Dimensions of intonation. Wiktor Jassem's contribution to the studies on the melody of speech

Maciej Karpiński

Institute of Linguistics, Adam Mickiewicz University in Poznań
e-mail: maciejk@amu.edu.pl

Abstract: Intonation is often considered the most problematic component of prosody. It may largely contribute to the meaning of an utterance and provide rich indexical information. The complexity of the intonation-related phenomena, from the acoustic level up to the cognitive process of semiosis, resulted in a variety of conceptualisations and divergent theoretical approaches. In the present text, an attempt is made to sketch the late Wiktor Jassem's contribution to intonation studies, starting from his early work on colloquial British English to the studies focused on the Polish language as well as his general perspective on prosody and on the methods of its exploration. Some of his unpublished contribution to research projects on various aspects of prosody is also acknowledged. The present overview is based both on available written materials (publications, references, reports, reviews) and the author's personal communication with Wiktor Jassem and his collaborators.

Keywords: Wiktor Jassem, phonetics, prosody, intonation, Polish, English

1. Conceptual divergences in the study of intonation

Intonation is often declared to be a linguistic universal but its form and function strongly varies across languages (for references and discussion, see Ladd 1996; Cruttenden 1981, 1997; Fox 2000, Gussenhoven 2004, and many others). Commonalities as well as differences are not always obvious, still posing challenges and raising discussions in phonetics and phonology among other areas. The complexity of the phenomena under question contributes to formulating models and theories based on different assumptions and conceptual frames. Although it is not intended nor possible to encompass this diversity here, major doubts and controversies are mentioned below in order to provide a minimum background for further discussion on Wiktor Jassem's work in the field of intonation.

Perceivable changes of pitch or a lack of them in the speech signal may occur due to a number of factors related to different stages of speech production. Some of them may result directly from the implementation of language system rules while some others, referred to as non-linguistic, extra-linguistic or paralinguistic, occur due to certain mental, especially attitudinal or emotional states of the speaker, as well as due to physiological

or interaction-related factors. The distinction between the two components of intonation is often not obvious (Gussenhoven 2004) and poses a number of fundamental questions (Karpiński 2012). Surprisingly, some relevant doubts were formulated in early works by Jassem (1952), foreshadowing later arguments.

Among contemporary approaches to the conceptualisation and modelling of intonation, one usually distinguishes between sequential (linear) and superpositional ones. While the "British school" (Jassem 1952, Kingdon 1966, Crystal 1969), the "Dutch school (also referred to as "IPO approach", cf. 't Hart, Collier & Cohen 1990), the approach based on Autosegmental-Metrical Phonology (Pierrehumbert 1980) as well as later "ToBI-inspired" systems, or the more technology-oriented Tilt model (Taylor 1998) fall into the former category, the latter includes e.g., Fujisaki's model (Fujisaki 1983), Linear Alignment Model (van Santen & Möbius 1997), SFC (Superposition of Functional Contours; Bailly & Holm 2005), and PRISM (Mishra 2008). There are also models that are more difficult to categorise along this dimension, e.g. STEM-ML (Soft-template Markup Language; Kochanski & Shih 2000) or Kiel Intonation Model (Kohler 1995). Kohler's work (as well as e.g. Auberge's, Cruttenden's or Bolinger's) is sometimes associated with the hierarchical approach distinguished by some other prosodists (e.g., Rossi 2000: 34-35). The hierarchy itself may be viewed and conceptualised in a number of ways. For example, Pierrehumbert's system features two levels of intonational units: intermediate phrases and intonational phrases. While there have been quantitative attempts to confirm this two-level structure for the Polish language on the grounds of the AM approach (Wagner 2008), there have also been works inspired by the British school and still implementing the two-level structure (e.g., Karpiński et al. 2008). Jassem, in principle, was reluctant to accept such an option, strictly honouring the condition of one *ictus* per intonational phrase and a non-hierarchical structure.

The "levels vs. configurations" argument (Bolinger 1951; for a more recent discussion, see Ladd 1983:49-59) provides grounds for another categorisation of different approaches. While some scholars assume that the pitch contour is produced as a result of moving to subsequent pitch levels (targets) that are available to the speaker of a given language, others claim that there exist more complex indecomposable entities – pitch configurations – typical of a given language and followed by its speakers. On the shallow level of comparison, the British School is often associated with a more holistic pitch contour analysis, while the American structuralist analysis and the Autosegmental-Metrical approach is viewed as more "atomistic", bound to a limited inventory of (two) tones. However, tones or their sequences form pitch accents, and the pitch contour is viewed as a sequence of pitch accents and boundary tones, the latter used to delimit intonational phrases. Pierrehumbert (1980) assumes that for a given language (AmE), there is a limited number of tone sequences that may occur in a well-built intonational phrase. All the acceptable sequences can be generated by a language-specific transition network. On the other hand, on may consider a structuralist approach proposed by Steffen-Batóg who bases her system on a limited number of potential phrase melody patterns – the set of 26 intonemes (Steffen-Batóg 1966, 1996). Revisiting the "levels vs. configurations" argument, Ladd (1983:57) points to some weaknesses of American Structuralist Level Analysis (ASLA) but he simultaneously offers ways to deal with them. He states that as

current "level analyses" take into account much of the criticism coming from the con-figurationists", and the character of the initial argument has certainly evolved, he also declares that "the chapter of intonation studies opened by the American structuralists is still far from concluded."

This brief discussion is meant to indicate that the study of intonation is plagued with serious conceptual and methodological controversies, some of them remaining unsolved for decades. It may also help to explain why some of Jassems ideas that emerged in the 1950s still recur in contemporary studies of prosody.

2. Jassem's approach to intonation analysis: The perceivable and the measureable

Perception-based analysis of intonation is central to the British School. Nevertheless, proper listening may require some specific skills. In his doctoral thesis, Jassem states that

"Phonetic training enables the student to cast off certain prejudices which he has quite natu-rally formed in consequence of his own speech habits." (Jassem 1952:18-19)

And further he writes:

"The auditory analysis is hampered, in the first place, by the influence of the spelling and the morphonological structure of the language." (*ibidem*)

He also argues that:

"[…] speech is primary perceived as consisting of segments, i.e. certain minimal auditory impressions, and not by reference to any enigmatic non-physical 'phonemic features'." (*ibidem*)

Finally, he declares:

"A phonetic or tonetic description is correct if the investigator has no hearing defect and is able to discriminate between phonetic or tonetic segments because he has cast off all preju-dices, and has been made to notice all such features of segments as can be noticed by a nor-mal human ear, and if he has adequate means of describing what he hears in unambiguous terms or graphic symbols whose values are known." (*ibidem*, pp. 22-23)

Before any attempts to evaluate these claims, one must take their historical context into account. According to Liberman (1996), the presumption that phoneme perception is accomplished by detecting the acoustic characteristics that corresponds to each pho-neme or by comparing the incoming signal to a phoneme template stored in memory, was present in the early days of SP since the 1940s. But while Fant's influential publica-tions supporting this approach appeared only in the 1960s (1960, 1962, 1967), the onset of the "object of speech perception" controversy dates back to Liberman's earlier studies (Liberman et al. 1952; Liberman 1957). Also the discussion on the units of speech per-ception came to the surface early in the seventies (e.g., Savin & Bever 1970, McNeill & Lindig 1973), finding its culmination in Mehler's (Mehler 1981; Mehler et al. 1981) and Cutler's (e.g., Cutler et al. 1986) works. Jassem puts his faith in expert listening, in

the domain of both segmentals and suprasegmentals. Although he mentions certain conditions (e.g., training, casting off all "auditory" prejudices) to be met by the listener, the requirements are not extreme. The idea of direct perception of "minimal auditory impressions", not involving abstract and "enigmatic" phonemic features, refers to the segmental level but one may assume that a similar mechanism may be applied in the perception of suprasegmentals.

Early Jassem claims that:

> "Thus it is possible to speak of two identical sounds. In order to say that sound (a1) is the same as sound (a2), it is *not* necessary to postulate an ideal sound (a) which exists in souls or in an immaterial world and is "materialized" or "realized on earth" in two, physically different shapes. It is in fact not necessary to leave "the earth" at all. Two sounds (a1) and (a2) are identical simply if a normal, phonetically trained ear can detect no difference between them. (Jassem 1952:23)

Again, perception is a central and almost absolute reference. Being fully conscious that there are no two acoustically identical sounds of speech, Jassem focuses here solely on their perception and the resulting perceptual impressions, rejecting the necessity of introducing any "ideal sounds" (sound categories or other intermediate categories) in the course of verification of this claim. Obviously, this does not mean the rejection of phoneme and phonemic analysis. However, as he points out, "Phonemic analysis is based on the consideration of contextual interrelations between phonetic segments." which suggests some type of distributional approach (Jassem 1952:24).

Nevertheless, Jassem carefully distinguishes between the phonetic and tonetic analyses, and puts much effort in his attempts to show that many methods and achievements of the former may be more or less directly employed in the latter. This distinction may be less emphasized nowadays (Port 2008) although in much recent overview of methods in intonation analysis, Vaissière (2005:240) mentions the "non-applicability of well-established research methods", including those applied in segmental phonetics.

Early Jassem summarises his approach to intonation analysis in six postulates. He assumes that only those sound features are linguistically significant that are detectable by ear and conventionally recurrent in a speech community. He stresses the functional aspect of segmental categorisation (referring to both phonetic and tonetic segments) and the fact that the meanings do not recur as identical situational elements but some of them "can and must be described as not different from each other" (Jassem 1952:24-25). Finally, he warns the reader against "speculating philologist's pitfalls" and points to neural activity (associative cerebral processes) as a reference to be relied on (*ibidem*). Again, while these ideas are perfectly sound and acceptable today, they may have sounded revolutionary and futuristic in the sixties.

While Jassem always remained bound to the British School and the belief that ears were the primary tools of a phonetician, from his early years in research he was open to all the technological achievements in computer-based speech signal measurements and statistical modelling. Accordingly, while applying the "close listening" approach, he was still exploring and testing facilitating methods such as using slowed-down playback, backwards listening, or listening to syllables in various orders. He used processed signals and

resynthesized speech in perception experiments (for examples, see section 5). He was also very interested in software that would help with automatic intonation annotation of speech corpora (INTSINT; Campione et al. 2000) and with modelling the perception of intonation (Prosogram; d'Alessandro & Mertens 1995; Mertens 2004). Fully understanding their limitations, he still considered them useful and very helpful tools that would play a crucial role in future corpus-based intonation research. He himself took part in the development of the tools that would automatically classify pitch traces using discriminant functions and artificial neural networks (see section 5).

3. Levels of representation and analysis

Early in the 21st century Wiktor Jassem returned to his work on Polish intonation to face the necessity of dealing with relatively large corpora of spontaneous speech (e.g., PoInt corpus, Karpiński & Kleśta 2001). The largely unchanged methodological and terminological situation sketched out in section 1 inspired him to dwell more on the technical and methodological aspects and conditions of such studies. At that time, he was diligently following and testing Hirst's and Merten's efforts in pitch contour modelling and the automatic annotation of intonation (see section 2). He believed those methods would be further developed and perfected but they could be also used immediately as auxiliary tools – at least by those who are conscious of how they worked and what kind of issues were to be expected. Pointing to the rapidly growing body of intonation studies, Jassem (2002:291) noticed the methodological discrepancy and variation that often make comparative meta-analyses impossible. He formulated his "research manifest", overtly different from those inspired by autosegmental-metrical phonology, and based on a "strictly distributional definition of intonational distinctiveness" (*ibidem*, 293). What may be important to mention, however, is that he did not understand the structural and autosegmental-metrical approaches as mutually exclusive but rather as complementary, the former being empirical and inductive while the latter – deductive and rationalistic (Jassem 1999:33-34).

According to Jassem, one of the weaknesses of the early AM-based analyses was the lack of a clear procedure for the translation from the pitch contour to the phonological level and back while such procedures should be available as a part of a model of intonation (e.g., Hirst et al. 2000:5). Moreover, an insightful look at theoretical perspectives, methods and techniques of studying intonation, led Hirst and his colleagues (Hirst et al. 2000; see also Hirst 2005) to the conclusion that they often operated on different levels of analysis and the studies pertain to fundamentally different phenomena from different levels of abstraction. Discussions over these issues as well the above mentioned publication by Hirst et al. itself induced Jassem to declare his own viewpoint. Slightly modifying Hirst's proposal, he postulated four levels of representation,
1. Acoustic
2. Acoustic-phonetic
3. Perceptual
4. Phonological

The acoustic level referred to the physical, acoustic sound parameters. At the second level, f_0 was smoothed and normalised for the speaker's individual mean pitch and pitch range. In this way, personal variability was largely eliminated. (One must note, however, that Hirst and his colleagues did not overlook the issue of speaker normalisation – they referred to relevant publications by Hermes, Campbell, Terken and others.). Only at the third level perception-related phenomena were taken into account. And finally, at the fourth level, one might discuss the phonology of intonation, keeping in mind the steps that have been made on the lower levels. Subsequent levels of representation and analysis may be viewed as a set of filters. The ultimate, phonological level abstracts from speaker-specific indexicals and from any non-linguistic intonational meaning (Gussenhoven 2004; but cf. Jassem's seemingly more flexible approach in (1962)).

The importance of this apparently obvious distinction, along with clear definitions of the levels of analysis and representation, is of paramount methodological and technical importance, going far beyond any schools and approaches. It is becoming especially clear when working with emotional speech, the speech of children, of the elderly, regional language varieties, and so on, where there are no clear indications of what is linguistic and what is not or which features of speech are linguistically relevant. Their relevance may be based on their place in the phonological system or on some social and cultural settings. Belonging to the same phonological category, actual units of speech may differ in ways that can be captured only on some lower levels of analysis.

4. Units of intonation analysis. Tunes and beyond.

Under the burden of years of written language studies, linguists often tend to neglect the differences between script and speech, and engage the analytic tools handcrafted for written texts in the analysis of spoken language (Linell 2005; Karpiński 2012). Even the basic units of analysis that are relatively easy to define in written texts become less obvious and sometimes incompatible with what is found in speech (e.g., a sentence). The problem lies not only in the conception and the internal structure of the unit but also in the practical means of determining its boundaries. Other issues may emerge when the hierarchy of units or their well-formedness is considered. Keeping these facts in mind, one can see at least some factors that make defining basic units of intonation analysis surprisingly difficult.

In most of his texts on English intonation, Jassem operates on the level of *tunes*. He assumes that "every language has a finite number of basic intonation patterns" (Jassem 1999:36) that correspond to what Jones (1956) termed *tunes*. He accepts the term *intonation phrase* as "a cover term for the different intonation patterns" but stresses that "a fully satisfactory definition of an intonation phrase is not yet available" (Jassem, *ibidem*).

Jassem's concept of the basic tune is closely bound to the British School and it is often compared to the model by O'Connor and Arnold (1961):

[prehead [head [[nucleus] tail]]]

Jassem's model is mostly known as incorporating an additional component, the anacrusis, i.e. a sequence of syllables that are pronounced extremely rapidly and do not show any relation between their durations and number in succession (Jassem 1952:40):

[anacrusis][[preictic/prenuclear intonation [ictic/nuclear intonation]]]

In this model, the constituents of the tune include both those that are optional (anacrusis, prenuclear intonation), and those that are obligatory, i.e. the nuclear intonation. The nuclear intonation contains one and only one *ictic* accent (main intonational accent) and, optionally, one or more post-ictic non-intonational accents. This is expressed symbolically in the following formula (Gibbon 1984):

(TRU (ANA [+syll, -stress]*) (NRU [+syll, +stress] [+syll, -stress]*)) RJ

The concept of anacrusis, however, emerged from Jassem's earlier work on speech rhythm in British English (Jassem 1949). He proposed basing the model on the Broad Rhythm Unit (BRU) which consisted of

- an optional Anacrusis, i.e. a sequence of unstressed syllables from a grammatical boundary up to (but not including) the next stressed syllable;
- an obligatory Narrow Rhythm Unit (NRU), consisting of a stressed syllable followed optionally by a sequence of unstressed syllables, extending to the next relevant grammatical boundary.

Tonal Unit is a domain of pitch movement while the *Narrow Rhythm Unit* and *Anacrusis* are the domain of segmental timing. The *Narrow Rhythm Unit* is a sequence of syllables, with the first one being rhythmically strong and the last one being followed by a rhythmical juncture (Jassem 1952:40). As Hirst (2012:29) points out, it is basically identical to the *Stress Foot* (introduced several year later by Abercrombie and Halliday).

Distinguishing between the *Tonal Unit* and the (*Narrow*) *Rhythm Unit* is of twofold importance. It suggests that efficient prosody modelling may require separate units for the tonal and the rhythmic domains. In the same time, the entire model shows the proximity and mutual influences of these two classes of prosodic phenomena and domains. Jassem elaborates this viewpoint further in (Jassem 1999), where he jointly discusses the domains of stress, accent and intonation.

In order to deal with the tone-group melody in English, Jassem adopts six classes of tonal units (level, falling, rising, falling-rising, rising-falling and rising-falling-rising) and proposes his inventory of twelve *nuclear* tunes. Prenuclear tunes are subdivided into eight classes. A simple and intuitive set of symbols is designed to represent the tunes: tiny arrows are meant for the nuclear ones while prenuclear ones annotated as short lines (Jassem 1952:58-60). In 1959, Jassem proposed an inventory of nuclear tunes for the Polish language that includes six melodies (low falling, high falling, low rising, high rising, low level, and high level; Jassem 1959:262). According to his view, Polish seems to be deprived of more elaborated (e.g., falling-rising) contours that may occur in English (Jassem 1972b).

Following the British tradition, Jassem mostly refers to each nuclear tune as a whole. However, in his later work he pays some attention to the inventory of potential pitch levels necessary to represent the intonation of a given language. In his last publication on intonation, Jassem (2002:294) proposes five distinctive pitch heights he assumes sufficient for describing intonation of any language:

• Extra Low, Low, Medium, High, Extra High (xL, L, M, H, xH)

Although operating multiple tone labels and replacing word labels with symbols may be viewed as a mere technical option, it offers some advantages over those traditional descriptive labels: Better clarity, technically easier comparative and quantitative analysis as well as more flexibility. For example, "HM" may refer to "high-falling", as Jassem's definition of the latter implies that the lowest pitch band is not reached. Jassem believed that statistical clustering might actually allow for the distinction of a set of typical pitch levels for a given person and, with a larger group of voices under analysis, to decide on the number of relevant pitch heights for a given language. This would be not only very necessary for the understanding of how intonation works but also important from the technological point of view (see section 5).

5. Speech technology and quantitative approach to intonation

Although Wiktor Jassem never gave up close listening as the primary instrument of the study of intonation, he also believed that abundant information provided by the acoustic signal can be somehow bound to the results of perception. With Demenko, Krzyśko and Dobrogowska, he made attempts to verify hypotheses on the inventory of pitch curves in Polish (Jassem & Dobrogowska 1974; Jassem 1987a; Demenko et al. 1987). Simultaneously, he worked on their off-line automatic recognition (Jassem 1987b). In 2011, he supervised M. Wypych's doctoral thesis on the automatic recognition of intonation patterns (Wypych 2011). Jassem was deeply convinced that such software should be widely available. In the eighties, with Demenko, he explored mutual relations between the segmental and suprasegmental structures and features (Jassem & Demenko 1989) but also returned to the very fundamental aspect of phonetic-linguistic studies (Jassem & Demenko 1986): the relation between form and meaning.

Early in the 1970s, Jassem pointed to four major problems in the linguistic interpretation of pitch traces: "(1) How accurate must the instrumental data be to make sure that no linguistically relevant information is missed? (2) Which of the details of the intonation curve are linguistically relevant? (3) How should the graph be scaled? (4) How should the curves be normalized so as to neutralize individual (speaker-dependent) differences?" (Jassem 1971). Most of his later works may be viewed as attempts towards solving these issues. Technology available today makes (1) less relevant even though top class equipment definitely does not guarantee the most accurate voice recordings. The remaining three questions are still not fully answered. The number of conceptually different approaches to pitch contour modelling available today may suggest that the second issue remains unsolved. Similarly, although the logarithmic scale has become the most popular when it comes to representing pitch, other units, including semitones, mels, etc., are still

in use and find applications. The question of normalisation is, to a certain extent, addressed by various modelling approaches. Nevertheless, invariants in pitch contours are still difficult to find and confirm.

In the middle and late seventies, Jassem was preoccupied with the problems of normalising pitch curves (1975) and distributional analysis of pitch phenomena (Jassem 1978). Although the immediate motivation for these studies was rather technological, they influenced his later views to the possibility of building a coherent symbolic system for the annotation of intonation that would somehow join the worlds of perception and acoustic parameters. Most probably, much later it may have induced the inclusion of the acoustic-phonetic level to his inventory of intonation representation levels (Jassem 2002). Jassem's interest in the role of pitch-related parameters in speaker recognition and identification is partially reflected in his papers dealing the average short-term f_0 values as personal voice characteristics (Steffen-Batóg et al. 1970; Jassem 1972; Jassem et al 1973).

Another aspect of his overall approach to prosody finds expression in e.g. (Jassem & Demenko 1997) – a text on the phonetic and grammatical coherence of the phrase. The authors approached the theoretical reconstruction of the syntax-prosody interface by looking for an interrelation between the intonational and grammatical phrase boundaries.

While Jassem's work on Swedish is devoted to its pitch accent and not to its intonation in the narrow sense, it is worth mentioning here as an example of his capability of employing speech technology to support experimental designs that deal with theoretical, phonological or linguistic issues. Jassem (1962) explores pitch as a correlate of Swedish pitch accent using perception tests. He points to the fact that the number of minimal pairs based on word accent is quite limited and some of them may not be primarily distinguished on the basis of different pitch configurations. Further, he argues that "the existence of such „minimal pairs" is not a necessary condition for the establishment of relevant (distinctive) word accent.", and it is sufficient to show "that (1) the distinction is not free, and (2) that the distinction is not predictable" (Jassem 1962:4). As a further step, he used synthetic speech (OVE II system) in order to explore various configurations of pre-tonics and accents, and shows that pitch manipulation alone may result in switching accent category (Jassem 1963:12). With Morton and Steffen-Batóg, he used speech-like synthetic stimuli in the perception study of stress in Polish (Jassem et al. 1968). Years later, he returned to this method in a study conducted with van Dommeln, devoted to the perception of accent in re-synthesized speech (Jassem & van Dommeln 1990).

6. Collaboration and impact

Most often mentioned as a major contributor to the British School of intonation analysis and an insightful explorer of British English prosody, Jassem never gave up his research on this langue. In some of his later works, he re-visited fundamental aspects of British intonation (1987c; 1994; 1996a, 1996b, 1999), including its linguistic function (Jassem 1994) and the inventory of nuclear tones (1996a, 1996b).

He also influenced many researchers working with a number of other languages. Even though intonation studies do not dominate his body of phonetic work confessed to the Polish language, his impact in this field is undeniable.

In the late 1960s and early 1970s, Jassem collaborated with Maria Steffen-Batóg. Her study of Polish intonation, although completed over fifty years ago (Steffen-Batóg 1966; fully published only in 1996), remains difficult to challenge in terms of scope, formal precision and deep linguistic grounding. Steffen-Batóg (1966) carefully analysed Jassem's approach, referring to his most influential publications (1952, 1959, 1962), but she built her own elaborated system. Although they co-authored a number of publications, their views on intonation modelling, especially on the phonological level, were seemingly different. Their common publications on intonation are focused rather on the acoustic or the acoustic-phonetic level than on the intonational structure modelling.

Since the late 1970s, Grażyna Demenko belonged to Jassem's closest collaborators in the area of prosody and speech technology. Her approach to intonation as presented in (Demenko 1999) stems directly from Jassems' ideas (1959 and later). She made further steps towards the implementation of intonation models in speech technology, and first of all in the field of speech synthesis. In the beginning of the 2000s, Jassem joined the *PoInt* project (Karpiński & Kleśta 2000) devoted to the intonation of Polish spontaneous speech, providing methodological foundation and sharing his practical experience in the annotation of intonation which proved immensely helpful to the young team. Jassem was asked to re-test his model and compare it with spontaneous, conversational Polish. The analysis was focused on the nuclear melody. Initially, it was decided not to exclude any potential configuration of five tones (xL, L, M, H, xH) so as the phonology of Polish intonation could be re-constructed from scratch on the basis of a new corpus of recordings. Selected material (ca. 1,100 phrases) was annotated by four phoneticians who, according to Jassems suggestions, held regular meetings for in-depth discussion of each annotation step, together decided on the final shape of annotation (choice of labels), extensively used phonetic and sound editing software as auxiliary tools. Later, the project team developed an approach in which, on the phonetic level, the nuclear syllable, the subsequent as well as the preceding one were annotated for their height using up to four levels. Capitalics were used for representing the tones on the nuclear syllable while lower case letters for the tones on the neighbouring ones. That was partly inspired by Grabe's system (Grabe & Post 2002) who, in the IViE project, annotated her recordings not only using a regular ToBI-based procedure but also on the phonetic level, in order to be able to track and analyse intonational variation among selected British dialects. Jassem also contributed to the Pol'n'Asia project, devoted to comparative studies of intonation in task-oriented dialogues (e.g. Karpiński & Szalkowska-Kim, 2012). Although Jassem's health was deteriorating at that time, he joined several discussions on the potential frameworks of comparative intonation studies and on intonation in tonal languages that were also included in the project (Vietnamese, Thai).

Jassem's work on intonation joined the worlds of close, tedious listening and high-tech instrumental analysis, the in-depth exploration of individual utterances and large corpus-based studies, a detail-focused bottom-up approach and a wide but insightful, theory-anchored view to many aspects of prosody. His modelling of speech prosody, elegantly

distinguishing between the domains of rhythm and intonation, simultaneously shows their interaction. What he had to offer to his collaborators was not only extensive knowledge but also countless "tricks of the trade", immensely helpful in prosodic research. Jassem's influence is not limited to specific findings, models and their adjustments, but it encompasses a peculiar, individual way of thinking and experience-based research intuition.

Acknowledgement

References

Bailly, Gerard & Holm, Bleike, 2005. SFC: A trainable prosodic model. *Speech Communications*, 45, 364-384.
Campione, Estelle; Hirst, Daniel & Véronis, Jean. Automatic stylisation and symbolic coding of F0: implementations of the INTSINT model. *Intonation. Research and Applications, Dordrecht: Kluwer*, 2000.
Cruttenden, Alan 1981. Falls and rises: meanings and universals. *Journal of Linguistics* 17, 77-91.
Cruttenden, Alan 1997. *Intonation* (2nd ed.). Cambridge: Cambridge University Press.
Crystal, David 1969. *Prosodic Systems and Intonation in English*. Cambridge: Cambridge
Cutler, Ann; Mehler, Jacques; Norris, D.G. & Segui Juan 1986. The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language* 25, 385400.
Demenko, Grażyna 1999. *Analiza cech suprasegmentalnych języka polskiego na potrzeby technologii mowy*. Poznań: Wydawnictwo Naukowe UAM.
Demenko, Grażyna; Jassem, Wiktor & Krzyśko Mirosław. 1987. Classification of basic fundamental frequency patterns using discriminant functions. *Phonetica* 44, 245-357.
Demenko, Grażyna & Jassem, Wiktor, 1999. Modelling intonational phrase structure with artificial neural networks. Proceedings of EuroSpeech: Sixth European Conference on Speech Communication and Technology.
Fant, Gunnar 1960. *Acoustic Theory of Speech Production*, Mouton.
Fant, Gunnar 1962. "Descriptive analysis of the acoustic aspects of speech", *Logos* 5, 3-17
Fant, Gunnar 1967. "Auditory patterns of speech", in *Models for the Perception of Speech and Visual Form*, ed. W. Wathen-Dunn, Cambridge, Mass.: MIT Press.
Fox, Anthony 2000. Prosodic Features and Prosodic Structure: The Phonology of Suprasegmentals. Oxford: OUP.
Francuzik, Katarzyna; Karpiński, Maciej; Kleśta, Janusz & Szalkowska, Emilia 2004. Nuclear Melody in Polish Semi-spontaneous and Read Speech. Evidence from the Polish Intonational Database *PoInt*. *Studia Phonetica Posnaniensia*, 97-128.
Fujisaki, Hiroya 1983. Dynamic characteristics of voice fundamental frequency in speech and singing. The Production of Speech (pp. 39-55). Ed. P.F. MacNeilage: New York, NY: Springer.
Goldsmith, John A. 1981. English as a Tone Language. In D. Goyvaerts, ed. *Phonology in the 1980's*. Ghent: Story-Scientia. Circulated in 1974.
Grabe, Esther & Post, Brechtje 2002. The transcribed IViE corpus. University of Oxford Phonetics Laboratory.
Gussenhoven, Carlos 2004. The Phonology of Tone and Intonation. Cambridge: CUP.
Hirst, Daniel J. 2001. Automatic analysis of prosody for multilingual speech corpora. In: Keller, E., Bailly, G., Monaghan, A., Terken, J., Huckvale, M. (Eds.), Improvements in Speech Synthesis. John Wiley, London, 320-327.
Hirst, Daniel 2005. Form and function in the representation of speech prosody. *Speech Communication*, 46, pp. 334-347.

Hirst, Daniel; Di Cristo, Albert & Espesser, Robert 2000. Levels of representation and levels of analysis for the description of intonation systems [In:] Merle Horne (Ed.) Prosody: Theory and Experiment. Kluwer Academic Press.

Hirst, Daniel 2012. Empirical models of tone, rhythm and intonation for the analysis of speech prosody, *Speech and Language Technology*, vol. 13-14, pp. 23-33.

Jassem, Wiktor 1959. The Phonology of Polish Stress. *Word*, 15:2, 252-269, DOI:10.1080/00437956.1959.11 659698

Jassem, Wiktor 1962. Pitch as a correlate of Swedish pitch accent. Quarterly Status and Progress Report, Speech Transmission Laboratory, Royal InstituteTechnology, Stockholm, 1/1962, 4-14.

Jassem, Wiktor 1963. Experiment in pitch accent perception using synthetic speech. Quarterly Progress and Status Report, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm 4/1963, 10-12.

Jassem, Wiktor 1971. Pitch and compass of the speaking voice. *Journal of the International Phonetic Association*, 1, 50-68.

Jassem, Wiktor 1972a. Statistical parameters of average short-term F0 values as personal voice charaterics. In Acta Universitatis Carolinae, *Phonetica Pragensia* 3. Praha, 123-126.

Jassem, Wiktor 1972b. The question-phrase fall-rise in British English. In A. Waldman, ed., Papers in Phonetics and Linguistics, In Memory of Pierre Delattre. Mouton: The Hague, 241-252.

Jassem, Wiktor 1975. Normalization of F0 curves. In: G. Fant and M.A.A. Tatham eds. Auditory Analysis and Perception of Speech. London: Academic Press, 523-530.

Jassem, Wiktor 1978. On the distributional analysis of pitch phenomena. *Language and Speech*, 21, 362-372.

Jassem, Wiktor 1987a. Computer-assisted classification of basic Polish intonations. In Proceedings of the XI international Congress of Phonetic Sciences, vol 6. Tallin, 2219-2222.

Jassem, Wiktor 1987b. Off-line automatic recognition of basic F0 curves. In European Conference on Speech Technology, vol 1. Edinburgh, 219-222.

Jassem, Wiktor 1987c. Structural units of intonation in the acoustical speech signal. In U.W. Drechsler. S. Luschotzki, O. E. Pfeiffer and J. E. Rennison (Eds), *Phonologica 1984*. London – New York – Rochelle: CUP.

Jassem, Wiktor 1994. Linguistic function(s) of intonation in Standard British English. In B. Halford and H. Pilch eds. Intonation. Tübingen: Gunter Narr Verlag, 89-95.

Jassem, Wiktor 1996a. A quantitative analysis of Standard British English nuclear tones. *Journal of Quantitative Linguistics* 3, 3, 229-243.

Jassem, Wiktor 1996b. Intonacje rdzenne w dialogu angielskim. Analiza akustyczna i statystyczna. In Cz. Basztura and K. Dobrogowska, eds. Podstawowe założenia fonetyczne i techniczne tłumaczenia różnojęzycznego dialogu w czasie rzeczywistym. Poznań: Polskie Towarzystwo Fonetyczne, 85-104.

Jassem, Wiktor 1999. English Stress, Accent and Intonation revisited. *Speech and Language Technology*, 3, 33-50.

Jassem, Wiktor 2002. Classification and organization of data in intonation research. In A. Braun and R. Masthoff eds. Phonetics and its applications. Festschrift for Jens-Peter Köster on the Occasion of his 60th Birthday. Wiesbaden, Stuttgart: Franz Steinet Verlag. 289-297.

Jassem, Wiktor and Demenko, Grażyna 1986. On extracting linguistic information from F0 traces. In C. Johns-Lewis [Eds.], Studies in of Intonation in Discourse. London: Croom Helm. 1-17.

Jassem, Wiktor & Demenko, Grażyna. 1989. Zależność przebiegu parametru F0 od długości frazy i dźwięczności segmentalnej (The effect of phrase duration and segmental voicing on the shape of the F0 curve). Prace IPPT PAN 29/1989.

Jassem, Wiktor & Demenko Grażyna 1997. Fonetyczno-gramatyczna spójność frazy. In Speech and Language Technology. Poznań: Polskie Towarzystwo Fonetyczne, 125-140.

Jassem, Wiktor & Dobrogowska, Katarzyna 1974. Inwarianty w przebiegach parametru F0 (Invariant F0 curves). *Biuletyn Polskiego Towarzystwa Językoznawczego* XXXII, 159-171.

Jassem, Wiktor & van Dommeln, Wim 1990. Perception of Polish accent in a re-synthesized speech signal. *Archives of Acoustics* 15, 325-348.

Jassem, Wiktor; Morton, John, & Steffen-Batóg, Maria. 1968. The perception of stress in synthetic speechlike stimuli by Polish listeners. In W. Jassem,ed., Speech Analysis and Synthesis 1. Warszawa: Państwowe Wydawnictwo Naukowe, 289-308.

Jassem, Wiktor; Steffen-Batóg, Maria & Czajka, S. 1973. Statistical characteristics of short term average F0 distributions as personal voice features. In W. Jasssem, ed., Speech Analysis and Synthesis 3, Warszawa: Państwowe Wydawnictwo Naukowe, 209-225.

Jones, Daniel 1956. Outline of English Phonetics. Cambridge: Heffer.

Karpiński, Maciej; Jarmołowicz-Nowikow, Ewa & Malisz, Zofia 2008. Aspects of Gestural and Prosodic Structure of Multimodal Utterances in Polish Task-oriented Dialogues. In: G. Demenko, K. Jassem, St. Szpakowicz (Eds.) Speech and Language Technology, vol. XI, 113-122.

Karpiński, Maciej & Szalkowska-Kim, Emilia 2012. On Intonation of Korean and Polish Questions. *Speech and Language Technology*, vol. 14/15, 137-148.

Karpiński, Maciej 2012. The Boundaries of Language: Dealing with Paralinguistic Features. *Lingua Posnaniensis*, vol. LIV (2)/2012. PL ISSN 0079-4740, ISBN 978-83-7654-252-2, 37-54.

Karpiński, Maciej & Kleśta, Janusz 2001. The Project of Intonational Database for the Polish Language. In St. Puppel, G. Demenko (Eds.), *Prosody 2000*. Poznań: Faculty of Modern Languages and Literature, UAM.

Kingdon, Robert 1966. The groundwork of English intonation. London: Longmans.

Kochanski, Greg P. & Shih, Chillin 2000. STEM-ML: Language independent prosody description. Proceedings of the International Conference on Spoken Language Processing, Beijing, China, 239-242.

Kohler, Klaus J. 1995. The Kiel intonation model (KIM), its implementation in TTS synthesis, and its application to the study of spontaneous speech, http://www.ipds.uni-kiel.de/kjk/forschung/kim.en.html, based on a paper presented at the ATR Workshop on Computational Modelling of Prosody for Spontaneous Speech Processing, Kyoto/Japan, 12-14 April 1995

Ladd, Dwight R. 1996. *Intonational Phonology*. Cambridge University Press.

Ladd, Dwight R. 1983. Levels vs. Configurations, Revisited. In: F.B. Agard, G. Kelley A. Makkai, V.B. Makkai (Eds.) Essays in Honor of Charles F. Hockett. Leiden: E.J.Brill, 49-59, ISBN 90 04 07039 7.

Liberman, A.M., Delattre, P.C., & Cooper, F.S. 1952. The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *American Journal of Psychology*, 65, 497-516.

Liberman, Alvin M. 1957. Some results of research on speech perception. *The Journal of the Acoustical Society of America*, 29.1, 117-123.

Linell, Per 2004. *The Written Language Bias in Linguistics: Its Nature, Origins and Transformations*. New York: Routledge.

Mehler, Jacques 1981. The role of syllables in speech processing. *Philosophical Transactions of the Royal Society*, 295, 333-352.

Mehler, Jacques; Dommergues, Jean Yves; Frauenfelder, Uli & Segui, Juan 1981. The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior* 20, 298-305.

McNeill, D., & Lindig, K. 1973. The perceptual reality of phonemes, syllables, words, and sentences. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 419-430.

d'Alessandro, Christophe & Mertens, Piet 1995. Automatic pitch contour stylization using a model of tonal perception, Computer Speech & Language, Volume 9, Issue 3, July 1995, pp. 257-288, ISSN 0885-2308, http://dx.doi.org/10.1006/csla.1995.0013.

Mertens, Piet 2004. The Prosogram: Semi-automatic transcription of prosody based on a tonal perception model. Speech Prosody Conference Proceedings, Nara, Japan.

O'Connor, J.D. & Arnold, G.F. 1961. *Intonation of Colloquial English*, London: Longman.

Pierrehumbert, Jannet B. 1980. *The Phonology and Phonetics of English Intonation*. PhD dissertation, MIT. [IULC edition, 1987].

Port, Robert F. 2008. All is prosody: Phones and phonemes are the ghosts of letters. In *Proc 4th Speech Prosody Conference,* pp. 7-16.

Rossi, Mario 2000. Intonation: Past, Present, and Future. In Antonis Botinis (Ed.) Intonation: analysis, Modelling and Technology. Kluwer Academic Press.

Savin, H. B., & Bever, T. G. 2007. The nonperceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior*, 9, 295-302.

Steffen-Batóg, Maria, Jassem Wiktor & H. Gruszka-Kościelak. 1970. Statistical distribution of short-term F0 values as a personal voice characteristic. In W. Jassem, (Ed.), Speech Analysis and Synthesis 2. Warszawa: Państwowe Wydawnictwo Naukowe, 197-206.

Steffen-Batóg, Maria 1966. Versuch einer strukturellen Analyse der polnischen Aussagemelodie. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforchung*, 19, (6), 397-440.

Steffen-Batóg, Maria 1996. *Struktura przebiegu melodii polskiego języka ogólnego* (The Structure of Speech Melody Sequences in Standard Polish). Poznań: Wydawnictwo SORUS.

't Hart, Johan; Collier, Rene, & Cohen, Antonie 1990. A perceptual study of intonation: An experimental-phonetic approach to speech melody. Cambridge, UK: CUP.

Vaissière, Jacqueline 2005. Perception of intonation. in D. B. Pisoni and R. E. Remez (Eds.) Handbook of
     Speech Perception. Oxford: Blackwell.
van Santen, Jan & Möbius, Berndt 1997. Modeling pitch accent curves. Intonation: Theory, Models and Ap-
     plications (Proceedings of the ESCA Workshop, 321-324, Botinis, A., Kouroupetroglou, G., & Carayiannis,
     G.(Eds.) ESCA & the University of Athens.
Wagner, Agnieszka 2008. *A comprehensive model of intonation for application in speech synthesis*. Unpub-
     lished PhD Thesis defended at Adam Mickiewicz University in Poznań, Poland.
Wypych, Mikołaj 2011. Układ rozpoznający struktury intonacyjne w sygnale mowy. A doctoral thesis de-
     fended in 2012 at the Institute of Fundamental Technological Research, Polish Academy of Sciences.
     Unpublished.