Microprosodic Aspects of Vowel Dynamics – an Acoustic Study of French, English and Czech

Tomáš Duběda* & Eric Keller**

*Institute of Phonetics, Charles University in Prague, nám. J. Palacha 2,

116 38 Prague, Czech Republic, dubeda@ff.cuni.cz

**LAIP – IMM, Université de Lausanne, 1015 Lausanne, Switzerland,

eric.keller@imm.unil.ch

Abstract

The article gives an account of microdynamic behaviour of vowels in three languages. The microdynamic profile, defined as a set of ten normalized values measured at ten equidistant points for each sound, shows systematic relations with the articulatory properties of the sound and its immediate context, but also with its duration and macrointensity. No systematic correlation with fundamental frequency was found. Apart from its theoretical value, research on microintensity can find a successful outlet in speech synthesis.

1. Introduction

Microprosody is understood as a complex of local prosodic variations due to articulatory processes (Silverman 1986, Di Cristo 1986). It cannot be directly controlled by will nor convey primary phonological information. It interacts with the segmental level, as well as with macroprosody in various ways: for instance, Kohler (1987, p. 80) reports that microintensity helps German listeners in discriminating fortis and lenis stop consonants, and House (1989) identifies conflicts between macromelodic peaks and microintonation. In casual usage, microprosody may be confounded with microintonation, which is only a part of it. Microprosody should be understood as covering at least two prosodic parameters, namely fundamental frequency and intensity. Duration as a microprosodic parameter makes little sense, and the term ˝microduration' could apply at the very most to the study of different parts of sounds, such as the closure and the burst in plosives. Unlike microintonation, which has been the topic of a certain number of studies (Di Cristo 1985, Kohler 1987, Monaghan 1992, Whalen et al. 1990), microintensity has not been granted extensive attention so far.

We define microintensity (the corresponding adjective being ˝microdynamic' in our usage) as energetic properties of speech linked to sound articulation. By contrast, the term ˝macrointensity' is reserved to suprasegmental changes in speech energy, due mainly to the respiratory and phonatory activity, not to articulatory processes. A typical macroprosodic instance of intensity is the intensity maximum or mean of a sound, frequently used in prosodic description, when speaking, for example, about intensity decrease at the end of prosodic constituents; the location of the maximum within the sound and the evolution of the intensity curve over the sound would then be a microprosodic issue.

The evolution of the intensity curve throughout an ideal vowel follows a rising-descending profile, whose first (pre-peak) part corresponds to an opening gesture, and the second (post-peak) part to a closing gesture. This default form is well observable especially in isolated vowels or vowels between stops. The nature of adjacent sounds has an impact not only on the overall intensity of the vowel (House & Fairbanks 1974), but also on its microdynamics, especially in its edge parts. Therefore, microdynamic changes are an integral part of coarticulatory processes. For instance, depending on the degree of coarticulation between a voiced fricative and a following vowel, we can observe different shapes of the dynamic line at the sound transitions. Similarly, most instances of gesture overlapping between vowels and neighbouring sounds, be it in the domain of lips, tongue tip, blade, body or root, velum or larynx (Farnetani 1997), have an indirect impact on the microdynamic profile of the vowel. Even if the present research is not focused on coarticulation, one of its objectives is to show how the microdynamic changes are affected by intrinsic properties of the sound (inherent to the sound itself) on the one hand, and by its co-intrinsic properties (resulting from gesture overlapping) on the other.

The opening-closing gesture in vowels constitutes a fundamental component of the syllabic structure of speech, which is language-universal (Hála 1966). Microdynamic properties of the syllable's constituents may provide useful evidence for the perception of segmental contrasts underlying the syllable or syllable boundaries, and, more particularly, help to explain why the opening gesture (transition from a consonant to a vowel) is a universally dominant syllabic feature (MacNeilage and Davis 2003, p. 383).

Fig. 1

2. Method

2.1 Objective

Our analysis is centred on microprosodic behaviour of intensity, sometimes considered to be the *most neglected* (Rossi 1981, p. 46), and sometimes *the most overrated* (Bolinger 1978, p. 486) prosodic parameter. On a speech corpus covering three languages, we shall first consider the possibility of extracting microintensity from the measured values, choosing suitable descriptors and formalizing them. Thereafter, we shall relate these values to intrinsic features of the sound in question and to several contextual and macroprosodic parameters, with the aim of providing a comprehensive account of the microdynamic behaviour of vocalic segments, on a cross-linguistic basis.

Our study is limited to vowels. This category of sounds is of particular interest for their mutual comparability (despite differing intrinsic duration and macrointensity), their utmost importance in prosody (Rossi et al. 1981) and their constant voicing. The study, essentially acoustic in nature, deliberately leaves out details on articulatory aspects influencing microintensity.

2.2 Corpora

We made use of three corpora of approximately the same size (3:30–3:50 minutes of speech including pauses, 2382–2725 sounds per language, out of which 1034–1040 vowels). The French speaker (adult male, non-professional, recorded in a studio at 16 kHz, previously

trained to reduce head movements, so as to exclude external variability in intensity; average f_0 130 Hz) is representative of non-dialectal Swiss French, with only a very small perceptible difference to standard French. The English speaker (adult male, professional BBC speaker, analogue studio recording digitized to 16 kHz; average f_0 106 Hz) is originally from New Zealand, but lives in England and uses RP pronunciation (speaker BP of the MARSEC corpus, section b – Knowles et al. 1996). The Czech speaker (adult male, non-professional, recorded in a soundproof booth at 48 kHz, downsampled to 16 kHz, previously trained to reduce head movements, so as to exclude external variability in intensity; average f_0 100 Hz) comes from Central Bohemia and uses standard Czech pronunciation. All vowels of all the languages are represented in the corpus, except for the Czech marginal diphthong [EY]. The segmentation and the analysis of the corpora were performed specifically for the purposes of this experiment, so as to keep the methodology as comparable as possible. On the other hand, this practical aspect also led us to the decision to keep only one speaker per language. All conclusions drawn must therefore be understood as related to the speakers analyzed.

The texts are of informative character, taken partly from newspapers and partly from broadcasting programmes and read at a medium speech rate (articulation rate 5.25–5.50 syllables/second). This type of speech has been chosen with regard to its representativeness and its potential use of the results in speech synthesis. The choice of languages was not only dictated by the experience of the authors, but it also offers a good sampling of European languages (one Romance language, one Germanic and one Slavonic) providing an interesting cross-section of suprasegmental typologies:

French: bound non-lexical stress (at least often in default readings) on the last (full) syllable of a stress group, less regular secondary stress near the beginning of a stress group;

little vowel reduction; 14 vocalic phonemes (excluding [A] and [ɹ̩)]); phonotactic and prosodic vowel length.

British English: unbound lexical stress; systematic vowel reduction in unstressed syllables; 12 vocalic and 8 diphthongal phonemes; intrinsic, phonotactic and prosodic vowel length.

Czech: bound non-lexical stress on the first syllable of a stress group; almost no vowel reduction; 10 vocalic phonemes falling into two length categories (5 short and 5 long), 3 (rare) diphthongal phonemes; phonological and prosodic vowel length.

2.3 Analysis

After semi-automatic segmentation and manual adjustment with tools developed at LAIP Lausanne, the sounds were transcribed in broad SAMPA notation according to their perceptual value rather than to their normative form, including the notation of perceived stresses (including initial, secondary stresses in French). The term ˈperceived stressˈ is meant as a perceptual amalgam where any of the three prosodic parameters (duration, intensity and pitch) may play a role, according to the accentual behaviour of the language (cf. section 3.3). The global perception of the transcriber was decisive, and no distinctions or restrictions in terms of e. g. ˈpitch accentˈ or ˈlength prominenceˈ were made. The English part of the corpus was labelled by an external transcriber, and later checked by the first co-author of this article, who also labelled the Czech and the French parts. Priority was given to labelling by a single person, speaking all the three languages and phonetically trained, even if native only in Czech.

The segment boundaries were located with respect to spectrum, wave form, voicing and intensity. Where two or more features entered in conflict (e. g. the friction of a voiceless fricative and the periodicity of the following vowel), the solution adopted aimed for a reasonable compromise between them. Diphthongs in Czech (about 0,4% of all speech sounds in the present corpus) and English (about 6%) were categorized as single segments.

Thereafter, a detailed prosodic analysis by means of the Praat software was performed. For each vowel sound, 10 equidistant points (at 5, 15, 25... 95% of sound duration) were taken as the basis for intensity and f_0 mapping, so as to grant a good accuracy covering fine intensity changes. Temporal normalization using ten equidistant points may blur certain instances of dynamic non-linearity with respect to the time axis, but it offers good comparability between sounds of different durations. Besides, the influence of sound duration on microdynamic profiles is studied as a separate kind of interaction in section 3.4; this prevents us from excluding the absolute time dimension from our observations.

The intensity was analyzed without any previous normalization, with a time step of 2 ms, the f_0 was extracted at 1 kHz by a two-pass autocorrelation algorithm developed at LAIP, Lausanne. The algorithm determines negative-going peaks with great precision and calculates local f_0 values from this information.

To render the microdynamic variations of intensity comparable throughout the corpus, it was essential to normalize the measured values so as to exclude (or at least strongly reduce) the contribution of macrointensity. The average intensity of the sound was taken as reference macrodynamic value, and the ten measured values in the sound were then normalized with respect to this average. We considered two different ways of normalizing: the multiplicative

approach (i. e. expressing microintensity values as ratios of the macrodynamic average over the sound), and the additive approach (i. e. expressing microintensity values as distances in dB from the macrodynamic average). The first method was supported by the fact that the generated results provide better microdynamic comparability of different realizations of the same sound: for instance, the microdynamic profiles of all occurrences of an [a] vowel in a language had a lower standard deviation with the multiplicative than with the additive approach. On the other hand, the idea of an additive normalization seemed more plausible because the decibel itself is already a logarithmic unit. Finally, we adopted the latter, i. e. additive, method. Nevertheless, the comparison of the microdynamic profiles showed that all tendencies described in this article are well detectable with both methods, and that the curves are in fact not very different (a discussion of multiplicative vs. additive approaches to microprosody can be found e. g. in Monaghan (1992)).

To sum up, the ten values were expressed as dB distances from the mean intensity measured within the sound. These values are called ʻrelative dynamic excursionsʼ in the figures, and they include both negative and positive values. The output of the analysis is, for each sound, a ten-point curve which should be interpreted with reference to its two-fold normalization: the temporal one and the dynamic one. It is on the basis of such normalized micro-dynamic curves that we attempt to describe interactions of intensity and other prosodic and linguistic factors.

The research, centred around the theme of microdynamics, also addresses the problem of microintonation. The determination of this parameter was based on the assumption that microintonation consists of local f_0 excursions which are superimposed on larger macromelodic changes (Di Cristo & Hirst 2002, p. 316). The separation of the two

components was based on the methodology proposed in Monaghan 1992, with the exception that smoothing was obtained by calculating a moving average instead of fitting a spline. The detailed curve was then divided by the smoothed one, with the micromelodic residuum as result (values higher than 1 denote micromelodic rises, those below 1 micromelodic falls). The multiplicative approach in frequency grants better perceptual adequacy of the obtained values (Monaghan 1992). The degree of smoothing was selected so that the standard deviation of the obtained set of values does not exceed 2% of their mean value, which corresponds e. g. to an average micromelodic deviation of 2.5 Hz for a neutral voice register of 125 Hz.

3. Results and Discussion

3.1 Intrinsic microintensity of vowels

The typical microdynamic profile of a vowel, as defined in section 2.3, is a smooth convex curve with the beginning mostly higher than the end and the maximum generally shifted to the left. Fig. 2 illustrates average microdynamic profiles of English sounds [I] and [A⁻]. The standard deviation of the values (Fig. 3) calculated for each of the ten equidistant points increases towards both ends, which is due to variations in phonetic context. Correlates of this general microdynamic profile are both articulatory and distributional: the opening gesture is faster than the closure, the onset is usually more tense than the coda (Hála 1966), and the end of the vowel, especially in an open syllable, can bear the marks of prosodic finality (Duez 1987, p. 26). Nasal vowels in French show generally a flatter and less smooth shape, with a plateau-like tendency in the middle part, due to the constant shape of the nasal cavity, which is coupled to the oral cavity.

Fig. 2

Fig. 3

3.2 Position of the microintensity peak

The articulation of a vowel can be roughly described as an opening-closing gesture. The acoustic energy released in this gesture should show a corresponding time-course. As we have observed for English [A͡], the peak is often left-asymmetric, following an onset which is steeper than the tail. Fig. 4 shows the distribution of vocalic intensity peaks throughout the corpus, plotting the frequency with which any of the ten points coincides with the intensity peak.

Fig. 4.

The first information provided by the graphs is that the intensity peak can fall on any of the ten points, though with very different probabilities. The point where the intensity maximum occurs with the highest frequency is point 5 for English and Czech, and 4 for French. In both cases, this point lies in the first half of the vowel, near its middle point. The evolution of the English curve is rather different from that of the two others, namely in the first part, where it is more or less steady: the peak can be thus located with about the same probability throughout the first two thirds of the vowel, which can be explained mainly by the high occurrence of diphthongs in English (cf. Fig. 6). If we isolate long vowels or diphthongs and analyze them separately, we can see that the intensity peak falls generally closer to the left edge of the vowel. In French, the data for long vowels are more disparate due to a smaller number of analyzed items, but even there, the distribution curve is more skewed to the left.

The distributional density of peaks, falling towards both ends of the curve, rises again in close vicinity to the adjacent sounds. A contextual analysis of vowels having an intensity peak at

these extreme points showed that both left-eccentric and right-eccentric peaks are mostly triggered by an immediately preceding or following vowel, diphthong, semi-vowel, nasal or approximant (all these categories having in common high sonority). Initial peaks are typical of front unrounded or nasal vowels in French, diphthongs, some long vowels or [↔] in English, and long vowels in Czech. Final peaks occur mostly in high and nasal vowels in French, high vowels, [↔] or [A⁻] in English, and long vowels, [Y] or [a] in Czech. Context seems to be a more powerful factor than the inherent character of the sound, since the set of sounds that favour adjacent edge peaks is quite coherent, and corresponds to sounds with relatively high intensity. Among the features of the sound itself, length is the most frequent triggering factor of initial or final intensity peaks.

A certain percentage of the studied vowels differ from the typical convex signal shape insomuch as they show a secondary microintensity peak, with a concave space separating it from the primary peak (intensity maximum for the given vowel), cf. Fig. 1. The percentage of vowels with secondary intensity peaks is given in Tab. I, together with the prevailing characteristics of sounds belonging to this category. It is important to say at this point that the result of the analysis obviously depends on the number of points analyzed (10 per speech sound in the present case, corresponding to slightly different time intervals in shorter and longer vowels), as well as on the time step chosen for intensity computing (2 ms).

Table I.

Secondary peaks to the right rather than the left of the main peak predominate in all three languages; this is in congruence with the typical left-eccentric position of the primary peak and with the less steep dynamic evolution at the vowel tail. French seems to be the language

with the most ʻbroken' microintensity, in which the nasal vowels, with their tendency to exhibit a medial dynamic plateau, certainly play a key role.

Three categories of sounds are particularly likely to exhibit a secondary peak: long, closed and (in French only) nasal vowels, their common denominator being the relatively elongated and flattened shape of the microdynamic profile.

3.3 Influence of stress on microintensity

Before analyzing the links between stress and microintensity of vowels, a rough sketch of macrodynamic and durational correlates of stress in French, English and Czech should be given.

For French, it has been demonstrated that syllables perceived as bearing final stress are always longer and mostly less intense than other syllables (Delattre 1966, p. 67, Duez 1987, p. 129). For such vowels, one can thus expect profiles resembling those of long vowels. Initial stress, emphatic in origin but currently gaining ground in other than emphatic contexts, is ascribed mostly melodic and (macro)dynamic correlates (Lacheret-Dujour & Beaugendre 1998, p. 41).

In English, suprasegmental correlates of stress are closely tied to the segmental ones, which means that the distribution of vowels is different in stressed and unstressed syllables (Lehiste & Peterson 1967); by virtue of reduction principles, we can expect, on the average, longer duration and higher intensity of stressed vowels (Pamies 1996).

In Czech, duration as a stress-marking parameter is inhibited by the phonological role of vocalic length (Palková 1995, p. 170), and stressed syllables (always word-initial) should thus not differ significantly from syllables in the rest of the utterance with respect to this point. A recent study (Duběda 2002, p. 348) has shown that positive deviations in intensity on stressed syllables in Czech, often overestimated in the past, are mostly limited to emphatic stress; otherwise, the macrointensity is the same or even lower than in the following syllable.

Our data showed the expected tendencies in French, English and Czech (Fig. 5). In addition to that, we can see that syllables bearing initial stress in French are more easily grouped with unstressed than with stressed vowels with respect to their microdynamic form.

Fig. 5

We verified the statistical significance of pairs of average values at the ten durational points. The performed t-tests (significance threshold 0.05) were followed by a Bonferroni correction for series of values. As for the difference between stressed and unstressed vowels, the number of points where the two curves satisfy the Bonferroni condition is 9 for French, 10 for English and 1 for Czech. In French, the difference between initial-stressed and unstressed vowels does not satisfy the Bonferroni condition in any point.

Inasmuch as the pairs of curves have at least one crossing point, it is nearly impossible to satisfy the Bonferroni condition in all points. Therefore, we can consider the results describing microdynamic differences between stressed and unstressed vowels as highly significant for English and French, and almost insignificant in Czech. Initial-stressed vowels in French seem to have no microdynamic specificities as compared with unstressed vowels.

The observed behaviour seems to be in congruence with macroprosodic properties of the compared languages: Czech stress has very weak prosodic correlates (Duběda 2003), French durational variables have a clear impact on the microdynamics, and English combines temporal changes with articulatory salience of the stressed syllables.

3.4 Interaction of microintensity and duration

A further point of interest was the difference between short and long vowels. In Czech, where vowel length is phonological, and even marked in spelling, the categorization was simple. In English, the vowels [ɪ͡, ʊ͡, □͡, ɛ͡, A͡] were taken as long, and the diphthongs were treated separately. In French, where length is mostly phonotactic and prosodic (it had not been marked in the labelling), the differentiation was made according to Léon (1992, p. 98): 'long' vowels were defined as stressed and followed by a lengthening consonant [®, ϖ, ζ, Z] at the same time, e. g. *futur* [ϕψ∪τψ͡®], *progressive* [π®□γ®ε∪σɪ͡ϖ], or as [o] in a stressed closed syllable, e. g. *autre* [∪o͡τ®]. French nasal vowels were not included in the category of long vowels, because their greater duration is mostly intrinsic and they are absent from English and Czech. When reflecting on the relation between microintensity and vowel length, one should bear in mind that the distinction short/long is based on different criteria in the different languages, and that the two types of vowels also have very different frequency (see Fig. 6).

Long vowels, whose macrodynamic behaviour is different from short vowels also seem to have different microdynamic profiles. According to our data for French and Czech, long vowels are less intense in terms of both average and maximum intensity, mainly by virtue of the law of energy summation – Pamies 1996; English long vowels are more intense that short

vowels, mostly because of their co-occurrence with rhythmically strong syllables. Fig. 6 shows the average microdynamic values separately for the two classes (English diphthongs were analyzed as a separate category).

Fig. 6

As in the preceding section, we verified the statistical significance of the tendencies presented in the graph by means of a t-test (significance threshold 0.05) followed by a Bonferroni correction. Each time, the t-test was calculated for ten pairs of microdynamic excursions, corresponding to ten equidistant points where the values were analyzed. As for the difference between short and long vowels, the number of points that satisfy the Bonferroni condition is 7 for French, 8 for English, and 6 for Czech. The difference between long vowels and diphthongs in English satisfies this condition at 1 point only.

The microdynamic differences between short and long vowels show a step-by-step increase in the order Czech – French – English. Taking into account the fact that the compared curves cross each other at least once (cf. Fig. 6), the default significance threshold should lie somewhere near the value 7. Diphthongs in English show only a small difference when contrasted with long vowels, but they are necessarily more distinct from short vowels than long vowels.

The inspection of Fig. 6 shows that for French, the energy peak of long vowels is shifted to the left; the wider range of the curve only accounts for greater deviations from the mean value, not for greater energy as compared to short vowels. In English and Czech, long vowels also show a steeper onset and an intensity break after the second point, which has the effect of

flattening the contour. This plateau-like tendency is stronger in Czech. English diphthongs show a more pronounced form of the profile peculiar to long vowels. Furthermore, French and English on the one hand and Czech on the other are differentiated by the relative span of the profiles: in the first two languages, the relative dynamic deviations of long vowels are greater than those of short vowels, whereas Czech long vowels show relatively smaller deviations from their average intensity. The difference can probably be related to the flattened shape of the profile, and thus a more 'compact' form, which makes the relative dynamic excursions smaller.

As explained in section 2.3, the dynamic evolution of vowels is studied on a relative time axis, which is a product of temporal normalization. However, it is interesting at this point to turn our attention to the absolute time dimension of the obtained data. For instance, the relative positions of microdynamic peaks in Fig. 6, combined with the respective average durations of the concerned vowel classes, result in the following absolute peak positions, expressed as distances in ms from the beginning of the sound (Table II.):

Table II. Absolute peak positions in short vowels, long vowels and diphthongs (NA = does not apply).

	French	English	Czech
Short vowels	41 ms	34 ms	39 ms
Long vowels	58 ms	37 ms	41 ms
Diphthongs	NA	33 ms	NA

In absolute terms, the microdynamic peak is always reached later in long vowels than in short vowels. In the case of Czech and English, however, this difference is far smaller than the one

which is displayed in Fig. 6. This means that the opening gesture requires about the same time to achieve its target, and that temporal changes affect the vowel tail (post-peak part) much strongly than the onset. In French, the temporal variability of the pre-peak part seems to be greater, which means that the peak occurs significantly later in long vowels (in absolute terms). But this peculiar difference may well be an artefact of the method, since the set of long vowels analyzed was rather small, and included, by definition (cf. section 3.4), only vowels in a stressed closed syllable. We suppose that the systematic presence of the coda in the syllables analyzed may have an influence on the position of the peak.

3.5 Interaction of microintensity and macrointensity

Apart from the temporal characteristics, the microdynamic profile can also be influenced by the macrointensity. An analogy to this relationship can be seen for instance in the interaction between global speech rate and local proportion, say, of closures and bursts in plosives: the internal temporal structure of speech sounds does not remain the same when the duration changes. In the domain of intensity, we can suppose, among other things, that a more intense vowel would require more time to achieve its dynamic maximum, all things being equal.

There are several plausible candidates for the reference macrodynamic value: we can take either the average intensity over the whole sound, or the maximum value, supposedly having a considerable perceptive impact, or the value at a given time point within the sound. Rossi (1981, p. 46), who favours the last method, has found that the ʼperceptual centre' for intensity of French vowels lies at two thirds of their duration when the fundamental frequency is rising, and at one third of the duration when f_0 is falling.

However, in order to keep the study uni-parametric, we chose the average vowel intensity as the representative value for the analysis of possible interactions between micro- and macrodynamics. The use of this value is also supported by the fact that average intensity has been set as reference for relative microdynamic excursions. Fig. 7 shows the microdynamic behaviour of vowels having different average macrointensity.

Fig. 7

While most of the vowels tend to have a uniform microdynamic shape, those with macrointensity below 65 dB exhibit forms resembling patterns identified for long vowels (cf. section 3.4). This similarity is particularly strong in French, where longer vowels, standing mostly in final syllables, also bear marks of prosodic finality, among which lower intensity. On the other hand, the data for English show a much smaller microdynamic differentiation as to the vowels' macrointensity. This is unlikely to be a coincidence, because in English vowels, length and smaller intensity show a much weaker correlation (cf. average duration of the classes given in the graph caption). Czech presents an intermediate case according to our data.

Going back to the principle of pre-peak-phase temporal rigidity (see section 3.4), we can extend it to macrointensity: in a vowel with smaller overall intensity, the dynamic peak will be reached relatively earlier, since the velocity of the opening gesture is not adapted to the vowel's macrointensity.

3.6 Interaction of microintensity and fundamental frequency

Having investigated microdynamic correlates of duration and macrointensity, we now move to the interaction between microintensity and fundamental frequency. The problem can be subdivided into two categories: interaction of microintensity and macromelody (global, suprasegmental changes in f_0) and that of microintensity and micromelody (local changes in f_0 triggered by segmental properties of speech).

In the first category, two parameters were taken as potentially having an influence on microdynamics: one relative – the direction of the f_0 change within the vowel, and one absolute – average fundamental frequency of the vowel.

The f_0 course was expressed as the difference in Hz between the highest and the lowest f_0 measured in the vowel; a negative value corresponds to a fall, a positive one to a rise. The potential perceptual incomparability of melodic intervals defined in Hz was largely compensated by the fact that the average fundamental frequencies of the three speakers were not too far from each other. The study of the interaction between this parameter and microintensity has shown that steep rises in Czech and both steep rises and falls in English correspond to a left-eccentric, long-tailed microdynamic profile, resembling the one described above for long vowels. This analogy is unlikely to be coincidental, since major melodic prominences usually occur at the ends of prosodic units, which are concurrently lengthened. In French, on the other hand, little variability was found, except for major falls. There, however, the microdynamic profile does not correspond to that of long vowels, as it would be expected; this difference could indicate that the correlation between melodic falls and lengthening is less important in French than in English.

When contrasted with the absolute f_0 values of the individual vowels, the microdynamic profiles seem to deviate from the default shape only for low frequencies in French and English, but not in Czech. The observed left-eccentric profile can be, here again, ascribed to final lengthening.

On the whole, no striking regularities were found in the relationship between macromelody on microdynamics, except for those which result from other, supposedly stronger prosodic features.

The second defined category of interactions between intonation and intensity involves micromelodic changes, which were filtered from the detailed f_0 curve as described in section 2.3. However, most of the attempts to correlate microintonation and microintensity were blocked by two important facts: First, the magnitude of micromelodic changes being very small, we cannot expect clear results if the previous step of our analysis showed that macrointonation, with its much more pronounced changes, has provided only limited evidence. Second, vocalic microintonation is mostly described as a function of consonant context (Lehiste & Peterson 1961, Silverman 1986, Kohler 1987), whose clear impact on microintensity was demonstrated above (section 3.2). Thus, the observed microdynamic changes, if any, would rather be explained by the context than by microintonation, which is in itself a consequence of context. The difficulty of formulating a testable hypothesis concerning the interaction of microintensity and microintonation led us to the decision of leaving out microintonation as a source of microdynamic variation.

4. Conclusion

The study of microdynamic profiles of vowels has provided numerous pieces of evidence which form a relatively stable framework. Within the defined methodological background, the default microdynamic profile of a vowel appears as a smooth convex curve with its highest point often shifted to the left. In certain cases (nasal vowels in French, long vowels in Czech), an intermediate plateau is a systematic constituent of the profile.

The left- or right-eccentric position of the microintensity peak corresponds to a class of conditions which are in mutual relation: the intensity maximum can be attained at a relatively earlier stage of articulation a) because it corresponds to a smaller dynamic deviation (as it is the case for less intense and close vowels); b) because the vowel is longer, and the opening reaches its maximum relatively earlier; c) because the vowel is preceded by a sound with relatively great intensity; d) because the dynamic profile of the vowel itself implies a plateau (mostly in long or nasal vowels). More rarely, a right-eccentric peak can be triggered by a) a relatively small dynamic maximum; b) a plateau; c) a following sound with relatively great intensity. It was demonstrated that pre-peak phases of Czech and English vowels in our corpus exhibit a certain temporal rigidity, i. e. they tend to have about the same duration irrespective of the vowel length; the post-peak phases, on the other hand, are more open to elongation or compression. In the French recording, this tendency exists, but it was probably obscured by the constraints relative to the definition of long vowels. The observed behaviour may lead us to the assumption that the timing of the opening phase in a vowel is planned in absolute terms, while the closing phase offers the ground necessary for temporal adaptations.

Going further in the interpretation of the findings relative to average microdynamic profiles, and more particularly to the position of the intensity peak, we may try to relate these to some universal aspects of syllabic structure: CV is considered to be the default syllabic structure, and this assumption is supported by phylogenetic (Hála 1966), ontogenetic (McNeilage 2003) and statistic (Hyman 1975, p. 188) evidence. In the same vein, the phonetic interdependence of the syllable nucleus and the following coda (both forming a unit referred to as ‘rhyme’ – Goldsmith 1990, p. 108) is tighter than the one which exists between the onset and the nucleus, and the onset is granted more independency and less variability, both phonological and phonetic, than the coda. The same kind of behaviour can be observed in vocalic microintensity: the pre-peak phase displays less temporal plasticity, while the post-peak phase is more likely to undergo contextual and prosodic adaptations. From the perspective of the syllable, the left-eccentric intensity maximum of the vowel may be explained, in part, by the tendency to keep this maximum near the middle of the syllable, knowing that syllables without coda are universally more frequent than syllables without onset.

Secondary intensity peaks can be seen as another form of the behaviour described in the paragraph above: indeed, a primary peak shifted to either edge of the vowel as well as a dynamic plateau appears to favour the occurrence of a secondary peak, which then leads to a concave profile of the vowel.

The microdynamic profile has turned out to be a good indicator of stress properties: the microdynamic impact of stress is negligible in Czech and in the case of French initial stress; French final stress reveals its strong durational correlates; English stress corresponds, on the whole, to durational and macrodynamic changes (i.e., systematic intensity differences between stressed and unstressed vowels), both of them being based on segmental constraints.

The confrontation of microintensity and macrointensity has shown a behaviour which is related to durational changes: less intense vowels are often lengthened, but in reduction languages like English, this correlation is significantly weaker. However, microdynamic specificities have only been found in vowels with low overall intensity, where the dynamic peak is reached earlier, presumably because the velocity of the opening gesture is not adapted to the overall intensity. This observation, together with the principle of pre-peak temporal rigidity, can be seen as two instances of the same general tendency.

Macrointonation, generally accepted as a parameter with dynamic correlates, has shown only little impact on microintensity. Some of the observed tendencies are better explained by other, non-intonational phenomena.

The microdynamic analysis, covering a well-defined set of possible acoustic variables, has revealed important tendencies which are of great interest both in speech description and modelling. In this paper, we have not arrived at a stage where the individual variables would be fully separated; instead, we are witnesses of a prosodically founded, multiparametric behaviour, manifested by clusters of features.

Although perceptual validation of the observed tendencies remains to be performed, informal observation shows that different microdynamic proportions in vowels can be identified by the ear, and are necessarily part of the phonetic form of speech. Despite their extraphonological status, these small-sized variations may well turn out to be indispensable in speech processing, and particularly in speech synthesis. Indeed, one of the most important issues in synthesis research nowadays is the search for naturalness (Keller 2001, p. 14). It has been

demonstrated (e. g. Monaghan 1992) that a rich microprosody can contribute to the perceptual naturalness of synthesized speech. Attempts have even been made to introduce random microintonation on a smooth intonation curve, with a result that is judged to be perceptually better than the original synthetic signal (Sorin et al. 1987).

In the domain of concatenative speech synthesis, as represented by the diphone-based models *LAIP TTS*, developed in Lausanne (Keller 1998), and *Epos*, developed in Prague (Horák 1999), we can make two predictions as to the implementation of microdynamic algorithms: The first potential improvement concerns cases of incompatibility between adjacent units at concatenation points, where diphones (or other units) taken from different contexts do not have the same amplitude. At the perception level, this introduces disagreeable acoustic changes which deteriorate the signal quality (Duběda & Machač 2003). Although smoothing can ease the perceptual transition, it would clearly be preferable if a reliable model of microintensity could be developed. The second domain where microdynamic variations may interfere is the duration control of individual segments. As it has been shown, the intensity of vowels (and other speech sounds presumably as well) does not change in a linear manner when the sound is compressed or expanded. A striking example of this can be an utterance-final, pre-pausal vowel whose intensity peak is off centre, in its left portion, and whose tail has a concave rather than a convex form (cf. Fig. 1).

The analysis described in this article opens a certain number of perspectives, both theoretical and application-driven. Prior to the implementation of any of our conclusions in speech synthesis systems, the data should be subjected to perceptual validation. Also, it is important to emphasize that all the results presented in this analysis provide information about speech

behaviour of one speaker per language. Therefore, they must be taken as authentic and based on a well-chosen corpus, but not covering variability across speakers.

Acknowledgements

The first co-author of this article thanks the Swiss Confederation for granting him a research attachment at the Lausanne University during the winter semester of 2002–2003. He is no less thankful to have been given a warm welcome at the LAIP Lausanne. This research was supported by a research grant to the second author by BBW/OFES Berne under COST 277. Thanks are also extended to Mr. L. Wiget for his manual segmentation of subject BP (English), and to the three anonymous referees for the help on our way to the final version of this article.

References

Bolinger, D. (1978). Intonation Across Languages. In: J. H. Greenberg (Ed.), *Universals of Human Language*, Vol. 2 – Phonology. Stanford: Stanford University Press.

Caelen, G. (1979). Quelques remarques sur les relations entre fréquence fondamentale et énergie chez deux locuteurs. *Séminaire Larynx et parole*. Grenoble. Offprint.

Chlumský, J. (1928). *Česká kvantita, melodie a přízvuk*. Prague: Czech Academy of Sciences and Arts. [Czech quantity, melody and stress].

Delattre, P. (1966). Accent final en français: accent d'intensité, accent de hauteur, accent de durée. In: *Studies in French and Comparative Phonetics*. London – The Hague – Paris: Mouton & Co., pp. 65–68.

Di Cristo, A. – Hirst, D. (1986). Modelling French Micromelody: Analysis and Synthesis. *Phonetica 43*, pp. 11–30.

Di Cristo, A. (1985). *De la microprosodie à l'intonosyntaxe*. Aix-en-Provence: Publications de l'Université de Provence.

Di Cristo, A. & Hirst, D. (2002). De l'acoustique à la phonologie, représentations et notations de l'intonation: une application au français. In: A. Braun & H. R. Masthoff (Eds.), *Phonetics and its Applications. Festschrift for Jens-Peter Köster on the Occasion of his 60th Birthday*. Stuttgart: Steiner.

Dolbec, J. & Rogers, S. (1996). Caractéristiques microprosodiques de durée et d'intensité en lecture et en conversation semi-dirigée. In: J. Dolbec & M. Ouellet (Eds.), *Recherches en Phonétique et en phonologie au Québec*. Québec, pp. 19–35.

Duběda, T. (2002). Structural and quantitative properties of stress units in Czech and French. In: A. Braun & H. R. Masthoff (Eds.), *Festschrift for Jens-Peter Köster on the Occasion of his 60th Birthday, Phonetics and its Applications*. Stuttgart: Steiner.

Duběda, T. (2003). De l'acoustique au conventionnel: une vue configurative et multiparamétrique de l'accent en français et en tchèque. *Language Design 5/2003*. Granada: Método Ediciones, pp. 1–10.

Duběda, T. & Machač, P. (2003). Constructing and Optimizing a Diphone Database for Czech Speech Synthesis. *Acta Universitatis Carolinae, Phonetica Pragensia* (in print)

Duez, D. (1987). *Contribution à l'étude de la structuration temporelle de la parole en français*. Thèse pour le doctorat d'Etat ès lettres et sciences humaines. Aix-en-Provence: Université de Provence, Aix-Marseille 1.

Farnetani, E. (1997). Coarticulation and connected speech processes. In: *The Handbook of Phonetic Sciences.* Hardcastle, W. J. & Laver, J. (eds.). Blackwell Publishers, pp. 371–404

Goldsmith, J. A. (1990). *Autosegmental & metrical phonology*. Oxford: Basil Blackwell.

Hála, B. (1966). *La sílaba: su naturaleza, su origen y sus transformaciones*. Madrid: Consejo Superior de Investigaciones Científicas.

Horák, P. (1999). Implementation of original Czech TTS system on the Epos speech synthesis platform. In: *Proceedings of the 8th Czech-German Workshop.* Prague: Academy of Sciences of the Czech Republic, pp. 45–46.

House, A. C. & Fairbanks, G. (1974), The influence of consonant environment upon the secondary acoustical characteristics of vowels. In: *Experimental Phonetics*. N. J. Lass (ed.). MSS Information Corporation, New York, pp. 79–99.

House, J. & Johnson, M. (1987). Enlivening the Intonation in Text-to-Speech Synthesis: An 'Accent-Unit' Model. In: *Proceedings of the ICPhS, Tallin*, Vol. 1, pp. 134–137.

House, J. (1989). Syllable structure constraints on f_0 timing. Poster presentation. Edinburgh: LabPhon II.

Hyman, L. M. (1975) *Phonology – Theory and Analysis*. New York: Holt, Rinehart and Winston.

Keller, E. (1987). The variation of absolute and relative measures of speech activity. *Journal of Phonetics*, 15, pp. 335–347.

Keller, E., & Zellner, B. (1998). Motivations for the prosodic predictive chain. Proceedings of ESCA Symposium on Speech Synthesis. Paper 76, pp. 137–141. Jenolan Caves, Australia.

Keller, E., Bailly, G., Monaghan, A., Terken, J. & Huckvale, M. (Eds.) (2001). *Improvements in Speech Synthesis.* Chichester: Wiley and Sons.

Knowles, G., Williams, B. & Taylor, L. (Eds.) (1996). *A Corpus of Formal British English Speech*. London: Longman.

Kohler, K. (1987). Microprosody in segment perception. In: *Proceedings of the ICPhS, Tallin,* Vol. 1, pp. 80–83.

Lacheret-Dujour, A. & Beaugendre, F. (1998). La prosodie du français. Paris: CNRS.

Lecuit, V. & Demolin, D. (1998). The relationship between intensity and subglottal pressure with controlled pitch. In: *Proceedings ICSLP.* Sydney, pp. 3079–3082.

Lehiste, I. (1970). *Suprasegmentals*, Boston: M. I. T Press.

Lehiste, I. & Peterson, G. (1961). Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America,* 33/4, pp. 419–425.

Lehiste, I. & Peterson, G. E. (1967). Vowel Amplitude and Phonemic Stress in American English. In: I. Lehiste (ed.), *Readings in Acoustic Phonetics*. The M. I. T. Press, pp. 183–190.

Léon, P. R. (1992). *Phonétisme et prononciations du français*. Paris: Nathan.

MacNeilage, P. F. & Davis, B. L. (2003). Intersyllabic and word-level regularities in early acquisition. In: *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, pp. 383–386.

Möbius, B., Zimmermann, A. & Hess, W. (1987). Microprosodic fundamental frequency variations in German. In: *Proceedings of the ICPhS, Tallin,* Vol. 1, pp. 147–149.

Monaghan, A. (1992). Segmental Effects on Prosody for Synthetic Speech. In: *Proceedings of the International Conference on Spoken Language Processing*. Banff, pp. 1159–1162.

Nishinuma, Y. & Santi, S. (1992). Effect of Intensity Slopes on the Perception of Vowel Duration. *JASA*, vol. 92, n° 6, pp. 3425–3427.

Palková, Z. (1994). *Fonetika a fonologie češtiny*. Praha: Karolinum. [Phonetics and Phonology of Czech].

Pamies, A. (1996). Consideraciones sobre la marca acústica del acento fonológico. In: *Estudios de Fonética Experimental VIII*. Barcelona: Universidad de Barcelona, pp. 11–49.

Potužáková, L. (1999). *Mikrointonace v melodické oblasti*. M. A. thesis. Prague: Charles University in Prague. [Microintonation in the domain of melody].

Rossi, M. (1978). Interactions of intensity glides and frequency glissandos. *Language and Speech*, vol. 21, n° 4, pp. 384–396.

Rossi, M., Di Cristo, A., Hirst, D., Martin, Ph. & Nishinuma, Y. (1981). *L'intonation: de l'acoustique à la sémantique*. Paris: Klincksieck.

Séguinot, A. (1976). L'accent d'insistance en français standard. In: F. Caron, D. Hirst, A. Marchal & A. Séguinot (Eds.), *L'accent d'insistance. Emphatic stress*. Montréal, Paris, Bruxelles: Didier.

Silverman, K. (1986). Fo segmental cues depend on intonation: The case of the rise after stops. *Phonetica 43*, pp. 76–91.

Sorin, Ch., Larreur, D. & Llorca, R. (1987). A rhythm-based prosodic parser for text-to-speech systems in French. In: *Proceedings of the ICPhS, Tallinn,* Vol. 1, pp. 125–128.

Traber, C. (2000). Spectral smoothing of diphone boundary mismatches. *COST 258 Workshop*. Stockholm.

Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1990). Gradient effects of fundamental frequency on stop consonant voicing judgments. *Phonetica, 47*, pp. 36–49.

Zellner, B. (1998). *Caractérisation et prédiction du débit de parole en français. Une étude de cas.* Thèse de Doctorat. Lausanne: Université de Lausanne.

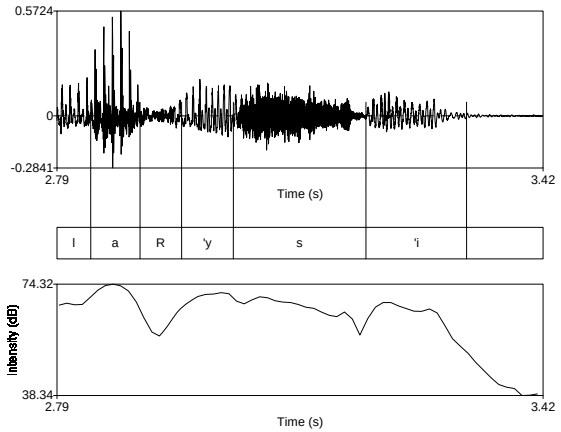

Fig. 1 Example of three French vowels (*la Russie* 'Russia') illustrating inherent intensity (the open vowel [a] is intrinsically more intense than the close vowel [y]), as well as different microdynamic shapes (e. g. the two-peaked profile of the final vowel [i])

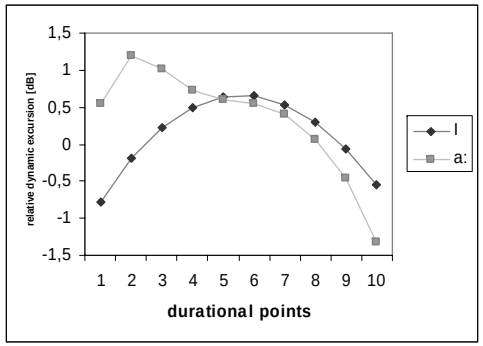

Fig. 2 Average microdynamic profile of the English vowel segments [I] and [A¯]. Number of observations: 183 and 51 respectively.

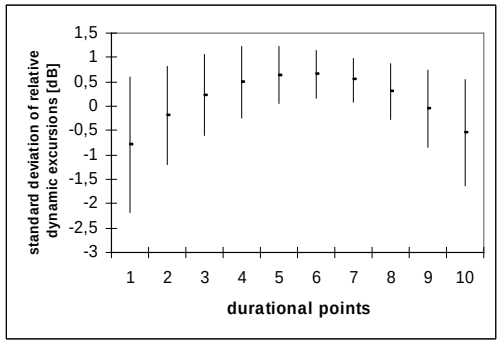

Fig. 3 Average microdynamic profile of the English vowel [I] with the standard deviation of the individual points. Number of observations: 183.

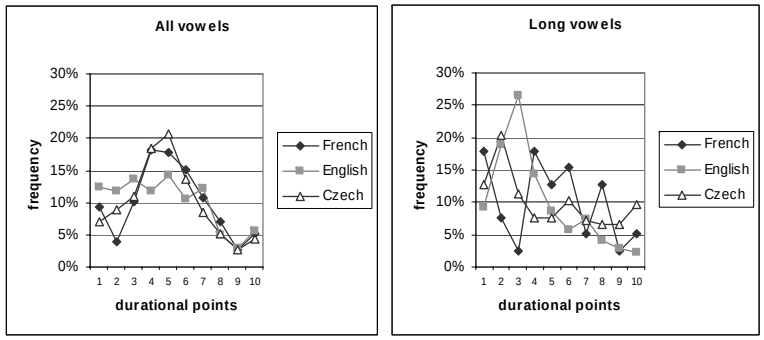

Fig. 4 Distribution of intensity peaks in French, English and Czech vowels (including diphthongs). For the definition of long vowels, see the beginning of section 3.4. Number of observations/average duration:

all vowels – French: 1034/78 ms, English: 1040/76 ms, Czech: 1034/79 ms;

long vowels – French: 39/129 ms, English: 174/107 ms, Czech: 196/118 ms.

Table I. Characteristics of vowels with secondary intensity peaks

	French	English	Czech
% of vowel segments with left secondary peak	8.61%	3.46%	6.18%
% of vowel segments with right secondary peak	10.44%	7.31%	7.25%
Total (left or right secondary peak)	19.05%	10.78%	13.43%
Prevailing categories	ɪ, ʏ, o), A)	ɪ˺, ʊ˺, A˺, diphth.	ɑ˺, E˺, ʊ˺, ɪ˺
Average duration of the vowels with left or right secondary peak	103 ms	131 ms	144 ms

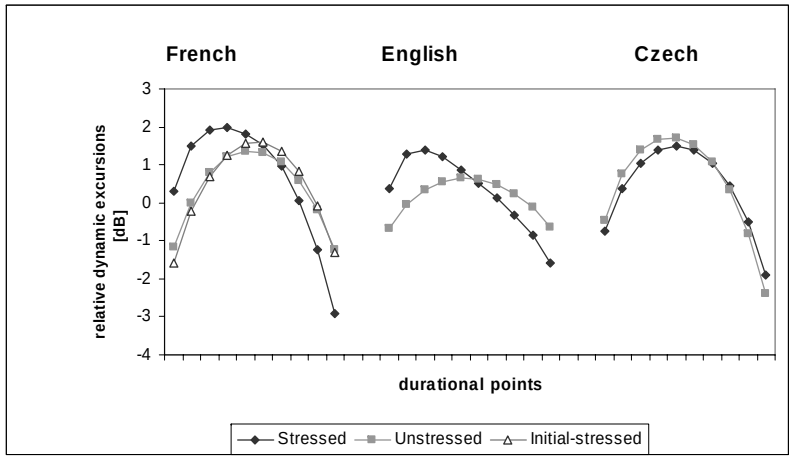

Fig. 5 Average microdynamic profiles for vowels perceived as stressed and unstressed (French, English, Czech). For French, it proved useful to distinguish initial vs. non-initial stress, since the microdynamics of stressed vowels in initial position resembled more those of vowels perceived as unstressed than those perceived as stressed. Number of observations/average duration:

stressed – French: 276/106 ms, English: 361/114 ms, Czech: 326/75 ms;

unstressed – French: 643/67 ms, English: 604/48 ms, Czech: 694/80 ms;

initial-stressed – French: 115/68 ms.

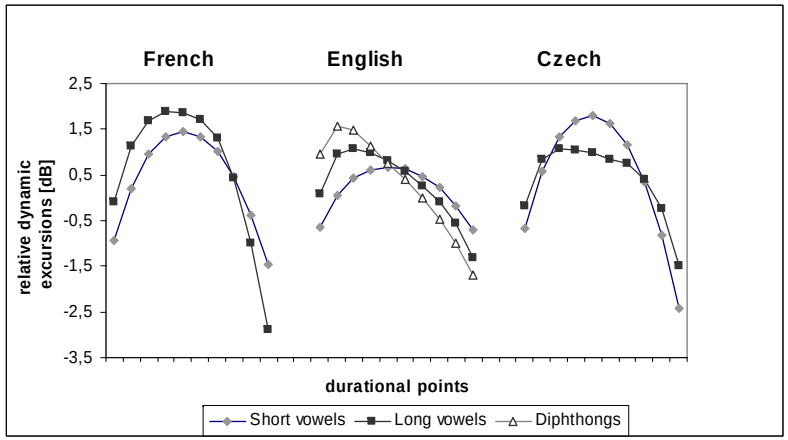

Fig. 6 Average microdynamic profiles for short vowels, long vowels and diphthongs. Czech diphthongs are not represented for their very low frequency. Number of observations/average duration:

short – French: 995/76 ms, English: 697/53 ms, Czech: 554/70 ms;

long – French: 39/129 ms, English: 174/107 ms, Czech: 196/118 ms;

diphthongs – English: 169/133 ms.

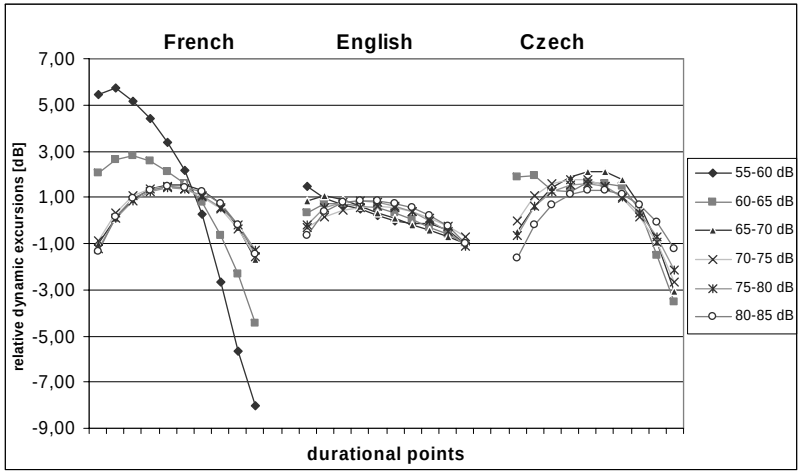

Fig. 7 Average microdynamic profiles for vowels of different average macrointensity (as given in the box on the right). Number of observations/average length for the 6 categories in the order as shown in the graph:

French – 12/97 ms; 36/116 ms; 190/83 ms 499/75 ms; 274/73 ms; 23/73 ms;

English – 5/68 ms; 20/74 ms; 57/61 ms; 274/59 ms; 461/79 ms; 215/92 ms; 8/78 ms;

Czech – 20/126 ms; 56/98 ms; 240/85 ms; 573/74 ms; 126/67 ms.

Note: The visible difference in overall amplitude between the three languages is to be imputed to recording conditions (cf. note after Fig. 5).