions when their final constituent was (slightly) raised and (considerably) shortened.
(4) Not surprisingly, statements are generally perceived as questions if a lexical marker is added; but statements with *yo* are still perceived as statements. Question stimuli without lexical question markers are perceived as 'imperfect' statements. However, when *kah* is deleted in confirmative questions these tend to be perceived as echo questions.

References

Boersma, P. & D. Weenink (1996) *PRAAT : A system for doing phonetics by computer*. Report of the Institute of Phonetic Sciences of the University of Amsterdam 132. [http://www.praat.org]

Collins, J.T. (1992) Bahasa Melayu di sungai Mahakam siasatan fonologi Dialek Kutai [The Malay language along the Mahakam river: A study on the phonology of the Kutai dialect]. *Sari* 10, 3−24.

Haan, J. (2001) Speaking of questions: An exploration of Dutch question intonation. LOT Dissertation Series 52. Utrecht: LOT.

Heuven, V.J. van & J. Haan (2000) Phonetic correlates of statement versus question intonation in Dutch. In A. Botinis (ed) *Intonation: Analysis, modelling and technology*. Amsterdam: Kluwer, 119−143.

Heuven, V.J. van & E. van Zanten (2005) Speech rate as a secondary prosodic characteristic of polarity questions in three languages. *Speech Communication* 47, 87−99.

Lehiste, I. (1970) *Suprasegmentals*. Cambridge: The MIT Press.

Mursalim, M. & Gazali (1995) *Geografi Dialek Bahasa Kutai di Kabupaten Kutai [Geography of the Kutai dialect spoken in Kutai regency]*. Samarinda: Universitas Mulawarman.

Pernyata, S. (1992) *Hubungan kekerabatan bahasa-bahasa daerah non-Dayak di Kalimantan Timur [Kinship relations between non-Dayak regional languages in Eastern Kalimantan]*. Samarinda: Universitas Mulawarman.

Sluijter, A.M.C. & J.M.B. Terken (1993) Beyond sentence prosody: Paragraph intonation in Dutch. *Phonetica* 50, 180−188.

Small, A.M. (1973) Psychoacoustics. In F.D. Minifie, T.J. Hixon & F. Williams (eds) *Normal aspects of speech, hearing, and language*. New Jersey: Prentice Hall.

Sugiyono (2003) *Kajian psikoakustik terhadap frekuensi fundamental pada kontras deklarativitas dan interogativitas [Psychoacoustic background of the role of fundamental frequency in the contrast between statement and question intonation]*. PhD dissertation, Universitas Indonesia.

Suryadikara, F., Sjahrial S.A.R. Ibrahim & M.P. Lambut (1979) *Bahasa Kutai [The Kutai language]*. Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.

Thorsen, N. (1978) An acoustical investigation of Danish intonation. *Journal of Phonetics* 6, 151−175.

Wurm, S.A. & S. Hattori (eds) (1981) *Language atlas of the Pacific area*. Pacific Linguistics C-66, C-77.

Appendix

The 48 target sentences are printed in **bold**

A-broad. Answer-Statements; focused words are **CAPITALIZED**
Focus on entire VP:

1. Q: Napa garang sida busu tu?
Doing what PAR HON uncle that
'What is Uncle doing?'
A: Sida busu tu kah? **Sida busu MANCING**.
HON uncle that PAR HON uncle fishing
'Uncle? Uncle is fishing'

2. Q: Napa garang sida busu tu?
Doing what PAR HON uncle that
'What is Uncle doing?'
A: Sida busu tu kah? **Sida busu MANCING JUKUT PATIN**.
HON uncle that PAR HON uncle fishing catfish
'Uncle? Uncle is fishing catfish'

3. Q Napa garang sida busu wayah ni yo?
Doing what PAR HON uncle now PAR
'What is Uncle doing?'
A: Sida busu Onoi kah? **Sida busu MANCING DI MAHAKAM**.
HON uncle NAME PAR HON uncle fishing on Mahakam.
'Uncle Onoi? Uncle is fishing on the Mahakam (river)'

4. Q: Napa garang sida busu wayah ni yo?
Doing what PAR HON uncle moment this PAR
'What is Uncle doing at the moment?'
A: Sida busu Onoi kah? **Sida busu MANCING JUKUT PATIN DIMAHAKAM**.
HON uncle NAMA PAR HON uncle fishing catfish in Mahakam
'Uncle Onoi? Uncle is fishing on the Mahakam (river)'

5. Q: Apa dipolah sida busu tu?
what is done HON uncle that
'What is Uncle doing?'
A: Nya dipolah sida busu tu kah? **Sida busu MANCING**.
what is done HON uncle itu PAR HON uncle fishing
'What Uncle is doing? Uncle is fishing.

6. Q: Apa dipolah sida busu tu?
what is done HON uncle that
'What is Uncle doing?'
A: Nya dipolah sida busu tu kah? **Sida busu MANCING JUKUT PATIN**.
what is done HON uncle that PAR HON uncle fishing catfish
'What Uncle is doing? Uncle is fishing catfish.

A-narrow. Focus on a smaller constituent:

7. Q: Apa dipolah sida busu di Mahakam tu?
what is done HON uncle on Makakam that
'What is Uncle doing on the Mahakam?'

A: Nya dipolah sida busu di Mahakam? **Sida busu MANCING di Mahakam**.
What is done HON uncle on Mahakam HON uncle fishing on Mahakam
'What Uncle is doing on the Mahakam? Uncle is fishing on the Mahakam.'

8. Q: Apa dipolah sida busu di Mahakam ngan jukut patin tu?
what is done HON uncle on Mahakam with catfish that
'What is Uncle doing on the Mahakam with the catfish?'
A: Nya dipolah sida busu di Mahakam ngan jukut patin tu?
what is done HON uncle on Mahakam with catfish that
Sida busu MANCING jukut patin di Mahakam.
HON uncle fishing catfish on Mahakam
'What Uncle is doing on the Mahakam with the catfish? Uncle is fishing catfish on the Mahakam.'

9. Q: Apa dipancing sida busu tu?
what is fished HON uncle that
'What is Uncle fishing?'
A: Nya dipancing sida busu tu kah? **Sida busu mancing JUKUT PATIN.**
what is fished HON uncle that PAR HON uncle fishing catfish
'What Uncle is fishing? Uncle is fishing catfish.'

10. Q: Apa dipancing sida busu di Mahakam tu?
what is fished HON uncle on Mahakam that
'What is Uncle fishing on the Mahakam?'
A: Nya dipancing sida busu di Mahakam? **Sida busu mincing JUKUT PATIN di Mahakam.**
What is fished HON uncle on Mahakam HON uncle fishing catfish on Mahakam
'What Uncle is fishing on the Mahakam? Uncle is fishing catfish on the Mahakam.'

11. Q: Dimana garang sida busu mancing tu?
where PAR HON uncle fishing that
'Where is Uncle fishing?'
A: Odah sida busu mancing kah? **Sida busu mancing DI MAHAKAM.**
place HON uncle fishing PAR HON uncle fishing on Mahakam
'Where Uncle is fishing? Uncle is fishing on the Mahakam.'

12. Q: Dimana garang sida busu mancing jukut patin tu?
where PAR HON uncle fishing catfish that
'Where is Uncle fishing the catfish?'
A: Odah sida busu mancing jukut patin kah? **Sida busu mincing jukut patin DI MAHAKAM.**
place HON uncle fishing catfish PAR HON uncle fishing catfish on Mahakam
'Where Uncle is fishing catfish? Uncle is fishing catfish on the Mahakam.'

B. Confirmation statements (4) and declarative questions (4)

1. Q: Apa garang? **Sida busu mancing?**
what PAR HON uncle fishing
'What? Uncle is fishing?'
A: Ya leh. **Sida busu mancing.**
yes PAR HON uncle fishing
'Yes. Uncle is fishing.'

2. Q: Apa garang? **Sida busu mancing jukut patin?**
what PAR HON uncle fishing catfish
'What? Uncle is fishing catfish?'
A: Ya leh. **Sida busu mancing jukut patin.**
yes PAR HON uncle fishing catfish
'Yes. Uncle is fishing catfish.'

3. Q: Apa garang? **Sida busu mancing di Mahakam?**
what PAR HON uncle fishing on Mahakam
'What? Uncle is fishing on the Mahakam?'
A: Ya leh. **Sida busu mancing di Mahakam.**
yes PAR HON uncle fishing on Mahakam
'Yes. Uncle is fishing on the Mahakam.'

4. Q: Apa garang? **Sida busu mancing jukut patin di Mahakam?**
what PAR HON uncle fishing catfish on Mahakam
'What? Uncle is fishing catfish on the Mahakam?'
A: Ya leh. **Sida busu mancing jukut patin di Mahakam.**
yes PAR HON uncle fishing catfish on Mahakam
'Yes. Uncle is fishing catfish on the Mahakam.'

C. Contrastive statements

1. Q: Sida busu begubangan maha?
HONuncle boating only
'Is Uncle just boating?'
A: Endik leh. **Sida busu MANCING.**
no PAR HON uncle fishing
'No. Uncle is fishing.'

2. Q: Sida busu mancing jukut jelawat?
HON uncle fishing fish jelawat
'Is Uncle fishing jelawat fish?'
A: Endik leh. **Sida busu mancing JUKUT PATIN.**
no PAR HON uncle fishing catfish
'No. Uncle is fishing catfish.'

3. Q: Sida busu njala jukut patin?
HON uncle fishing with a net catfish
'Is Uncle fishing catfish with a net?'
A: Endik leh. **Sida busu MANCING jukut patin.**
no PAR HON uncle angling catfish
'No. Uncle is angling catfish.'

4. Q: Sida busu njala di Mahakam?
HON uncle fishing with a net on Mahakam
'Is Uncle fishing with a net on the Mahakam?'
A: Endik leh. **Sida busu MANCING di Mahakam.**
no PAR HON uncle angling on Mahakam
'No. Uncle is angling on the Mahakam.'

5. Q: Sida busu mancing di Semayang?
HON uncle fishing on Semayang
'Is Uncle fishing on the Semayang?'
A: Endik leh. **Sida busu mancing DI MAHAKAM.**
no PAR HON uncle fishing on Mahakam
'No. Uncle is fishing on the Mahakam.'

6. Q: Sida busu mancing jukut patin di Semayang?
HON uncle fishing catfish on Semayang
'Is Uncle fishing catfish on the Semayang?'
A: Endik leh. **Sida busu mancing jukut patin DI MAHAKAM.**
no PAR HON uncle fishing catfish on Mahakam
'No. Uncle is fishing catfish on the Mahakam.'

7. Q: Sida busu njala jukut patin di Mahakam?
HON uncle fishing with a net catfish on Mahakam
'Uncle is fishing catfish with a net on the Mahakam?'
A: Endik leh. **Sida busu MANCING jukut patin di Mahakam.**
no PAR HON uncle angling fish patin on Mahakam
'No. Uncle is angling catfish on the Mahakam.'

8. Q: Sida busu mancing jukut jelawat di Mahakam?
HON uncle fishing fish jelawat on Mahakam
'Is Uncle fishing jelawat fish on the Mahakam?'
A: Endik leh. **Sida busu mancing JUKUT PATIN di Mahakam.**
no PAR HON uncle fishing catfish on Mahakam
'No. Uncle is fishing catfish on the Mahakam.'

D. Declarative questions (no lexical marking)

1. Apa mbok? **Sida busu mancing?** Ah … ndik percaya.
what aunt HON uncle fishing INTJ not believe
'What is it, Aunt? Uncle is fishing? Ah… I don't believe so.'

2. Apa mbok? **Sida busu mancing jukut patin?** Ah … ndik percaya.
what aunt HON uncle fishing catfish INTJ not believe
'What is it, Aunt? Uncle is fishing catfish? Ah… I don't believe so.'

3. Apa mbok? **Sida busu mancing di Mahakam?** Ah … ndik percaya.
what aunt HON uncle fishing on Mahakam INTJ not believe
'What is it, Aunt? Uncle is fishing on the Mahakam? Ah… I don't believe so.'

4. Apa mbok? **Sida busu mancing jukut patin di Mahakam?** Ah … ndik percaya.
what aunt HON uncle fishing fish patin on Mahakam INTJ not believe
'What is it, Aunt? Uncle is fishing catfish on the Mahakam? Ah… I don't believe so.'

E. Tag questions; interrogative particle *kah*

1. Mbok-mbok… **Sida busu mancing kah?**
Aunt (RED) HON uncle fishing PAR
'Aunt…. Is Uncle fishing?'

2. Mbok-mbok… **Sida busu mancing jukut patin kah?**
Aunt (RED) HON uncle fishing catfish PAR
'Aunt…. Is Uncle fishing catfish?

3. Mbok-mbok… **Sida** **busu** **mancing** **di** **Mahakam** **kah?**
Aunt (RED) HON uncle fishing on Mahakam PAR
'Aunt…. Is Uncle fishing on the Mahakam?

4. Mbok-mbok… **Sida** **busu** **mancing** **jukut patin** **di Mahakam** **kah?**
Aunt (RED) HON uncle fishing catfish on Mahakam PAR
'Aunt…. Is Uncle fishing catfish on the Mahakam?

F. Tag questions; interrogative particle *kan*

1. Mbok-mbok… **Sida** **busu** **mancing** **kan?**
Aunt (RED) HON uncle fishing PAR
'Aunt…. Uncle is fishing, isn't he?'

2. Mbok-mbok… **Sida** **busu** **mancing** **jukut patin** **kan?**
Aunt (RED) HON uncle fishing catfish PAR
'Aunt…. Uncle is fishing catfish, isn't he?'

3. Mbok-mbok… **Sida** **busu** **mancing** **di** **Mahakam** **kan?**
Aunt (RED) HON uncle is fishing on Mahakam PAR
'Aunt…. Uncle is fishing on the Mahakam, isn't he?'

4. Mbok-mbok… **Sida** **busu** **mancing** **jukut patin** **di Mahakam** **kan?**
Aunt (RED) HON uncle fishing catfish on Mahakam PAR
'Aunt…. Uncle is fishing catfish on the Mahakam, isn't he?'

G. Tag questions; interrogative particle *yo*

1. O… **Sida** **busu** **mancing** **yo?**
INTJ HON uncle is fishing PAR
'0… So Uncle is fishing?'

2. O… **Sida** **busu** **mancing** **jukut patin** **yo?**
INTJ HON uncle fishing catfish PAR
'O… So Uncle is fishing catfish?'

3. O… **Sida** **busu** **mancing** **di** **Mahakam** **yo?**
INTJ HON uncle fishing on Mahakam PAR
'O… So Uncle is fishing on the Mahakam?'

4. O… **Sida** **busu** **mancing** **jukut patin** **di Mahakam** **yo?**
INTJ HON uncle fishing catfish on Mahakam PAR
'O… So Uncle is fishing catfish on the Mahakam?'

H. Wh-questions; question word *bila* 'when'

1. Mbok-mbok. **Bila** **kah** **sida** **busu** **mancing?**
Aunt(RED) when PAR HON uncle fishing
'Aunt. When did Uncle go fishing?'

2. Mbok-mbok. **Bila** **kah** **sida** **busu** **mancing** **jukut patin?**
Aunt (RED) when PAR HON uncle fishing catfish
'Aunt. When did Uncle go fishing catfish?'

3. Mbok-mbok. **Bila** **kah** **sida** **busu** **mancing** **di** **Mahakam?**
Aunt (RED) when PAR HON uncle fishing on Mahakam
'Aunt. When did Uncle go fishing on the Mahakam?'

4. Mbok-mbok. **Bila** **kah** **sida** **busu** **mancing** **jukut patin** **di Mahakam?**
Aunt (RED) when PAR HON uncle fishing catfish on Mahakam
'Aunt. When did Uncle go fishing catfish on the Mahakam?'

Abbreviations:

HON	honorific
INTJ	interjection
NAMA	name
PAR	particle
PREP	preposition
RED	reduplication