Detecting Anticipation Foci for Simultaneous Interpreting

Eva Kiktová, LICOLAB, Pavol Jozef Šafárik University in Košice
Július Zimmermann, LICOLAB, Pavol Jozef Šafárik University in Košice
Mária Paľová, LICOLAB, Pavol Jozef Šafárik University in Košice
Rudolph Sock, University of Strasbourg/LICOLAB, Pavol Jozef Šafárik University in Košice

The main thrust of this investigation is to unveil prosodic anticipatory cues in political speeches, for French. The general hypothesis is that, sensitised to these anticipatory prosodic patterns, the more or less beginner interpreter, interpreting from French to Slovak, would take advantage of these anticipatory prosodic patterns in French, and should hence be capable of integrating them in making strategic interpretation decisions. Experimental investigations concerned perceptual evaluation tests, acoustic analyses and speech recognition techniques. The expert interpreters who served as subjects in the investigation confirmed that a rising melody on the last syllable of a rhythmic group, together with the subsequent pause were useful phenomena in signalling crucial upcoming semantic information, contained in the following rhythmic group. Hence, the last syllable of the rhythmic group, with a rising melody and the subsequent pause were considered as the anticipation focus. Results obtained in this study also indicate that HMM modelling, in combination with the MFCC pitch affected features, can be effectively applied to the anticipation issue. On a long-term basis, results from this investigation should contribute to elaborating a conceptual model, which would combine the phonetic, syntactic and semantic levels, and serve as a "mental software" for optimal interpreting.

Keywords: *simultaneous interpreting, anticipation, nucleus, focus, prosody, Hidden Markov Model*

1. Aim and background

The current study aims at unveiling and analysing prosodic anticipatory cues for French. The general hypothesis is that, sensitised to these anticipatory prosodic patterns, the more or less beginner interpreter, interpreting from French to Slovak, would take advantage of these anticipatory prosodic patterns in French and should hence be capable of integrating them in making strategic interpretation decisions.

Experimental investigations will rely on perceptual evaluation tests, acoustic analyses and speech recognition techniques. On a long-term basis, results from this investigation should contribute to elaborating a conceptual multilevel model, which would combine phonetic, syntactic, semantic and pragmatic components, and serve as a "mental software" for optimal simultaneous interpreting.

Anticipation is a basic concept for our study. Reviewing the literature tells us that quite a good number of interpreter researchers approve its importance in simultaneous interpreting (see, e.g. Moser 1976; Kirchhoff 1976; Wilss 1978; Lederer 1981; Seleskovitch 1984; Adamowicz 1989; Van Dam 1989; Chernov 1992; Gile 1992; Kohn & Kalina 1996; Riccardi & Snelling 1997; Massaro & Shlesinger 1997; Zanetti 1999; Setton 1999).

However, there is no consensus as to whether anticipation plays a more significant role when interpreting between structurally different languages than when interpreting between structurally similar ones. This issue is not central to our investigation here, which focuses on intonational features rather than on cross-language differences in syntactic structures.

However, the matter will need to be addressed in our future investigations (for now, see (Seeber 2001) for a detailed overview devoted to the question). In short, the disagreement concerns two main "schools" of the interpreting research community.

Seeber (2001) reports that one school, called "the universalists" by Setton (1999) or "liberal arts community" by Moser-Mercer (1994), carry out research within the conceptual framework of interpretive theory (IT) or theorie du sens, and is mainly advocated by Lederer (1981) and Seleskovitch (1968, 1984). Proponents of this school posit that given sufficient linguistic skill in both languages involved, simultaneous interpretation would be just as problematic for any language pair. Indeed, the predictive nature of comprehension would annul structural asymmetries (Setton 1999). As concerns anticipation, the claim is that strategies would be deployed across all sentence components, hence there would be no language-specific factor relating to the verb.

According to Seeber (2001) still, the other school called "the bilateralists" (Setton 1999) or the "natural science community" (Moser-Mercer 1994), theorises within the information processing (IP) paradigm (see, e.g. Moser 1976, 1978; Gile 1992, 1995). Relying on knowledge in cognitive psychology and in the neurosciences, they posit that the supposedly sequential nature of language perception, and therefore of its comprehension would have direct impact on simultaneous interpretation, as the conversion processes involved in simultaneous interpretation is believed to be governed by linguistic structure (Moser-Mercer et al., 2000).

Although the debate on the role that anticipation plays during simultaneous interpreting has been extensive and rich, thorough experimental research and corpus-based accounts of anticipation are quite rare in the literature.

Let us cite one investigation carried out by Wilss (1978), based on data collected by Mattern (1974, cited by Seeber 2001). It indeed represents a pioneering work towards an empirical investigation of the anticipation issue in simultaneous interpretation. In his paper, Wilss (1978) posits that the "transfer on the basis of parallel syntactic structures can – at least on the syntactic level of the interlingual transfer – be regarded as easier to accomplish" (Wilss 1978: 343) and that "[s]yntactic divergences between SL [source language] and TL [target language] have clearly different implications for translation and for SI procedures" (Wilss 1978: 345). Such a statement is clearly in line with the advocates of the IP theory.

Zanetti (1999) hypothesises that anticipation may stem from compression of the time required for decoding, taking into account the fact that any instance of linguistic processing, the cognitive task ultimately consists of a decoding phase followed by an encoding one. Results of her investigation indeed demonstrated that the so-called "temporal anticipation" (Lederer 1981), refers mainly to the macroscopic aspect of a deeper phenomenon, which is ultimately generated by a compression in the time required for decoding. She further argues that compression is made possible by the continuous information flow between bottom-up (textual) and top-down information from long-term memory. This flow is purportedly continuous and simultaneous at all levels of analysis, from the acoustic-phonological level (Massaro 1975, 1978; Marslen-Wilson 1989) to the syntactic level (Flores d'Arcais 1991) and to the semantic-pragmatic level (Kintsch & van Dijk 1983).

Coming closer to the issue addressed in this paper, it turns out that the review of the literature indicates that very few authors have examined the role of prosody, or of intonation, in simultaneous interpreting.

Rare authors like Kohn & Kalina (1996) claim that for strategic anticipation, the interpreter exploits all available cues, from pragmatic inferences to lexical collocations, syntactic structures and prosodic features.

Riccardi (1997) postulates that prosodic features may help the interpreter to comprehend parenthetical sentences, even when they are very long, and also awkward constructions, and that anticipation can be triggered by both linguistic and non-linguistic factors.

Gile (1992) found a significant difference in PSE (predictable sentence endings) between Japanese or German, and French or English. The author claims that such a finding could be rationalised mainly by one single factor: the fact that in Japanese and German, determinant elements precede the main verb, which is in final position. Hence, as soon as a function word or tonal stress signals the exhaustion of such determinant elements, the listener knows that what follows is sentence termination and can thus use this information to anticipate the upcoming sentence ending.

Falbo (1999) has shown that conference interpreting is a particular form of orality, pinpointing the importance of prosody in this cognitive task and the need to carry out interdisciplinary research investigations in the area.

In some instances, professional interpreters do come up with anomalous intonation and stress patterns (Shlesinger 1994; Ahrens 2004, 2005) because of the cognitive load they encounter during simultaneous interpretation. Such deviant prosodic structures in the speaker's production may affect comprehension of the source text. However, research in the area is very little explored.

Martellini (2013) carried out an investigation in which she analysed the prosody of professional interpreters with the help of a perceptual method, assessing prosodic features as observed in interpreting practitioners. The results concerning interpreters' speech rate and intonation confirmed well-known theories in simultaneous interpretation. However, categories such as pauses, stress on words and the sub-category of syllable lengthening brought up novel issues, indicating that specific behaviour was intentionally produced by interpreters to cope with difficult portions of the text by making use of prosodic features.

Seeber (2001) attempted to probe further into the relationship between anticipation and prosody, by empirically investigating the role of source language prosody in simultaneous interpreting.

Wanting to go further than intuitive claims and to uncover evidence validating or challenging the purported role of prosody for the simultaneous interpreting process, Seeber (2001) designed an experiment aimed at exploring the effects of monotonous intonation on anticipation in simultaneous interpreting.

Experimental results did not support the author's initial hypothesis according to which monotonous intonation of the source text would have an inhibiting effect on the interpreter's capacity to anticipate the verb when interpreting simultaneously from German into English. Rather, subjects anticipated the verb more accurately and more rapidly during the interpretation of the monotonous speech than during the normal speech. Seeber (2001) thus concludes that interpreters tend to compensate for absence of intonation by optimally increasing their cognitive effort, and by adopting a more conservative interpreting strategy.

Anticipation in simultaneous interpreting also relies on pauses. Some simultaneous interpreting researchers say that pauses and non-fluencies could be exploited in a strategic way (Čeňková 1989; Gile 1995). They would not only serve the interpreter to monitor her/his own anticipations, but that the interpreter could also produce filled micro-pauses in order to slow down her/his delivery and hence focus on listening to the input. What Setton (1999) refers to as waiting, is usually considered as another strategy to cope with complex or temporally undetermined syntactic structures. The manoeuvre consists in inserting short pauses at

grammatical boundaries, in order to gain time without giving the listener the impression of supressing portions of the original message. Such strategies would sometimes help the interpreter to make easier semantic and syntactic choices. However, these conjectures still need to be verified experimentally in a situational and functional perspective.

The standing of our present study in the area of simultaneous interpreting and its originality lie in the fact that investigations carried out, for French as the source language, go beyond assumptions to empiricism, backed by processed quantitative speech data, laboratory experiments and results. Our research is thus in line with Ghelly Vassilievich Chernov's seminal works in interpreting studies (1929–2000), where he puts emphasis on the process of probability anticipation or forward inferencing in simultaneous interpreting, drawing on real data. For him, simultaneous interpreting is a complex goal-oriented cognitive activity, as postulated in Anokhin's general model of purposeful activity (Anokhin 1968). Simultaneous interpreting is thought to be dependent on perceptual attunement to remarkable changes in the external and internal environment (see, e.g. Anokhin 1968, 1978), cited in Chernov 2004). Since it is carried out in extreme cognitive conditions, it decisively relies on a minimum level of redundancy in the input, together with some additional knowledge, to enable anticipation and thus to facilitate an adequate on-line synthesis of the verbal output (Chernov 1994). According to Chernov's model (Chernov 2004: 169–171) of probability anticipation as a multilevel mechanism, semantic redundancy in the discourse arises on four major tiers that can be summarised as follows: 1) The prosodic tier, with aspects of phonotactics, intonation and rhythm. 2) The syntactic tier, which concerns units such as phrase and utterance. 3) The semantic tier, considered as the most important tier of the mechanism, encompasses the levels of the phrase or syntagm, the utterance, and the discourse. 4) The inferential tier, which deals with the entire situational context, where all levels of the anticipation mechanism converge.

Although the prosodic tier may be considered as being subsidiary, we posit here that prosodic anticipatory cues make up a crucial part of the redundant and critical components in the input. Within the abovementioned theoretical framework of attentional and perceptual attunement to significant changes to information from the environment, prosodic features accompanying syntactic and semantic structures, such as the 'theme' and the 'rheme', correspond respectively to idiosyncratic cyclical features of semantic backgrounding and foregrounding in the input discourse. Whereas the theme, accompanied by its specific prosodic pattern, may serve as redundant yet valuable semantic features in the discourse, the interpreter would focus attention on the critical rheme and on its particular prosodic contour. As we shall see later, it is, however, the far most right edge of the intonational contour of the theme and the inserted pause that somehow pave the way for semantic foregrounding in the upcoming rheme.

Anticipatory phonetic phenomena are consistently present in speech production. The extent of these anticipatory speech strategies goes from the micro-level, i.e. phonetic segments (vowels, consonants and syllables) to the macro-level, which concerns dissyllabic and multisyllabic words, syntagms, utterances or sentences, and an entire discourse. Thus, in speech production, anticipatory articulatory gestures and the associated acoustic features are understood, at the micro-level, as the expansion or the extension of certain gestures and acoustic features to adjacent or neighbouring phonetic segments (Perkell & Chiang 1986). At the macro-level, the extension of these anticipatory phenomena goes beyond segments and syllables; they are usually catered for by prosodic features, especially intonational patterns, and their domain is at least the word, but usually long stretches of sentences. In speech perception, listeners have learned to exploit these anticipatory phenomena inherent to the speakers' productions in order to adequately and optimally parse the linguistic message (Lubker &

Lindgren 1982; Vaxelaire et al. 2003; Sock & Vaxelaire 2004). In other words, listeners make use of precocious articulatory, acoustic and prosodic cues, linked to these anticipatory elements in the speech chain, to enhance the intelligibility of the on-going message.

Within this framework, it has been shown for French (Kleiber & Sock 2006) that certain demonstrative adjectives in relatively long sentences do reveal anticipatory phenomena, and are thus referred to as cataphoras. In such cases, the referential process is reflected in the prosodic contour, which facilitates online construction of the linguistic message. Indeed, the gradual anticipatory elevation of mainly the intonational pattern (F0 or fundamental frequency), together with a following pause, reflect an expectative phenomenon, suspense, recuperation of elements from a long-term memory. Thus, this anticipatory intonational contour somehow represents, iconically, the output referential process of immediate memory in order to recover the referent in long-term memory.

2. Defining notions and presenting a recognition model and related tools

As the material analysed in the current work is based on speeches of the European Parliament members delivered in French, anticipatory phenomena studied here will take into account the central notion of rhythmic groups in this language. In French, a rhythmic group can be defined as a cohesive group or chunk of words, which, to some extent, makes sense and may correspond to a syntagm or to an entire utterance. Specifying rhythmic groups may vary from one speaker or listener to the other. However, decisions as to consensually identifying rhythmic groups in French are usually consistent and robust among native French or Francophone speakers.

Here are two simple examples to help illustrate this notion of rhythmic groups:

(1) *Le Président de la République slovaque est arrivé à Paris.*
The President of the Slovak Republic has arrived in Paris (a sentence made up of one rhythmic group).

(2) *Lorsque le Président de la République slovaque est arrivé à Paris // il a été reçu par son homologue français*.
When the President of the Slovak Republic arrived in Paris // he was met by his counterpart (a sentence made up of two rhythmic groups as indicated by the "//" separating sign).

Since stress or pitch accent, in French, is usually on the last syllable of individual words, of rhythmic groups, of syntagms, of utterances or of long stretches of utterances, this accent is frequently accompanied by a pause (Delattre 1966; Wenk & Wioland 1982; Bailly 1989).

Thus in (2), a stress or pitch accent will fall firstly on the last syllable [Ri] of the disyllabic word "Paris", followed very likely by a pause, and secondly on the last syllable [se] of the disyllabic word "français". These two prosodic features – final stress and the subsequent pause – indeed serve as syntactic and semantic factors in structuring utterances. Hence, final stress, together with the pause, largely determine word assembling within rhythmic groups in utterances and may even contribute to parsing along with the rhythm and melody pattern (Chernov 2004). Let us cite the following simple and well-known examples of a holorhyme here, where two (or more) sentences are segmentally similar but differ in meaning due to differences in prosodic structuring:

(3) *La belle ferme le voile.*
The beautiful farm is obscuring him (with pitch/stress accent on the word “voile”).
(4) *La belle // ferme le voile.*
The beautiful [woman] shuts the veil (with pitch/stress accent on the words “belle” and “ferme”).

Sentence (1) comprises a single rhythmic group, whereas sentence (2) is composed of two rhythmic groups.

In longer stretches of sentences or in a discourse, there is usually a rising intonational pattern at the end of the first segment of, say a bi-segmental utterance, where the initial segment carries a relevant but low communicative load. Therefore, in (2) and (4), it is always the second segment or rhythmic group, which is predominant in foregrounding the linguistic message. However, it is the last syllable of the first rhythmic group ([Ri] in (2) and [fERm] in (4)), which carries the pitch accent, with a rising intonational pattern, signalling continuity, and followed by a pause, both serving to introduce the expectative important linguistic phenomenon or suspense contained in the subsequent rhythmic group. We refer here to this last syllable of the first rhythmic group, accompanied by a pause as the focus in a given syntactic, semantic and pragmatic context. This focus corresponds to a central point, as of attraction, of attention, or of activity that serves to trigger anticipatory attention in the listener or the interpreter, in our case. In fact, preliminary examination of our experimental data has shown that anticipatory foci may not only concern last syllables of the first rhythmic groups, but may also spread backwards to antepenultimate syllables of those rhythmic groups, as fundamental frequency, F0, tends to rise remarkably within these last two syllables. We shall see later that the term sonantic nuclei will be retained to refer to these anticipatory foci, when it serves to better illustrate how absolute and relative measures were obtained for fundamental and intensity signal variations.

At this point, we would like to make it clear that there is no a priori evidence from previous studies that anticipatory supra-segmental or prosodic information is treated in a different way than segmental information as regards the architecture of the grammar of a specific language (Beaver et al. 2007). It is the combined effect of the purportedly two levels, which contributes to the emergence of meaning (Lederer 1978).

Nonetheless, the predominance of prosodic means in foregrounding semantic components within a given utterance or discourse is undeniable, as they are the most robust features that are always present, even with standard or neutral word order in an utterance. In an utterance with standard or neutral word order, the position of a word or of specific words in the utterance, without prosodic focus, does not contribute significantly to the semantic content of the sentence. We shall take up this issue later.

For now, let us briefly present the recognition model and related tools that will serve in the automatic detection of anticipation nuclei. They are mainly the Hidden Markov Model (HMM) and Mel-Frequency Cepstral Coefficients (MFCC).

A Hidden Markov Model (HMM) based classifier is a popular approach in different recognition tasks (Huang et al. 1990; Gales & Young 2008; Vozáriková et al. 2011; Bhardwaj et al. 2015). Tools from a Hidden Markov model ToolKit (HTK) (Young et al. 2002) are used to build an HMM based classifier system. In the training or learning process, the estimation of HMM parameters is carried out using training samples and their corresponding transcriptions. In the testing process, unknown samples are recognised using a Viterbi based decoding algorithm (Young et al. 2002; Bhardwaj et al. 2015). The goal of the acoustic processing is to provide an appropriate method in determining the maximal conditional probability P(O/W),

which means a probability that a word/event W will represent an acoustical vector/observation O. An HMM model is described by a number of emitting states and by a number of Probability Density Function (PDF) mixtures per state (Bhardwaj et al. 2015).

Mel-Frequency Cepstral Coefficients (MFCC) are very efficient features in characterising various kinds of non-speech and speech sounds such as diverse acoustic events, music, singing voice, birdsong, speech, speaker emotions, etc. (Vozáriková et al. 2011; Lalitha et al. 2015; Nalini et al. 2016). MFCC features are inspired by human perception through a triangular-shaped filter bank (Mel filter bank), where filters are spaced linearly at a low frequency and logarithmically at high frequencies. Such a spacing of filters corresponds to the better distinguishability of the ear in the low frequencies against the higher frequencies. MFCC coefficients are computed from signal segments, which are divided into short frames, the parameters of the signal being constant. The Hamming window is applied to the frames. Then, these frames are transformed to the frequency domain via discrete Fast Fourier Transforms (FFT), and then the magnitude spectrum is passed through a triangular Mel filter bank. The energy output from each filter is then log-compressed and transformed to the cepstral domain via the Discrete Cosine Transform (DCT).

Having defined all notions which we deem relevant for our study, and having presented the recognition model and related tools that will be used in this investigation, we may now go on to recalling the rationale of this study.

3. Hypothesis

As mentioned earlier, it was hypothesised that if the fundamental frequency, F0, rises within the last two syllables (the sonantic nuclei) of the rhythmic group, and is followed by a silent or a filled pause (hesitation pause or pause for breath), it is likely that the following rhythmic group would contain crucial semantic information. Hence, the anticipation nucleus should correspond to the last two syllables of the rhythmic group, with a rising melody and the subsequent filled or empty pause. Indeed, this nucleus would then be a useful phenomenon for an interpreter in constructing the overall prosodic rendering of the sentence.

4. Methodology

The corpus consisted of video productions of 74 speakers, 38 men and 36 women. All of the productions were speeches, delivered in French, by these seventy-four European Parliament members. Political speeches are read speeches or more or less improvised speeches, or impromptu speech (Dejean Le Féal 1978, 1982). In most cases, the acoustic output is a substrate that is largely based on an initial text, with a relatively high level of redundancy. Hence, such speeches are frequently well structured from a syntactic point of view, accompanied by marked prosodic devices for foregrounding specific segments of the discourse: pitch stress, with the aforementioned rising intonational pattern; emphasis, usually highlighted by a remarkable increase of intensity on the target syllable or syllables; and the specific use of a pause. So even, in rare cases, when a political speech is characterised by standard or neutral word order, there is salience of these prosodic argumentative devices for foregrounding specific segments of the discourse in order to render the speech more persuasive.

4.1. Data collection and processing

Altogether, 278 video recordings in MP4 format were obtained. The entire video speech stream for each speaker was recorded without interruption. Overall, 249 soundtracks were separated from the movies; they were saved in a wav format, in mono-mode at a sampling frequency of 44100 Hz and a 16-bit conversion width. The audio files were finally cut into 7366 sentences comprising identified rhythmic groups, as specified hereafter.

4.2. Perceptual determination of anticipation foci

The aim of the perceptual evaluation test was to determine rhythmic groups and indicate location of anticipatory foci within the sentences. The sentences were perceptually evaluated by 4 Slovak subjects, who are Assistant Professors of French at Pavol Jozef Šafárik University in Košice, 3 female linguists and 1 male linguist. They are all four professional interpreters, with an average of 20 years of experience (± 2 years) in simultaneous interpreting from French to Slovak, and had consequently been sensitised to the intonation structure of French and to rhythmic groups. The evaluation tests were carried out in a silent room, each subject wearing headphones to listen to the speeches.

This subjective perceptual detection of rhythmic groups and anticipatory foci location is a prerequisite in any speech recognition system with a detection algorithm. The acquired objective data and statistical evaluations of anticipatory nuclei of the rhythmic groups, would then serve in the speech recognition learning model.

From a total of 7366 sentences, we selected 200 representative samples of sentences. Among the 200 sentences, we excluded 66 sentences because of rare absence of consensus in the determination of anticipatory foci due to disagreement among the listeners, to fuzzy acoustic phenomena in the audio signals, and, in some cases, to extreme erratic values in the target variables. Such a finding underlines the necessity to use recognition models in order to objectively corroborate perceptual results obtained subjectively.

5. Measurements

Measurements were carried out within the last two syllables of the rhythmic groups and for the following pause, which were perceptually detected and considered as anticipatory nuclei based on the perception evaluation tests. The parameters measured were the following: the time position of the mid-point of the sonantic nuclei of the first and second syllables; the duration of the pause; the absolute value of both the intensity (dB) and fundamental frequency F_0 (Hz) at the sonantic nuclei. The final data resulted in the relative values of an increase or of a decrease of intensity, and in the variations of F_0 within the last two syllables of the rhythmic group or, when possible, in the rate of increase of the fundamental frequency, F_0. As regards F_0 values, the data for female and male speakers were not processed separately, despite the usual average difference of one octave between a female and a male voice, as analyses will be based, in fine, on normalised relative values.

6. Experimental results

Results of statistical analysis of related prosodic features and also results of the automatic anticipation nucleus via classifier based on Hidden Markov Models are presented in this section.

6.1. Statistical analysis of key prosodic parameters

6.1.1. Statistical processing of measured F0 values

We first analysed differences in the fundamental frequency, F_0 of the last two syllables: $F_0 2 - F_0 1$, measured in Hertz. In order to determine the characteristics of the average values and the variance in the statistical data – *i.e.* the values of the differences in F_0, we sorted the measured values and constructed a histogram from the sorted data. Based on the Sturgers rule for determining the number of columns in the histogram, and on the basis of the variation range, the data were divided into 6 intervals according to:

$$k = 1+3{,}3.\log 10(n)$$

where k is the number of columns and n is the number of different character values.

Figure 1 shows a histogram of the measured differences $F_0 2 - F_0 1$. The graph portrays average values and the variance of the measured data, as well as distribution characteristics in the data. The contour of column height in the histogram indicates that it is a normal Gaussian distribution. The Shapiro-Wilk test served to verify the normality of data; results are given in the graph header in Figure 1. We verified the H0 (null) hypothesis using this test: the selection emanates from a sample basic file with only data characterised by a normal distribution. The test showed that this null hypothesis was not rejected. One can observe in the figure a normal distribution curve modelling the distribution of the values of the variable, based on its mean value and scatter, plotted in red.

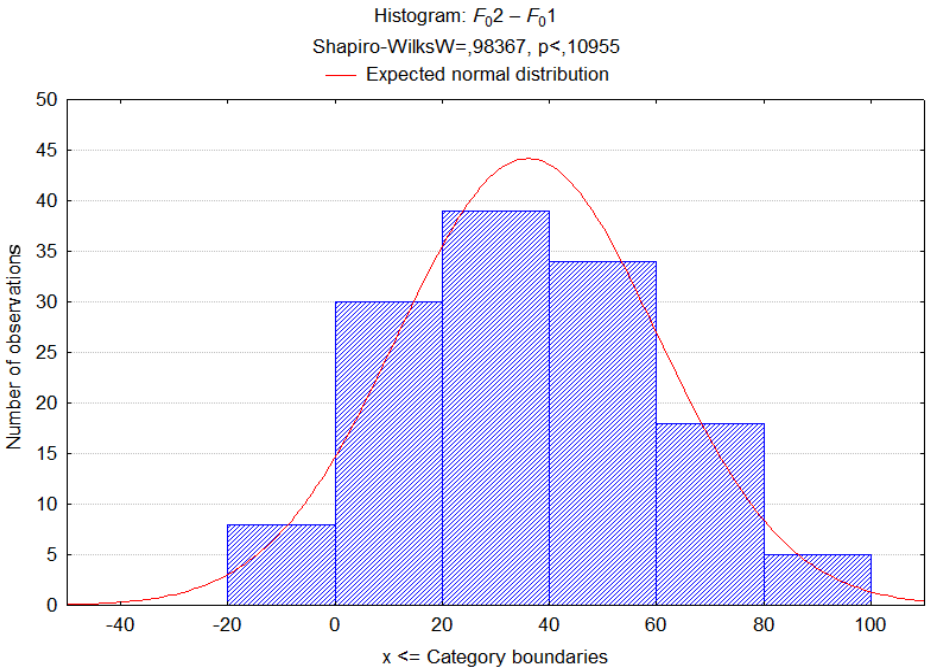

Figure 1. Histogram of the measured differences of fundamental frequency.

The basic descriptive statistics are shown in Table 1. Standard error of the mean value (SEM) is the standard deviation of the data sample diameter, which indicates the difference of the calculated average of a random sample diameter from the mean value of the basic set. It is calculated according to the formula:

$$SEM = \frac{s}{\sqrt{n}}$$

in which *s* is the standard deviation in the measured data, *n* is the amount of data.

The confidence interval for mean value indicates the probability with which the selected value of the diameter can be found in the interval of values. We usually choose a 95% interval of reliability, the significance level being 5% (p = 0.05) in this case. From the 5% level of significance, there are two critical areas, one for the left side of the Gaussian curve with p = 0.025, and one for the right side equally with p = 0.025. The confidence interval results from the relationship:

$$Conf.interval = \bar{x} \pm \left(1{,}96\frac{s}{\sqrt{n}}\right).$$

Table 1 shows that at a significance level of 5%, the mean value of the selected difference of fundamental frequency $F_0 2 - F_0 1$ will be greater than 31.84 Hz and less than 40.11 Hz.

Table 1. Measures of average values and variance of fundamental frequency differences.

Num. of samples	Mean $\bar{x}$	Confidence interval -95,00%	Confidence interval +95,00%	Min	Max	Standard deviation *s*	Standard error
134	35,98	31,84	40,11	-17,00	91,00	24,20	2,09

6.1.2. Statistical processing of measured values of intensity

We analysed the same sentences following the same procedure as that used for the F_0 (see part 6.1.1). We calculated the difference in the absolute value of the intensity of the last two syllables: $I2 - I1$, measured in dB. In Figure 2, a histogram shows the above mentioned intensity differences. To verify the normality of the data, the Shapiro-Wilk test was used here also, and results are shown in the graph header in Figure 2. The test indicates that the selection emanates from a basic set characterised by a normal distribution.

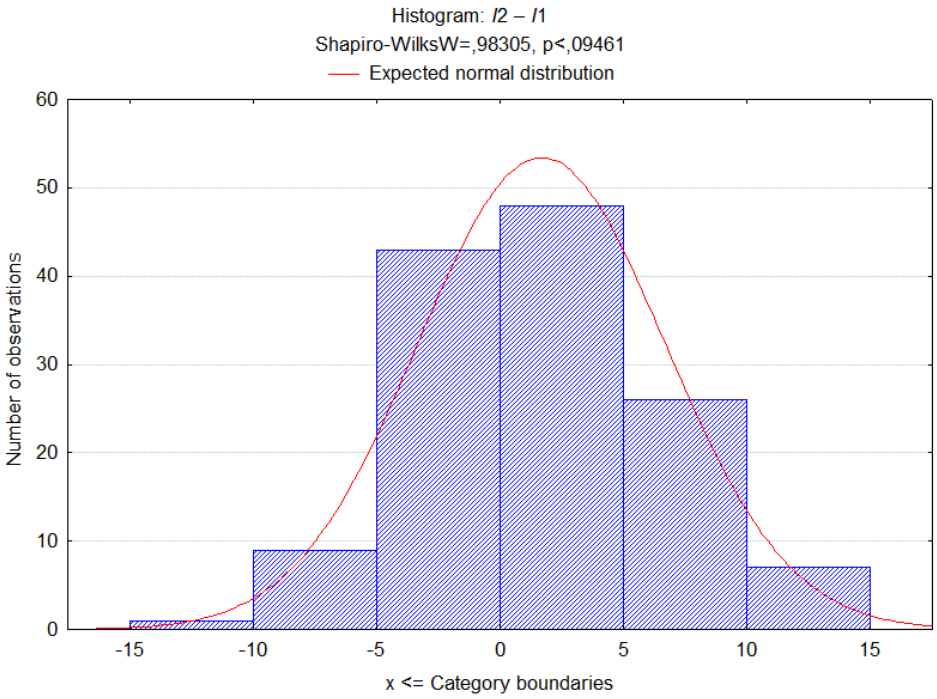

Figure 2. Histogram of the measured differences of the absolute intensity values.

The basic descriptive statistics are shown in Table 2. They have the same meaning as in the case of the F_0 analysis (see part 6.1.1). Table 2 shows that at the 5% significance level, the mean value of the selective distribution of the effective value of intensity $I2$ - $I1$ will be more than 0.83 dB and less than 2.54 dB.

Table 2. Location and scatter statistics of differences of the effective value of intensity.

Num. of samples	Mean $\bar{x}$	Confidence interval -95,00%	Confidence interval +95,00%	Min	Max	Standard deviation s	Standard error
134	1,69	0,83	2,54	-10,90	14,60	5,00	0,43

6.1.3. Statistical processing of the measured duration values of the pause

Here also, we analysed the same sentences following the same processing protocol as that used for fundamental frequency, F_0 (see part 6.1.1). Pauses are measured in seconds (sec). Figure 3 shows a histogram of durational values of pauses. To verify the normality of the data, the Shapiro-Wilk test was used; results are given in the graph header in Figure 3. The test shows that the selection comes from a basic set with a normal distribution.

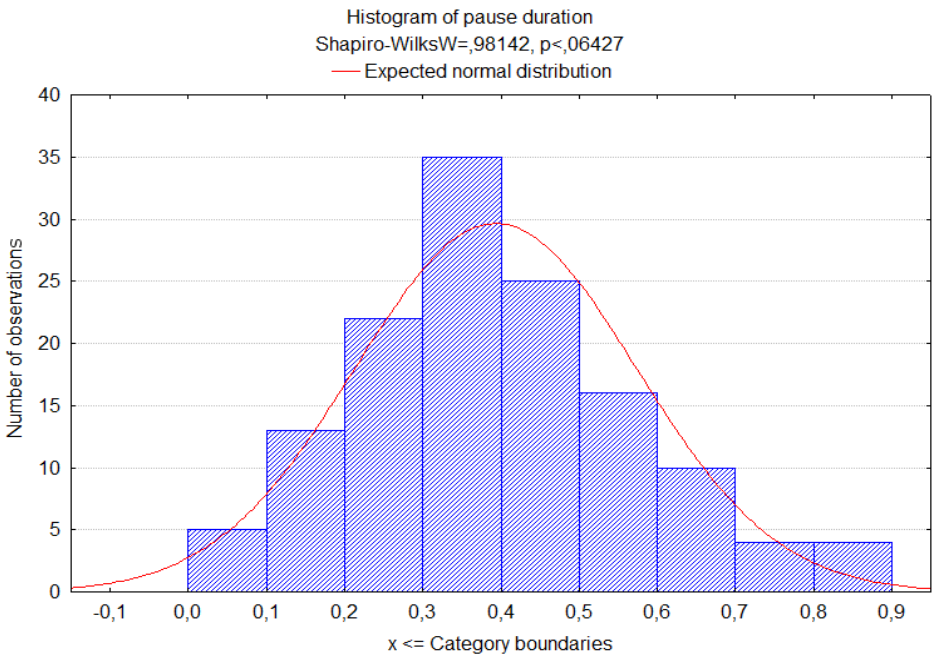

Figure 3. Histogram of pause duration values.

The basic descriptive statistics are shown in Table 3. They have the same meaning as in the case of the F_0 analysis (see part 6.1.1). Table 3 shows that at the 5% significance level, the average pause duration of the sample will be more than 360 milliseconds and less than 420 milliseconds.

Table 3. Location and scatter statistics of pause durational values.

Num. of samples	Mean $\bar{x}$	Confidence interval -95,00%	Confidence interval +95,00%	Min	Max	Standard deviation s	Standard error
134	0,39	0,36	0,42	0,03	0,89	0,18	0,01

6.1.4. Statistical processing of the rate of increase of measured fundamental frequency values

The rate or the steepness of the F_0 increase on the last syllable, compared to the F_0 value of the penultimate syllable before the pause, is derived from the speech tempo, more precisely from the time interval between the sonantic nuclei of the last two syllables $T2 - T1$. We determined this rate on the basis of the following relation:

$$Ascent\ speed\ F_0 = \frac{F_0 2 - F_0 1}{T2 - T1}.$$

The rate of increase of F_0 is determined in Hertz per second. Figure 4 shows a histogram of the increase of F_0 values.

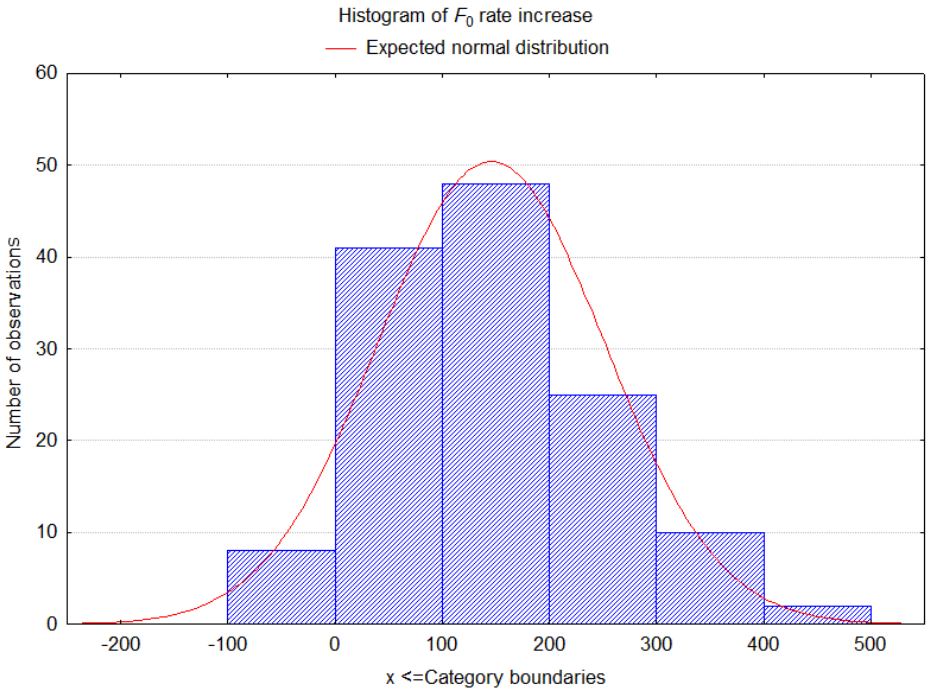

Figure 4. Histogram of measured F_0 rate increase values.

The basic descriptive statistics are given in Table 4, which shows that at a significance level of 5%, the average value of the rate of F_0 increase of the sample will be more than 127.71 Hz/s and less than 163.95 Hz/s.

Table 4: Position and spread statics of F_0 increase rate values.

Num. of samples	Mean $\bar{x}$	Confidence interval -95,00%	Confidence interval +95,00%	Min	Max	Standard deviation s	Standard error
134	145,83	127,71	163,95	-66,66	422,85	106,03	9,16

6.1.5. Examining the correlation between a pause, fundamental frequency increase and tempo

Speech tempo is usually measured by determining the number of syllables per second. As the last two sonantic nuclei represent the last two syllables, their time distance $T2 - T1$ can be considered as the duration of one syllable. For example, if this distance was 0.22 seconds, the speech rate would be 1/0.22 = 4.54 syllables per second.

In Table 5, the correlation coefficient values between a pause, fundamental frequency increase and tempo are shown, calculated at a 5% significance level. The table shows that there are no correlations between the stated parameters retained in determining anticipatory focus.

Table 5. Correlation coefficient values.

	Pause	$F_0 2 – F_0 1$	Tempo
Pause	1,00	0,02	-0,09
$F_0 2 – F_0 1$	0,02	1,00	0,04
Tempo	-0,09	0,04	1,00

6.2. Automatic anticipation nucleus detection by a Hidden Markov Model

The statistical learning method, based on a Hidden Markov Model (HMM), served to detect specific anticipatory foci points within an utterance. These points of interest, so-called anticipatory foci (or sonantic nucleus), were revealed by a similar procedure as applied in a speech recognition task.

In these experiments, the Mel-Frequency Cepstral Coefficients (MFCC) were computed by HTK tools (Lalitha et al. 2015) and the fundamental frequency (F_0) extraction was carried out in the PRAAT environment (Boersma & Weenink 2013). Our solution for the detection of anticipation foci is based on the HMM classification and the Viterbi based decoding algorithm. The principal block scheme of training and testing process is presented in Figure 5.

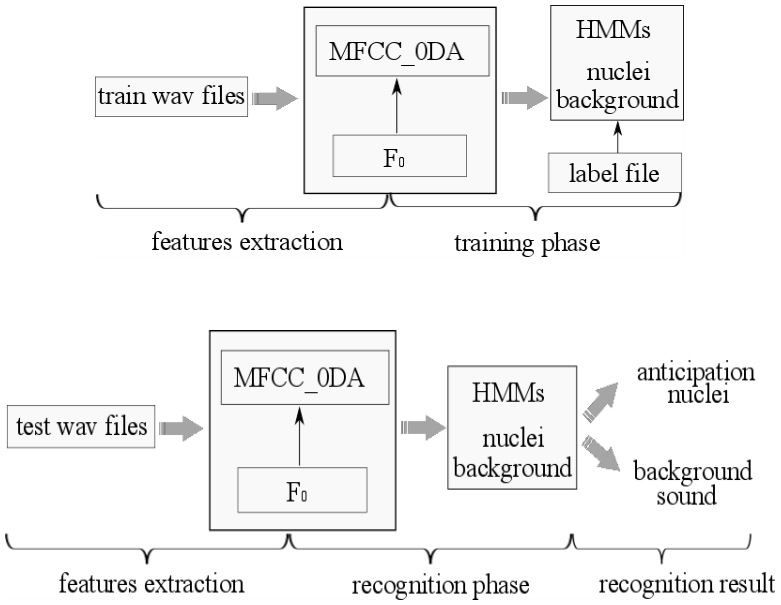

Figure 5. Principal block scheme of the training and testing process.

6.2.1. Experimental set-up

MFCC with zeroth-cepstral and time derivate features- delta and acceleration (MFCC_0DA) were computed from a 25 ms Hamming window, with a 10 ms frame shift by HCopy in HTK. These features were transformed by multiplication of the pitch value; hence, each feature vector was affected by the corresponding F_0. This transformation was carried out in Matlab environment. Before any multiplication, an automatic verification of F_0 values was performed in order to eliminate pitch multiples that may occur, albeit rarely.

Other types of parametrisation based on linear prediction, such as LPC, LPCC (Young et al. 2002) and other approaches to using F_0 to create an effective feature set, were examined. However, the best results were achieved by the help of the abovementioned method.

During the training process, one state HMMs from 1 to 256 PDFs (Probability Density Function) mixtures for the background sounds and for the foreground sounds (anticipation foci) were created (see Figure 5). Globally, 140 anticipation foci were identified with a total duration of approximately 115 seconds. The rest of the sound data were used to train the background model (see

Figure 6). Because of lack of sufficient data (especially of anticipatory foci), cross-validations were carried out; thus 10 train and test sets were created. Obviously, the data that were used for model training were not retained in the test process.

6.2.2. Experimental results of the learning model

One state HMM models to 256 PDFs were successfully trained in an HTK environment with different types of features. In the case of MFCC_0DA and MFCC_0DA$_{F0}$ features, interesting results were obtained. Test files with the alternation of background and anticipatory foci were recognised during the test process. Each test set contains 10 randomly chosen foci and 10 occurrences of background sounds. Results obtained from each data set were then averaged. HResults tool evaluates the output of the HTK decoder with the Accuracy measure (Young et al. 2002) that is described as follows:

$$ACC[\%] = \frac{N - D - S - I}{N} \text{x}100$$

where *D* is the number of deletion errors, *S* the number of substitution errors, *I* the number of insertion errors, and *N* is the total number of labels in the reference label (transcription) files. Results obtained are depicted in Figure 7.

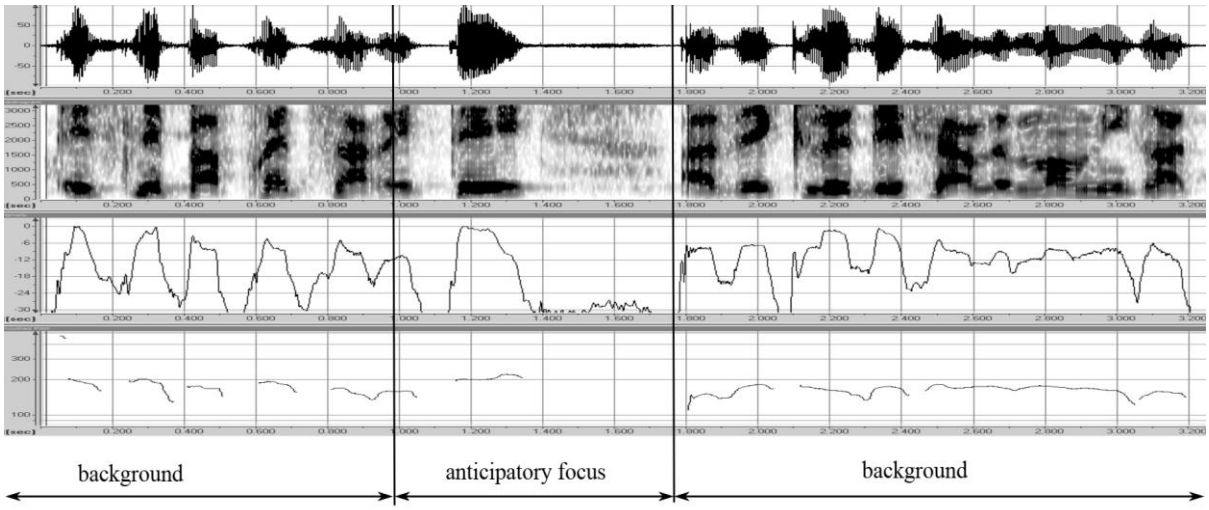

Figure 6. Example in Speech Analyzer 3.0.1. displays- raw waveform, spectrogram, relative intensity [dB] and smoothed F_0. In addition, it shows one sentence with an anticipation nucleus and background sounds.

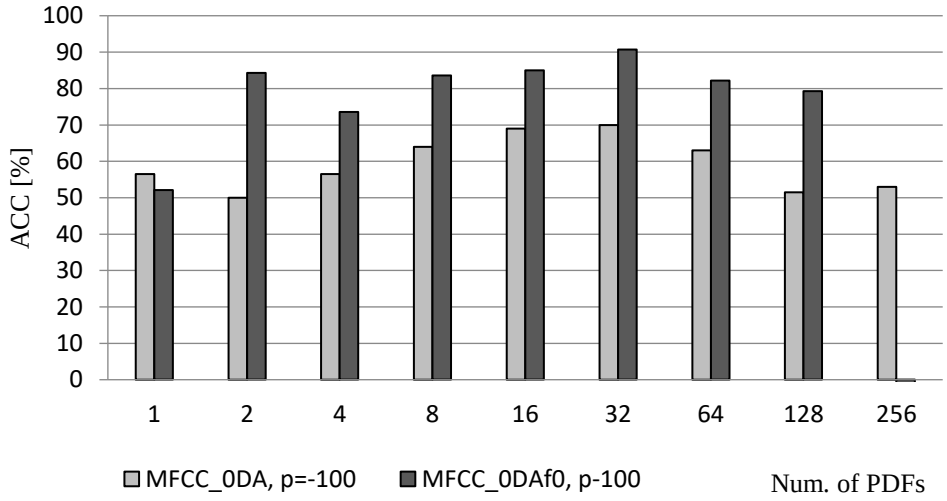

Figure 7. Recognition results for anticipatory focus detection based on the MFCC_0DA and MFCC_0DA$_{F0}$ affected features.

Based on the features used, the positive influence of F_0 was confirmed. Further experiments with a recogniser setup (see Figure 8) were also carried out with two different penalisation factors set by HVite tool with p = –100 and p = –300 (Young et al. 2002). This penalisation factor sets the word insertion probability. Balanced values of ACC measures were obtained by p = –100 to the HMM models with 64 PDFs and 128 PDFs mixtures, but in the case of 256 PDFs mixtures, too many insertion errors were generated. The successive increased values of ACC measures were observed in the case of p = –300.

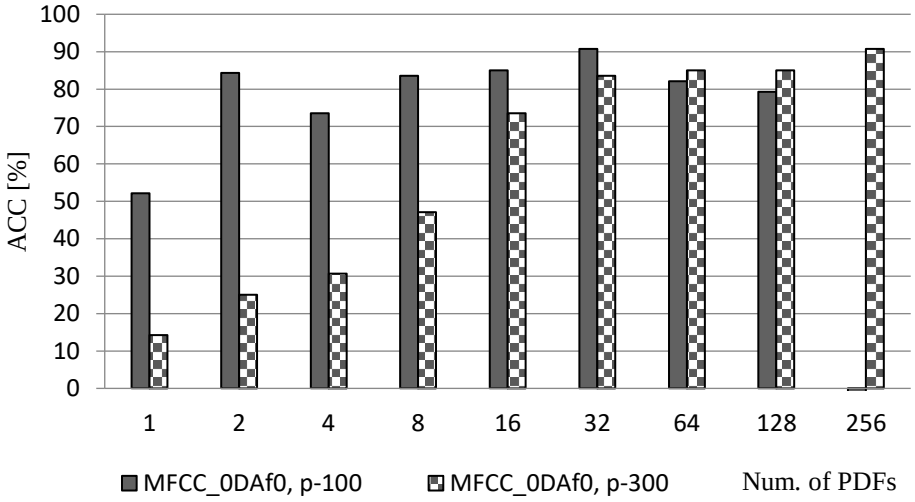

Figure 8. Recognition results for anticipation foci detection with different penalisation factors.

Presented results show that HMM models with 32 PDFs mixtures seem to be suitable in modelling anticipation foci and background sounds. Generally, they gave promising results in all tested scenario (p = –100, –300) in comparison with the other models. The maximal ACC = 90, 7% was yielded two times, for 32 PDFs (p = –100) and 256 PDFs (p = –300).

7. Conclusions

The current research aimed at uncovering and analysing prosodic anticipatory cues for French. The general hypothesis was that, sensitised to these anticipatory prosodic patterns, the interpreter, from French to Slovak, would take advantage of these anticipatory prosodic patterns in French and should hence be capable of integrating them in making strategic interpretation decisions.

We carried out perceptual evaluation tests, acoustic analyses and speech recognition experiments. The perceptual evaluation test served to determine rhythmic groups and to subjectively indicate location of anticipatory foci within the political speeches.

Our initial hypothesis was verified. Indeed, the fundamental frequency, F_0, rose generally within the last two syllables, the sonantic nuclei, of non-terminal rhythmic groups, and was usually followed by silent or filled pauses. The interpreters who served as subjects in the investigation confirmed that the rising melody and the subsequent filled or empty pauses were undeniably useful phenomena in signalling upcoming crucial semantic information contained in the following rhythmic group. Hence, the last two syllables of the rhythmic group, with a rising melody and the subsequent filled or empty pause were considered as the anticipation nucleus.

In order to objectively corroborate the perceptual evaluation experiments, acoustic data (time position of the mid-point of the sonantic nuclei of the first and second syllables, relative F_0 values, relative intensity values, rate of increase of $F_{0,}$ duration of pauses) and statistical evaluations of anticipatory nuclei of the rhythmic groups were used for the learning model.

Results obtained in this study indicate that HMM modelling, in combination with the MFCC pitch affected features, can be effectively applied to the anticipation issue. The proper parametric representation plays a very important role for each recognition system, because recognition performance strongly depends on the type and quality of the features used.

Since simultaneous interpretation is a cross-linguistic cognitive activity, performed under time pressure and in relatively adverse conditions, future research will necessarily have to take into account linguistic interdependent factors related to languages as structurally different as French and Slovak. Such investigations should allow underpinning, experimentally, both anticipatory segmental and supra-segmental cues, which the interpreter would be exploiting (Ardito 1999) in the source language, French, while elaborating strategic decisions in the target language, Slovak.

Acknowledgements

This work was supported by the Slovak Research and Development Agency under contracts No. APVV-15-0307 and APVV-15-0492.

References

Adamowicz, Alicja. 1989. The role of anticipation in discourse: text processing in simultaneous interpreting. *Polish Psychological Bulletin* 20(2). 153–160.

Ahrens, Barbara. 2004. *Prosodie beim Simultandolmetschen*. Frankfurt, Peter Lang.

Ahrens, Barbara. 2005. Analysing prosody in simultaneous interpreting: difficulties and possible solutions. *The Interpreters' Newsletter* 13. 1–14.

Anokhin, Pyotr K. 1968. *Biology and Neurophysiology of the Conditioned Reflex*. Moscow: Meditsina.

Anokhin, Pyotr K. 1978. *Philosophical Aspects of the Functional System Theory*. Moscow: Nauka.

Ardito, Giuliana. 1999. The systematic use of impromptu speech in training interpreting students. *The Interpreters' Newsletter* 9. 177–189.

Bailly, Gérard. 1989. Integration of rhythmic and syntactic constraints in a model of generation of French prosody. *Speech Communication* 8(2). 137–146.

Beaver, David & Clark, Brady Zack & Flemming, Edward & Jaeger, Tim Florian & Wolters, Maria. 2007. When Semantics Meets Phonetics: Acoustical Studies of Second-Occurrence Focus. *Language* 83(2). 245–276.

Bhardwaj, Sanjay & Pathania, Sunil & Akela, Rajesh. 2015. Speech Recognition using Hidden Markov Model and Viterbi Algorithm. *International Journal of Advanced Research in Electronics and Communication Engineering* 4(5). 1179–1183.

Boersma, Paul & Weenink, David 2013. *Praat: doing phonetics by computer*. (Computer program, version 5.3.51) (http://www.praat.org/)

Čeňková, Ivana. 1989. L'importance des pauses en interprétation simultanée. In Bothorel, André (ed.). *Mélanges de phonétique générale et expérimentale offerts à Péla Simon*, 249–260. Strasbourg: Publications de l'Institut de Phonétique de Strasbourg.

Chernov, Ghelly Vassilievich. 1992. Conference interpretation in the USSR: History, theory, new frontiers. *Meta* 37(1). 149–162.

Chernov, Ghelly Vassilievich. 1994. Message redundancy and message anticipation in simultaneous interpreting. In Lambert, Sylvie & Moser-Mercer, Barbara (eds.), *Bridging the Gap: Empirical Research in Simultaneous Interpretation* (Benjamins Translation Library 3), 139–153. Amsterdam: John Benjamins.

Chernov, Ghelly Vassilievich. 2004. *Inference and Anticipation in Simultaneous Interpreting: A Probability-prediction Model* (Benjamins Translation Library 57). Amsterdam: John Benjamins.

Dejean Le Féal, Karla. 1978. *Lectures et improvisations : incidence de la forme de l'énonciation sur la traduction simultanée: francais-allemand*. Paris: University of Paris III. (Unpublished doctoral dissertation.)

Dejean Le Féal, Karla. 1982. Why impromptu speech is easy to understand. In Enkvist, Nils Erik (ed.), *Impromptu speech: A Symposium*, 221–239. Åbo: Åbo Akademi.

Delattre, Pierre. 1966. *Studies in French and comparative phonetics*. The Hague: Mouton.

Falbo, Caterina. 1999. Interprétation: une forme particulière d'oralité. *Revue française de linguistique appliquée* 4(2). 99–112.

Flores d'Arcais, Giovanni B. 1991. Sintassi e psicolinguistica: i processi di elaborazione sintattica durante la comprensione del linguaggio. *Sistemi Intelligenti* 3(3). 315–346.

Gales, Mark & Young, Steve. 2008. The Application of Hidden Markov Models in Speech Recognition. *Foundations and Trends in Signal Processing* 1(3). 195–304.

Gile, Daniel. 1992. Predictable sentence endings in Japanese and conference interpretation. *The Interpreter's Newsletter* 1(Special issue). 12–24.

Gile, Daniel. 1995. *Regards sur la recherche en interprétation de conférence*. Lille: Presses universitaires de Lille.

Huang, Xuedong D. & Ariki, Yasuo & Jack, Mervyn A. 1990. *Hidden Markov Models for Speech Recognition*. New York: Columbia University Press.

Kintsch, Walter & van Dijk, Teun A. 1983. *Strategies of Discourse Comprehension*. London: Academic Press Inc.

Kirchhoff, Helene. 1976. Das Simultandolmetschen: Interdependenz der Variablen im Dolmetschprozess, Dolmetschmodelle und Dolmetschstrategien. In Drescher, Horst W. & Scheffzek, Signe (eds.), *Theorie und Praxis des Überstzens und Dolmetschens*, 59–71. Frankfurt: Lang.

Kleiber, Georges & Sock, Rudolph. 2006. Ce+N+Relative: Sémantique et Prosodie. *Lingvisticae Investigationes* 29(2). 251–273.

Kohn, Kurt & Kalina, Sylvia. 1996. The strategic dimension of interpreting. *Meta* 41(1). 119–138.

Lalitha, S. & Geyasruti, D. & Narayanan, R. & Shravani, M. 2015. Emotion Detection Using MFCC and Cepstrum Features. *Procedia Computer Science* 70. 29–35.

Lederer, Marianne. 1978. Simultaneous Interpretation – Units of Meaning and Other Features. In Gerver, David & Sinaiko, H. Wallace (eds.), *Language Interpretation and Communication*, 323–332. New York: Plenum Press.

Lederer, Marianne. 1981. *La traduction simultanée: expérience et théori*e. Paris: Minard.

Lubker, James F. & Lindgren, Rolf. 1982. The perceptual effects of anticipatory coarticulation. In Hurme, Pertti (eds), *Papers in Speech Research*, 252–271. Jyväskylä: Institute of Finnish Language and Communication, University of Jyväskylä.

Marslen-Wilson, William D. 1989. Access and Integration: Projecting Sound onto Meaning. In Marslen-Wilson, William D. (ed.), *Lexical Representation and Process*, 3–23. Cambridge (Mass.): MIT Press.

Martellini, Sara. 2013. Prosody in Simultaneous Interpretation: a Case Study for the German- Italian Language Pair. *The Interpreters' Newsletter* 18. 61–79.

Massaro, Dominic W. 1975. *Understanding language: An Information-Processing Model of Speech Perception, Reading, and Psycho-Linguistics*. New York: Academic Press.

Massaro, Dominic W. 1978. An information-processing model of understanding speech. In Gerver, David & Sinaiko, H. Wallace (eds.), *Language Interpretation and Communication*, 299–312. New York: Plenum Press.

Massaro, Dominic W. & Shlesinger, Miriam. 1997. Information processing and a computational approach to the study of simultaneous interpretation. *Interpreting* 2(1/2). 13–53.

Mattern, Natalie. 1974. *Anticipation in German-English Simultaneous Interpreting*. Saarbrücken: University of Saarland. (Unpublished MA thesis.)

Moser, Barbara. 1976. *Simultaneous translation: Linguistic, Psycholinguistic and Human Information Processing Aspects*. Innsbruck: University of Innsbruck. (Doctoral dissertation.)

Moser, Barbara. 1978. Simultaneous Interpretation: A hypothetical model and its practical application. In Gerver, David & Sinaiko, H. Wallace (eds.), *Language Interpretation and Communication*, 353–368. New York: Plenum Press.

Moser-Mercer, Barbara. 1994. Paradigms gained or the art of productive disagreement. In Lambert, Sylvie & Moser-Mercer, Barbara (eds.), *Bridging the Gap: Empirical Research in Simultaneous Interpretation* (Benjamins Translation Library 3), 17–23. Amsterdam: John Benjamins.

Moser-Mercer, Barbara & Frauenfelder, Ulrich Hans & Casado, Beatriz & Künzli, Alexander. 2000. Searching to define expertise in interpreting. In Dimitrova, Birgitta Englund & Hyltenstam Kenneth (eds.), *Language Processing and Simultaneous Interpreting: Interdisciplinary Perspectives* (Benjamins Translation Library 40), 107–132. Amsterdam: John Benjamins.

Nalini, N. J. & Palanivel, S. 2016. Music emotion recognition: The combined evidence of MFCC and residual phase. *Egyptian Informatics Journal* 17(1). 1–10.

Perkell, Joseph S. & Chiang, C. 1986. Preliminary support for a 'hybrid model' of anticipatory coarticulation. *Proceedings of 12th International Congress of Acoustics*. A3–A6.

Riccardi, Alessandra. 1997. Lingua di Conferenza. In Gran, Laura & Riccardi Alessandra (eds.), *Nuovi orientamenti negli studi sull'interpretazione*, 59–74. Padova: CLEUP.

Riccardi, Alessandra & Snelling, David C. 1997. Sintassi tedesca: vero o falso problema per l'interpretazione?. In Gran, Laura & Riccardi Alessandra (eds.), *Nuovi orientamenti negli studi sull'interpretazione*, 143–158. Padova: CLEUP.

Seeber, Kilian G. 2001. Intonation and Anticipation in Simultaneous Interpreting. *Cahiers de linguistique française* 23. 61–97.

Seleskovitch Danica. 1968. *L'interprète dans les conférences internationales: problèmes de langage et de communication*. Paris: Lettres Modernes.

Seleskovitch, Danica. 1984. Les anticipations de la compréhension. In Seleskovitch, Danica & Lederer, Marianne (eds.), *Interpréter pour traduire*, 273–283. Paris: Didier Erudition.

Setton, Robin. 1999. *Simultaneous interpretation: A Cognitive-pragmatic Analysis*. (Benjamins Translation Library 28). Amsterdam: John Benjamins.

Shlesinger, Miriam. 1994. Intonation in the production and perception of simultaneous interpretation. In Lambert, Sylvie & Moser-Mercer, Barbara (eds.), *Bridging the Gap: Empirical Research in Simultaneous Interpretation* (Benjamins Translation Library 3), 225–236. Amsterdam: John Benjamins,

Sock, Rudolph & Vaxelaire, Béatrice. 2004. Le diable perceptif dans les détails sensori-moteurs anticipatoires. In Sock, Rudolph & Vaxelaire, Béatrice (eds.), *L'anticipation à l'horizon du présent*, 141–157. Brussels: Mardaga.

Van Dam, Ine Marie 1989. Strategies of simultaneous interpretation. In Gran, Laura & Dodds, John M. (eds.), *The Theoretical and Practical Aspects of Teaching Conference Interpretation*, 167–176. Udine, Companotto Editore,

Vaxelaire, Béatrice & Sock, Rudolph & Roy, Johanna-Pascale & Asci, Aline & Hecker, Véronique. 2003. Audible and inaudible anticipatory gestures in French. *15th International Congress of Phonetic Sciences*. 447–450.

Vozáriková, Eva & Juhár, Jozef & Čižmár, Anton. 2011. Acoustic Events Detection Using MFCC and MPEG-7 Descriptors. *Multimedia Communications, Services and Security: 4th International Conference, MCSS 2011, Krakow, Poland, June 2-3, 2011* (Communications in Computer and Information Science 149). 191–197.

Wenk, Brian J. & Wioland, François. 1982. Is French really syllable timed?. *Journal of Phonetics* 10. 193–216.

Wilss, Wolfram. 1978. Syntactic Anticipation in German-English Simultaneous Interpreting. In Gerver, David & Sinaiko, H. Wallace (eds.), *Language, Interpretation and Communication*, 343–352. New York & London: Plenum Press.

Young, Steve & Evermann, Gunnar & Kershaw, Dan & Moore, Gareth & Odell, Julian & Ollason, Dave & Povey, Dan & Valtchev, Valtcho & Woodland, Phil. 2002. *The HTK Book (for version 3.2.)*. Cambridge: Cambridge University Engineering Department.

Zanetti, Roberta. 1999. Relevance of anticipation and possible strategies in the simultaneous interpretation from English into Italian. *The Interpreter's Newsletter* 9. 79–98.

Eva Kiktová
LICOLAB (Language Information and Communication Laboratory)
Department of Slovak Studies, Slavonic Philologies, and Communication
Faculty of Arts
Pavol Jozef Šafárik University in Košice
Slovakia
eva.kiktova@upjs.sk

Július Zimmermann
LICOLAB (Language Information and Communication Laboratory)
Department of Slovak Studies, Slavonic Philologies, and Communication
Faculty of Arts
Pavol Jozef Šafárik University in Košice
Slovakia

julius.zimmermann@upjs.sk
Mária Paľová
LICOLAB (Language Information and Communication Laboratory)
Department of British and American Studies
Faculty of Arts
Pavol Jozef Šafárik University in Košice
Slovakia
maria.palova@upjs.sk

Rudolph Sock
University of Strasbourg,
Linguistics, Languages and Speech Research Unit - LiLPa
France
LICOLAB (Language Information and Communication Laboratory)
Department of British and American Studies
Faculty of Arts
Pavol Jozef Šafárik University in Košice
Slovakia
sock@unistra.fr
rudolph.sock@upjs.sk

In SKASE Journal of Theoretical Linguistics [online]. 2020, vol. 17, no. 4 [cit. 2020-10-27].Available on web page http://www.skase.sk/Volumes/JTL45/pdf_doc/04.pdf. ISSN 1336-782X