AMUR STATE UNIVERSITY

PROCEEDINGS

of the 1st International Conference

Phonetics without Borders

June 13 – 15, 2011

Blagoveshchensk, 2011

ББК 81.2 Англ

УДК 811.11+801.4

Т 44

PHONETICS WITHOUT BORDERS

Phonetics without borders: Proceedings of the 1st International Conference / Editor S. V. Androsova. -- Blagoveshchensk: Amur State University, 2011. – . 134.

The Introduction presents a brief summary of the Conference "Phonetics without Borders", that was held on June 13-15th 2011 in Blagoveshchensk within the Amur State University co-organized between the Departments of Linguistics and International Relations of AmSU.
The Conference was devoted to 300 anniversary of Mikhail Vasilyevich Lomonosov – the great Russian Scientist (November 19, 1711 – April 15, 1765) who was one of the first founders of University education in Russia.
The Conference aimed at pursuing international communication between specialists in different fields of Phonetics. Integration of the Russian Federation into Global Educational Space brings new opportunities of sharing experience and enrichment of knowledge in the Theory of Phonetics and Phonology, acoustic and perceptual analysis, computer speech modeling, methods of teaching pronunciation for foreign learners, cross-language studies and in other directions.
All the participants claimed that the Conference contributed to working out more efficient ways of dealing with challenging issues in the ares mentioned.

This project was made possible by an award from the Russian Humanitarian Scientific Foundation (РГНФ) of the Government of Russian Federation, through a program of conference projects support (project number 11-04-14060г), administered by RHSF (РГНФ).
None of these organizations is responsible for the views expressed herein.

ISBN 978-5-93493-163-7

THE SYLLABLE AS A MINIMAL UNIT OF COORDINATION

Svetlana Androsova

Amur State University, Russia
androsova_s@mail.ru

ABSTRACT

There are two levels – language and speech. This is what the division into Phonetics and Phonology stems from. Phonological units differ drastically from their phonetic manifestations. The interaction between these requires much deeper insight than traditional concepts can suggest. The paper aims at studying syllables behavior in American English spontaneous speech.

The traditional concept of Phonological/Canonical Syllable in English says: checked vowels cannot form open syllables, single consonants and sequences disallowed at the beginning of words cannot function as onsets, word boundaries terminate the possibility of forming open syllable even if distribution rules allow it. Thus closed canonical syllables (C)VCs are postulated to be most characteristic of English. However there is much evidence concerning closer pronunciation ties in CV despite canonical syllable boundaries. It leads to the search of "the other" phonological syllable. The hypothesis is that (i) there is the level called phonological Coordination which function is to bridge the gap between language and speech (ii) the Syllable is a minimal unit of Coordination. Acoustic analysis American spontaneous speech is used to trace the evidences of coordination which demonstrate that unlike the Canonical Syllable, the Syllable as a minimal unit of coordination is open.

Keywords: Syllable, Coordination, stress, duration, allophonic variation.

1. INTRODUCTION

1.1. Criticism on the Syllable functions

There are two levels – Language System (further referred to as Level 1) and speech (further referred to as Level 2), in other words, system and its material representation. This Saussure's Dichotomy is what the division of Phonetics and Phonology stems from. The two important questions here are (i) Where is the Syllable in this dichotomy? and (ii) What function/functions does the Syllable perform?

There is no problem answering these question for languages like Chinese where the Syllables is a minimal meaningful unit. However it appears to be a challenging task for languages like English where syllable has no meaning of its own. The Syllable can be considered one of the most challenging units in languages like English as opposed to the ones like Chinese. The Syllable is more often appealed to than defined [Price, 1980: 327]. There are almost as many ideas of the Syllable as phoneticians and phonologists speaking about the issue. Unfortunately many of them either lack experimental support or that support is gained through strictly controlled laboratory experiments that are far from natural speech.

There is a list of well-known functions suggested for the syllable to perform:

1. Providing continuity of speech.
2. Phonemes distribution patterns arrangement.
3. Expiratory function.
4. Sonority arrangement.
5. Word stress and sentence accent phonetic manifestation.
6. Creating a minimal context for phonemes relevant features phonetic manifestation.

As far as *function 1,* it seems to reflect assumptions based on rather vague subjective parameters. *Function 3* did not find any reliable experimental support. *Function 4* does not solve the problem of syllabification [Malmberg, 1963: 67]. Therefore I will not turn to them any more in the present article.

Function 2 enables to view the Syllable as a phonological unit. The traditional concept of Phonological/Canonical Syllable belonging to Level 1 in English says: checked vowels can not form open syllables, single consonants and sequences disallowed at the beginning of words can not function as onsets, word boundaries are often said to terminate the possibility of open syllable forming even if the distribution rules allow it. Thus phonological syllables of CVC structure – closed syllables are postulated to be most characteristic of English including American English.

The first and the main problem connected with such Syllable is no necessity for introducing a new phonological unit — the Syllable — to explain phonemes distribution rules, which can easily be done with the already existing units – the Word and Morpheme. Therefore function 2 does not form an independent phonological unit and thus disables to view the Syllable as belonging to Level 1.

The second problem is total disconnect between Canonical Syllable and its material representation. There is much evidence about closer pronunciation ties of a consonant or a consonant sequence with the following vowel despite phonological syllable boundaries. The gap between Syllable as a phonological unit its phonetic correlate has become too large to bridge.

Function 5 is often referred to [Bailey, 1980: 27-28; Marshall, Nye, 1983: 433; MacNeilage, 1998; Wells, 1990: 80] and can be viewed as making the Syllable belong to Level 1. However the existing syllabification rules based on stress are either terminated by morpheme/word boundaries [Wells, 1990] or variate depending on tempo and speech style [Bailey, 1978; Redford, 2007: 1666]. The lack of universal criterion for syllabification casts doubt on the Syllable's reality, its possibility to function as an independent unit and disallows to view the Syllable as belonging to Level 1 again.

If there is no possibility to view the Syllable as belonging to Level 1 there are two options: (i) to look for it at Level 2 or (ii) to suggest another level.

The first option was realized by Liya Bondarko who synthesized *functions 5-6* in her concept of the Syllable for the Russian language: (1) syllable is a unit characterized by integrated articulatory performance of its constituents; (2) syllable is a unit of speech having linguistic value: it is the minimal unit where phonemes relevant features acquire their phonetic shape; (3) allophonic variation of phonemes is a result of integrated articulatory performance within the Syllable. Bondarko [Бондарко, 1981: 57, 107-126] argued that (i) the Syllable is a co-articulatory unit – a unit of speech (belongs to Level 2 in my interpretation), (ii) the acoustic cues of syllable integrity are syllable contrasts, (iii) the Syllable in the Russian Language is open. Studies in co-articulation demonstrate that such syllable can be universal [Pickett, 1999: 147] (see also studies of Peggy Mok using the material of the Thai language and British English [Peggy Mok, 2010: 1351, 1353]).

For the English language the concept of the Syllable as a real phonetic unit which boundary is determined by certain articulatory and acoustic cues was developed by B. Malmberg [Malmberg, 1963: 65] who wrote that it would be a mistake that the Syllable does not exist as a phonetic phenomenon and that phonemes grouping is done entirely conventionally without any phonetic grounds.

However linguistic value of such Syllable for the English language is mostly neglected. It is said that:

1. Co-articulation has nothing to do with the Syllable. It is trivial. It is no more than physiological adjustment of the neighboring sounds. Even if co-articulation is stronger in CV structures CV is not a real Syllable.
2. Elements are united within the Syllable by higher level relations of coordination.

Coordination is understood as governing of duration of element within the Syllable. Coordination is present only in VC which is considered the basic unit of English speech planning [Fowler, 1986; Jong, 1991: 8; Lehiste, 1970; Lehiste, 1971: 160].

The basic experimental proof of closed syllable predominance in English is negative correlation of durations in VC [Lehiste, 1971: 160].

Negative correlation in VC being efficient for monosyllabic words however did not work in disyllabic words where in V1CV2 structure C demonstrated stronger correlation with V2 compared to V1 (see Table 1) [Lehiste, 1971: 165].

Table 1: Duration correlation in disyllabic words (after I. Lehiste).

Word	V1C	CV2
steady	–0,13	+0,18
skiddy	–0,32	–0,61
skitty	+0,01	–0,37

Blaming the advocates for the Syllable as a minimal unit of coarticulation for triviality and oversimplification of the concept, duration-correlation-based-closed-Syllable supporters commit the same mistake: limiting speech production mechanisms to duration governing in monosyllabic isolated words is an obvious oversimplification.

No matter what function to choose – co-articulation or co-ordination as duration governing – the Syllable is an alien at Level 2. The first reason is the absence of an invariant phonetic parameter as a cue of syllable integrity. The second

is obvious variation of the Syllable: different expression of syllable contrasts for stressed and unstressed CVs, weakening of contrasts in connected speech as opposed to isolated words uttering [Бондарко, 1981: 116-124]. Is it one syllable or two different syllables in "Where are you going *to*? – I'm going *to* the movies" (the same question should be asked about Bondarko's examples of stressed and unstressed /di/ in two Russian words *дико* and *дикарь*).

If the Syllable is an entirely phonetic unit, it should be two different syllables and it should be accepted that every time we realize a sequence of phonemes we deal with different syllables. Thus the number of syllables is unlimited. As there is no invariant phonetic cue of the syllable boundary, reliable phonetic cues of syllabic integrity are non-existent. How to syllabify if phonetic parameters are vague or absent? The reality of such a syllable is doubtful.

If /di/ in both examples is one syllable, then we deal with variation known to be characteristic of Phonological units (Remember, we do not have an adequate one performing an independent function and being adequately connected with phonetic reality!). It seems a vicious cycle unless another phonological option is suggested. At this point we turn to coordination again but in its initial meaning suggested by N. Bernstein.

1.2. Three blocks vs two levels

It is obvious that besides language as a system of category structures, which Formal Phonology deals with, and their material representation as result of speech production there is another constituent that is called by R. Port [Port, 2008] "Language Pattern" or "speech activity" by L. V. Scherba [Щерба, 1974: 24]. The basic principle of any type of activity is coordination. Speech production is not the exception.

The term "Coordination" was first defined by the Russian physiologist N. Bernstein in 1967: Co-ordination is an activity which guarantees that a movement shall have the homogeneity, integration and structural unity. This activity is principally based not on particular process of individual neurons, but on the determinate organization of their common activity. This organization must necessarily be reflected in the anatomical plan in the form of localization [Bernstein, 1967: 30]. It is mastering the very many degrees of freedom involved in a particular movement or, in other words, of reducing the number of independent variables to be controlled [Bernstein, 1967: 127].

Two of the ten basic principles of coordination are picked up at the present research as being of primary importance to speech production (i) the kinematic (spatiotemporal) details of any coordinated state are not determined at the outset, in a single step by a single system but are contributed gradually, by many subsystems working together (ii) a plan for a coordinated act is defined functionally in an abstract manner that refers neither to body segments nor to actual motions (see the analysis of the ten principles in M. T. Turvey [Turvey, 1990: 939]).

I hypothesize that

I. Coordination is the basic principle of speech activity. It is coordination of articulatory movements while planning syntagma and phrase. Coordination can refer both to the process of speech production and the schemes and rules according to which that process is performed.

II. These schemes and rules form a block – the block of Coordination (further referred to as Coordination) which function is to bridge the gap between the System and its material representation during speech production process. Thus instead of two levels three blocks are suggested: Language System-Coordination-Speech.

III. The Syllable is the minimal (not basic!) unit of Coordination in which the way from the System to its material representation is programmed or, in Port's words, the patterns of language are revealed. It is the minimal unit that shows how phonological segments overlap in time. This syllable is of primary importance to the speaker.

IV. Characteristics of a syllable are determined by the interaction of a number of intrinsic and extrinsic factors: 1) arrangement of phonemes in the utterance; 2) stress; 3) relations with other syllables; 4) interaction with morphemes and words; 5) phrase position; 6) pauses; 7) tempo of speech.

V. The Syllable in American English as a minimal unit of Coordination unlike the Canonical Syllable is open because

a) temporal characteristics of consonants (correlation and covariation between consonant duration and surrounding vowels duration) and assimilatory variation, all of them demonstrating stronger dependence upon the formant structure and the stress pattern of the following vowel rather than the previous one across word boundaries.

b) Consonant sequences demonstrate close coarticulatory ties between the constituents across morpheme and word boundaries, which disallow to split them by the syllable boundary. The whole

consonant group is tautosyllabic with the following vowel across morpheme and word boundaries as coarticulatory effects with it are stronger than those of the previous vowel.

VI. In order to express the relations between the Syllable as the minimal unit of Coordination and the Canonical Syllable the notion of resyllabification (which deals with reassignment of a phonetic constituent from one syllable to another [Greenberg, 1999: 162]) is necessary.

This paper reports on a set of experiments to prove the hypothesis.

1. Experiment 1 – analyzing syllable timing: a) correlation and covariation between consonant length and surrounding vowels length; b) consonant length and surrounding vowels stress;
2. Experiment 2 — acoustic analysis of consonant allophonic variation.

2. EXPERIMENT 1

2.1. Method

The major experimental proof of the Closed Syllable being the basic unit of speech production in English is considered to be Ilse Lehiste's experiment with single consonant's duration in monosyllabic CVC words [Lehiste, 1970; Lehiste, 1971]. It resulted in revealing negative correlation between V and C [Lehiste, 1971: 160]. However it did not work in disyllabic words where in V1CV2 structure C demonstrated stronger correlation with V2 compared to V1 [Lehiste, 1971: 165].

As far as the correlation of C with V1 failed in disyllabic words it was logical to assume that the chance of its failure would be even higher in the flow of speech that will lead to resyllabification in terms of relationship between the Syllable as the minimal unit of Coordination and the Canonical syllable. Spontaneous speech seemed a perfect object to demonstrate the failure.

2.1.1. Participants and materials

Three volunteer native male speakers of American English (D1, D2, and D3) aged from 50 to 28 were seated in a quiet room individually. They were presented a list of questions (about family, hobbies, home town, cooking, favorite books etc), which none of the speakers had seen before, and were asked to speak as long as they could answering each question. None of the speakers reported of speech or hearing disorders. All the participants had Bachelor's Degrees. The speakers demonstrated no sharp dialectal features in their speech and can be considered as representing Standard American English (SA). The recording was made with the help of PRAAT that was also used for acoustic measurements. The total duration of spontaneous monologues recorded was around 3 hours.

2.1.2. Procedure

Duration measurements were done on wave form and dynamic spectrogram. Duration of single intervocalic consonants was measured. The consonants included stops /d b p k/, tap [ɾ] substituting /t/ and /d/, nasals /m n/, and fricatives /s z f v/. Measurements involved the following stress patterns: 'V1CV2, V1C'V2, V1CV2, 'V1C'V2. Ten tokens of each consonant were measured from each of the three speakers. The total number of single consonants measured is 660 (220 from each speaker). Duration of the previous vowel (V1) and following vowel (V2) was also measured for further calculation of correlation and covariation of duration in the patterns mentioned above. 'V1CV2 and V1C'V2 patterns (ten tokens of each consonant in each of the two stress patterns) were further considered for the effect of stress on consonant duration. Correlation (CorrV1 and CorrV2), covariation (CovarV1 and CovarV2), minimum (min.), maximum (max.) and mean duration of consonants were calculated using the corresponding functions in OpenOffice.org Calc.

Marginal duration values of word-final intervocalic consonants before a) unstressed vowels, b) stressed vowels and c) strongly stressed vowels were further considered for the degree of prosodic prominence effect.

2.2. Results

2.2.1. Correlation and covariation of durations in V1CV2

Table 2 in the appendix shows the result of correlation study for the three speakers. As expected, there is certain negative correlation in V1C especially for D2 and D3. D1 results show positive CorrV1 for 6 of the 11 consonants. The magnitude, however, turns out very low in the majority of cases. CorrV2 values are positive for all consonants and all the three speakers proving higher magnitude compared to CorrV1 in general and for most of the single consonants studied except for /b k/ in D1, /s f n/ in D2 and /d p/ in D3.

Table 3 (see the Appendix) demonstrates the results of covariation study for the three speakers.
As shown in Table 3 there is no regularity in duration interdependence as far as V1C structure is concerned for D1 and D3: covariation for certain consonants is positive, while for others it is negative. In D2 speech CovarV1 was negative for most consonants. At the same time CovarV2 turned out positive for the three speakers for all the consonants involved. Values of CovarV2 demonstrate higher magnitude compared to CovarV1 except for /f/ for D2. However, the range of values is very high within each speaker (286.04-3424.19 for D1, 167.35-1445.76 for D2, and 133.69-1004.8 for D3) and for the same consonants different speakers speech (e.g. /z/: D1 – 3424.19, D2 – 285.45, D3 – 852.65).

2.2.2. The effect of the following vowel prominence in spontaneous speech

Tables 4-6 present the results of prominence effect of V2. The magnitude of the effect is rather large.
The results of Experiment 1 demonstrate that in American spontaneous speech there is no systematic difference of consonant duration for word-internal vs word-initial vs word-final single consonants all other conditions being equal (the same stress pattern and approximately the same tempo). The tendency is especially strong for consonants before unstressed vowels. The following examples demonstrate the tendency.
Before stressed vowel: I **d**on't – i**d**ea (104 msec – 108 msec), for **p**eppers – o**p**inion (141 msec – 142 msec), we **f**inish – i**f** any (80 msec – 81 msec), stea**k** is – new **c**ountry (127 msec – 129 msec), a**n** interesting – phe**n**omenon (65 msec – 62 msec), very **m**any – amalgamation (76 msec – 78 msec) etc.
Before unstressed vowel: musi**c** is – electroni**c**a (85 msec – 87 msec), for pe**pp**ers – kee**p** in (66 msec – 66 msec), pro**c**ess – gla**ss** of water (100 msec – 100 msec), ama**z**es me – tho**se** are (61 msec – 62 msec), famous – dream of (55 msec – 58 msec), one of – anything (37 msec – 38 msec).

2.2.3. The degree of prosodic prominence effect of the following vowel

Being the minimal unit of coordination the Syllable has to be subjected to higher level influences of the Syntagm for example expressing the degree of prominence while differentiating not only between stressed and unstressed syllables but between strongly stressed, weaker stressed, and unstressed ones. Word-final intervocalic consonants were chosen as an object. It was difficult to find statically representative material under the limited conditions, however the results for the consonants that occurred in the positions under consideration enables to speak about the tendency of gradual reduction of consonant duration before strongly stressed vs weaker stressed vs unstressed vowels (see Table 7 in the Appendix) and that tendency is observed across word boundaries.

3. EXPERIMENT 2

3.1. Syllable centric approach to allophonic variation

It is considered that the choice of a particular allophone is determined by its position in the Syllable as it's onset or coda (see S. Greenberg about syllable-centric approach to allophonic variation [Greenberg, 1999]).
The issue of allophonic variation is too broad to be discussed in one article. Therefore I will turn to the three most controversial types of allophones as far as their position in the Syllable: (i) weak voiceless stops as a function of aspiration loss, (ii) intervocalic taps, and (iii) glottal burst.
John Wells [Wells, 1990] offered a list of allophones that appear only as syllable codas among which we can see the ones mentioned.
However, they can be much better explained by stress pattern withing a word or syntagm/phrase.

3.1.1. Aspiration and word stress and/or stress in the syntagm or phrase

The alternative explanation is that unaspirated single voiceless stops are tautosylabic with the following unstressed vowel because its unstressed character determines the lack of aspiration. This rule works well for the word-final /k/ and /p/ when words come together to build a syntagm or phrase. Figures 1 and 2 show the examples of aspiration loss within word and in the syntagm.
If the following vowel appears to be stressed in the Syntagm or Phrase it is the condition for aspirated word-final voiceless stop like the one in figure 3.
In the phrase the word *music* is located on the syntagmatic boundary; there is no pause between the syntagms; single /k/ finishes the first syntagm; the second syntagm begins with the strongly stressed vowel: *I like all kinds of musi**c, a**ll kinds of different music*. Aspiration for /k/ is strong and it reflects the formant structure of the following vowel rather than the previous one.

Figure 1: Aspiration loss in word-internal /k/ in the word «pa**ck**age».

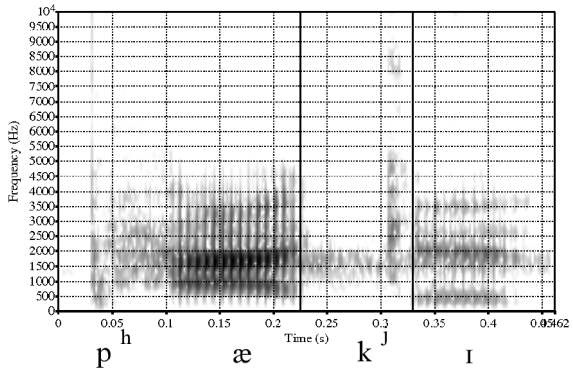


Figure 2: Aspiration loss in word-final /k/ as a result of following vowel stress lack in the syntagm.

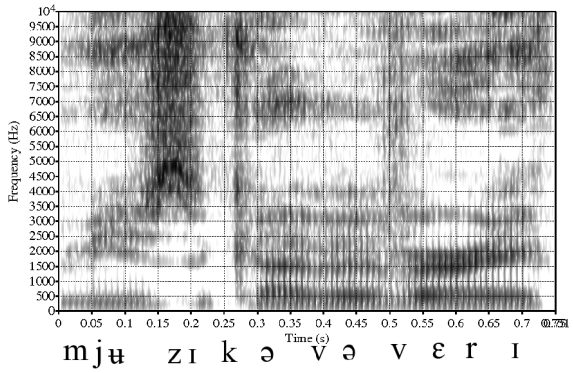
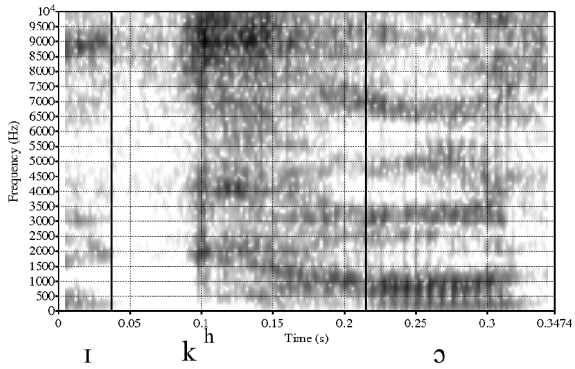


Figure 3: Aspiration gain in word-final /k/ in «music all» as a result of the following vowel being accented in the phrase.



Word-initial voiceless stops in the Syntagm and Phrase follow the same pattern. The following two examples (see Fig. 4-5) demonstrate the behavior of /p/ in the word *peer*. In the first one /p/ is unaspirated as a result of considerable stress weakening in the syntagm, while in the second one there is strong aspiration as a result of strong stress in the syntagm.

Building relations with Canonical Syllable it should be said that in the first three cases we obviously see resyllabification process when the following vowel stress pattern dictates the choice of unaspirated or aspirated phonologically voiceless stop across word boundaries. Analyzing the given examples that are typical for ASS we can see that word-internal pattern hardly differs from phrase-internal pattern at least for word-final /k/ and /p/.

Figure 4: Aspiration loss in word-initial /p/ as a result of following vowel stress lack in the syntagm.

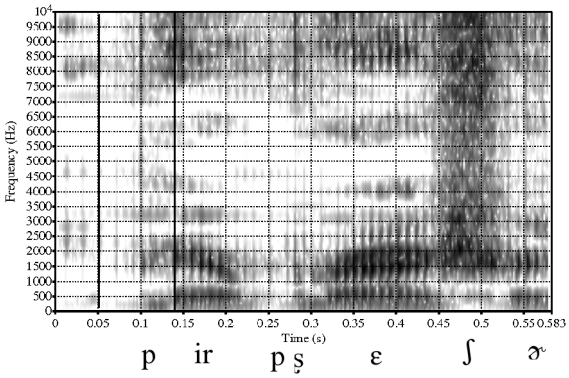
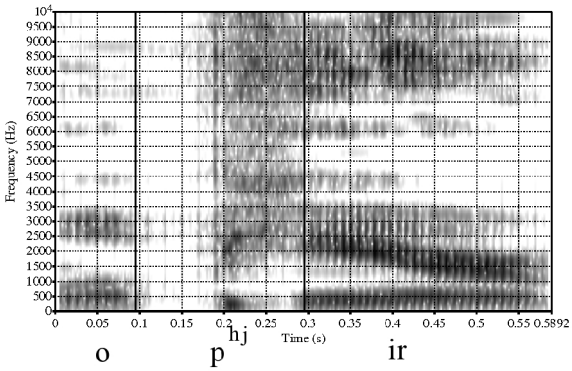


Figure 5: Aspiration gain in word-initial /p/ in «to a peer» as a result of the following vowel being accented in the phrase.



3.1.2. Taps

Taps are traditionally considered codas because they cannot occur word-initially. It is true when we deal with isolated words. In the flow of speech the restriction is canceled.

The omission of final /d/ in the word *told* from *told today* created intervocalic position typical for tap (see fig. 6).

Other typical examples revolve around *to* as particle, preposition or prefix: me **t**o be, go **t**o college, true **t**o a certain extent, be **t**ogether.

Another restriction about tap is that it cannot be onset of a stressed syllable both word-initially and

word-internally. However, the restriction does not work in case of tapped /d/. An interesting example would be the one from *four **d**ays* (see fig. 7).

Figure 6: Word-initial tap instead of /t/ in «told **t**oday».

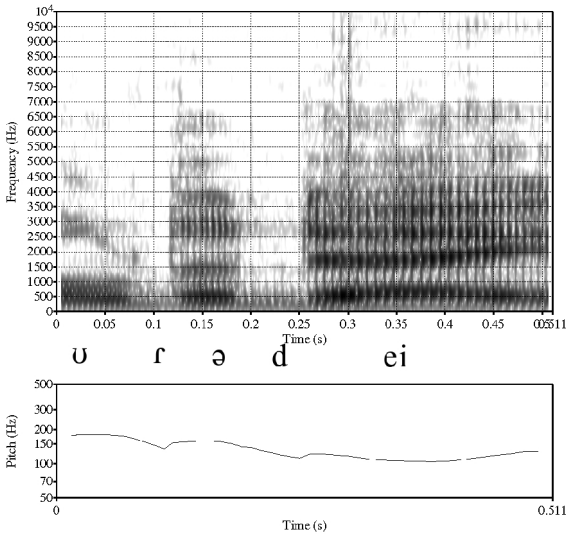
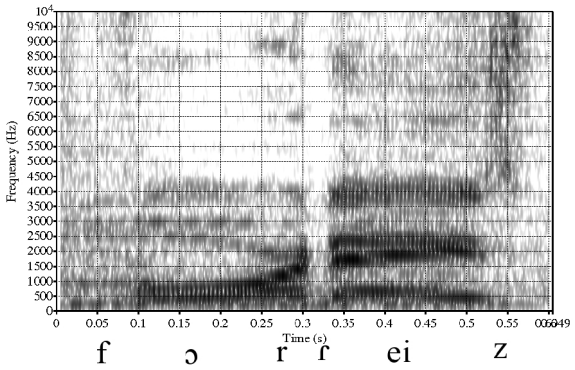


Figure 7: Word-initial pre-stress tap in «four days».



Other typical examples are a**dd**iction*, pro**d**uction, pre**d**ominant* and some others.

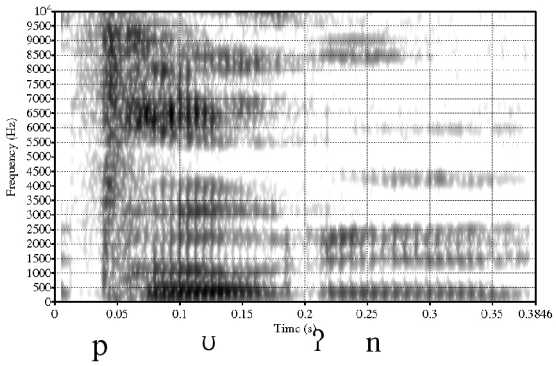
The restriction still acts in case of taps for word-initial and word-internal /t/: no tap in words like ***t**ea, a**tt**ention*.

True, both vowels are necessary for tapping. However, the results of acoustic analysis of ASS demonstrate that the stress pattern of V1 does not matter at all, while the stress pattern of V2 is relevant for /t/ tapping and mostly relevant for /d/ tapping at least word-initially and word-internally remaining irrelevant word-finally. In addition tap duration variates depending on the stress pattern of V2 showing no such dependence on V1; tap shows closer correlation and co-variation with V2 compared to V1 (see [ɾ] tables 2-6 in the appendix). Greater dependence of tap on V2 makes them tautosyllablic. In terms of co-ordination there is more regulation of tap on the part of V2 compared to V1. In terms of relations with the Canonical Syllable every morpheme- or word-final tap will be a resyllabification case.

3.1.3. Glottal burst

Up to the present moment only single consonants were viewed in the current paper. The only consonant cluster to be viewed here is *t+sonorant* (any except for /ŋ/). They are considered tautosyllabic from the point of view of coordination because the condition for glottal burst substituting [t] is the following sonorant presence. In most cases it occurs on morphological boundaries in examples like *lately, what* not etc. However it is not completely tied to the final position of the morphological unit in words like *Atlantic, atmosphere* or even the last name of the ex-president of Russia *Putin* (fig. 8).

Figure 8: Glottal burst in the word “Putin” (the last name of Russian ex-president, pronounced by D3).



In the cases mentioned here it is not mere coarticulation but determining the allophone choice as the main principle of syllabification within the syllable-centric approach to allophonic variation. The way *t+sonorant* cluster is viewed in the current study gives an implication for further viewing of other consonant clusters and their syllabic affiliation: the allophones of consonants constituting a certain cluster get better explanation being tautosyllabic.

3.2. Discussion

When we speak we mostly do not produce words as separate units. We produce syntagms and phrases. In this process word boundaries become illusive and phonemes variate according to their position in the syntagm and phrase. Canonical CVC syllable tied to morpheme and word

boundaries leaves many speech production patterns without any plausible explanation, while open syllable as a minimal unit of Coordination explains the choice of allophones in syntagms and phrases — phrasal allophones. Their temporal and spectral characteristic are acoustic cues of syllable integrity. These are found in CV across morpheme and word boundaries

4. CONCLUSION

Though it is impossible to deny that the experiments described here are far from being pure as probably all experiments based on spontaneous speech, their results still enable to speak about certain tendencies that turn out rather strong. It is clear that prosodic patterns of spontaneous utterance building put pressure on boundaries of such a linguistic unit as the Canonical Syllable and make them illusive. In this connection one of R. F. Port's articles' title is worth mentioning: "All is prosody: Phones and phonemes are the ghosts of letters" [Port, 2008]. According to R. Port [Port, 2008: 11] "the linguistic structures we model by using an alphabet (when either using orthography or when doing linguistic phonological analysis) are social structures or statistical patterns across a large corpus. They are not just psychological units or structures directly represented in the mind".

The same is true about syllables. Canonical CVC at least in American English is a perfect «social structure» that performs certain important functions (not connected with speech production process) that are well-known.

However, collecting empirical data to test phonological theories and models is logical. If the empirical data and the theory run too much contrary (which is certainly the case with Canonical closed syllable in American English) it seems to be the time to suggest a new theory.

Viewing the Open Syllable as a minimal unit of Coordination in American English enables to (i) consider the Syllable a phonological unit as far as coordination is of abstract, phonological nature; (ii) legitimately use speech patterns to build phonological model of the Syllable which turns out to be CV; (iii) bridge the gap between the theory and empirical data.

The empirical data — closer correlation and covariation of C with V2, great dependence of C on V2 stress pattern and the lack of such dependance on the one of V1, and closer coarticulatory ties that also contribute into the allophone choice — point to CV as a minimal unit of speech production in American English.

Allophonic variation presented in the current paper is only a part of the whole amount of phrasal allophones which choice can be explained from the position of the Open Syllable as a minimal coordination unit being a part of the Phrase imposing its effects on the Syllable.

Finally, performing measurements on spectrogram and waveform does not make syllable boundary a demarcation line on a spectrogram and waveform. It is still an abstract «line» that separates one phoneme realization from another to make their phrasal allophones most predictable.

5. REFERENCES

[1] Bailey C.-J. N. (1980) Evidence for Variable Syllable Boundaries. *The Melody of Language. Intonation and Prosody*. Baltimore, 25-39.

[2] Bailey C.-J. N. (1978). *Gradience in English syllabization and a revised concent of unmarked syllabization.* Indiana University Linguistic Club.

[3] Bernstein N. (1967). *The co-ordination and Regulation of movements*. Oxford-London...: Pergamon Press.

[4] Greenberg S. (1999). Speaking in shorthand – A syllable-centric perspective for understanding pronunciation variation. *Speech Communication*, 19, 159-186.

[5] Lehiste I. (1970). *Suprasegmentals.* Cambridge, Mass., L.: M.I.T. Press.

[6] Lehiste I. (1971). Temporal organization of spoken language. *Form and substance*, Odense, 159-169.

[7] MacNeilage P. F. (1998). The Frame/Content theory of Evolution of Speech Production. *Behavior and Brain Science*, 1998, (21), 499-511.

[8] Malmberg B. (1963). *Phonetics*. N-Y.

[9] Marshall C. W., Nye P. W. (1983). Stress and vowel duration effects on syllable recognition. *JASA*, 74, (2), 433-443.

[10] Peggy Mok P. K. Language-specific realization of syllable structure and vowel-to-vowel coarticulation. *JASA*, 2010, 128 (3), 1346-1356.

[11] Pickett, J. M. 1999. *The acoustics of speech communication*. Boston, London...: Allyn and Bacon.

[12] Port, R. F. (2008). All is prosody: Phones and phonemes are the ghosts of letters. *Speech Prosody 2008*, http://www.isca-speech.org/archive/sp2008/papers/sp08_007.pdf

[13] Port, R. F. (2007). How are words stored in memory? Beyond phones and phonemes. *New Ideas in Psychology, 25,* 143-170.

[14] Price P. J. (1980). Sonority and Syllabicity: Acoustic Correlates of Perception. *Phonetica*, 137, 327-343.

[15] Redford M. A. (2007). Word-internal versus word-peripheral consonant duration in three languages. *JASA*, 121 (3), 1665-1678.

[16] Turvey M. T. (1990). Coordination. *American Psychologist*, 45 (8), 938-953.

[17] Wells J. C. (1990). Syllabification and allophony. *Studies in the pronunciation of English (a commemorative volume in honour of A. C. Gimsom).* Great Britain, 76-86.

[18] Бондарко, Л. В. (1981) *Фонетическое описание языка и фонологическое описание речи*. Ленинград.

[19] Щерба Л. В. (1974). Об ударении. *Языковая система и речевая деятельность*. Ленинград, 176-177.

Table 4: *Prominence effect on intervocalic consonant duration in D1 speech.*

Consonant	Duration before stressed vowel			Duration before unstressed vowel		
	Min.	Max.	Mean	Min.	Max.	Mean
/d/	93	149	114	24	42	35
[ɾ]-/t/	41	63	51	21	39	26
/b/	94	136	111	54	90	68
/p/	128	191	155	91	132	110
/k/	113	159	130	79	110	99
/s/	117	182	155	109	177	140
/z/	83	224	123	71	102	90
/f/	122	169	142	100	133	114
/v/	61	108	89	45	67	57
/n/	70	122	92	37	62	50
/m/	90	130	110	45	78	65

Table 5: Prominence effect on intervocalic consonant duration in D2 speech.

Consonant	Duration before stressed vowel			Duration before unstressed vowel		
	Min.	Max.	Mean	Min.	Max.	Mean
/d/	81	122	104	23	46	32
[ɾ]-/t/	21	47	37	22	37	28
/b/	82	136	109	43	77	64
/p/	102	158	131	72	99	81
/k/	98	150	121	60	102	84
/s/	106	170	130	97	114	105
/z/	70	108	88	55	75	66
/f/	79	174	113	75	105	93
/v/	63	99	78	33	71	58
/n/	43	90	63	24	51	34
/m/	70	171	94	53	71	61

Table 6: Prominence effect on intervocalic consonant duration in D3 speech.

Consonant	Duration before stressed vowel			Duration before unstressed vowel		
	Min.	Max.	Mean	Min.	Max.	Mean
/d/	70	157	102	22	62	36
[ɾ]-/t/	30	48	40	16	26	22
/b/	70	128	95	39	82	56
/p/	75	198	125	65	89	73
/k/	92	184	142	73	102	89
/s/	104	183	128	85	124	102
/z/	73	167	104	59	81	69
/f/	80	157	104	75	109	92
/v/	58	111	74	27	62	45
/n/	41	106	64	18	46	33
/m/	75	153	92	49	61	56

Table 7: Single intervocalic word-final consonant duration decrease before stressed and unstressed vowel compared to intonation center (the material of three speakers).

Phoneme	Duration decrease (%)					
	D1		D2		D3	
	Before stressed vowel	Before unstressed vowel	Before stressed vowel	Before unstressed vowel	Before stressed vowel	Before unstressed vowel
/d/ – [ɾ]	15-36	53-64	13-36	26-53	18-41	57-69
/ŋ/			16-36	33-78	15-37	40-64
/n/	16-43	51-70				
/z/	8-48	39-49	8-36	32-51		
/p/					33-38	43-50
/k/					30-49	45-60

CHANGES IN PRONUNCIATION MODELS REPRESENTED IN MODERN ENGLISH-LANGUAGE COURSEBOOKS AND IMPLICATIONS FOR THE LEARNERS

Tatiana Anikhovskaya

Blagoveshchensk State Pedagogical University, Russia
taniho@inbox.ru

ABSTRACT

Until quite recently most British-English-oriented coursebooks provided students with listening tasks and pronunciation exercises based mostly on RP pronunciation. However the situation is changing rapidly and more and more coursebooks aim at giving students a broader view of the "pronunciation landscape" of Great Britain by incorporating various regional and non-native English accents into their listening activities. The article offers an overall analysis of three intermediate level English language coursebooks "*Enterprise"*, "*New Headway"* and "*Inside Out"*, revealing pronunciation peculiarities of the speakers recorded, and makes an attempt at finding the answer to the question whether the priorities for teaching English pronunciation should be changing.

Keywords: Teaching pronunciation, standard pronunciation, regional accents.

It is a well known fact that for many years the authors of British English language teaching materials have relied mostly on RP as a model for teaching pronunciation and developing speech comprehension skills. However RP, as any other standard, is not fixed and immutable, it changes with time and the new pronunciation tendencies are naturally finding their way to the recordings accompanying contemporary coursebooks. Moreover, in recent years there has been a noticeable change in the "pronunciation landscape" of Great Britain, with regional accents and non-native varieties of English gaining more power. On the one hand, more and more educated people are heard speaking with their regional accents (Northern English, Southern English, Estuary English) on radio and TV, and on the other, being a multicultural country, Britain buzzes with a whole set of different "Englishes", such as Indian, Greek, Japanese, etc. This peculiar current linguistic situation made the authors of teaching materials realize that in addition to gaining familiarity with standard pronunciation and the main trends of its development, overseas learners should be able to understand a range of British regional accents, as well as some non-native varieties of English.

The article offers an overall analysis of the audio materials for intermediate level English language coursebooks "*Enterprise"*, "*New Headway"* and "*Inside Out"*, which are quite popular and frequently used in this country for teaching English.

The main aim of the analysis is twofold: 1) to find out which of the current changes in standard pronunciation (described in detail by British phoneticians) are actually represented in the speech samples; 2) to reveal what dialects, or, to be more exact, what dialectal pronunciation features (also thoroughly described and classified) are used in various listening and pronunciation activities.

Out of the whole range of new pronunciation fashions and traditional dialectal peculiarities we've dwelt only on those which are frequently heard in the recordings analyzed and can be easily detected by means of auditory analysis. The latter is used as the main method of investigation, as it allows us to rely on what we can actually perceive by ear. It really seems to be indispensible for teaching purposes as the ultimate goal for us is to teach students to HEAR the difference. Still, experimental methods are equally important in such kind of investigation, allowing to check and clarify some controversial cases.

First of all, speaking about non-standard, purely regional forms of pronunciation, auditory analysis enabled us to reveal the following peculiarities:

1. Phoneme |Λ| is realized as |u|

In fact, vowel |Λ| does not occur in the Northern English accents. It is typically realized as |u| in Newcastle, Yorkshire and Bradford accents, for example. This peculiarity is quite frequently heard in the recordings to all coursebooks, mainly in dialogues imitating casual everyday conversations. It is interesting that the speakers who use it are not quite consistent about their accent, often pronouncing other sounds "in RP style". Here are the most vivid examples:

ENTERPRISE

- It will be another dull wet day tomorrow (ex.7 p. 22, a weather forecast).
- I was standing at the bus stop when it happened (ex. 10 p.48, a witness questioned).
- We had a wonderful meal and took a photograph of the bride and groom cutting their wedding cake (ex. 6 p.60, a woman shares impressions of her sister's wedding).
- It's more comfortable than being on a hot crowded bus (ex.11 p. 149, two people are talking about transport).
- If a person remains unemployed for a long time it becomes difficult for him to support himself (ex. 11 p.87, a man interviewed about the possible reasons of crimes).

INSIDE OUT

- We haven't been in touch since that day (tapescript 09).
- In the UK there's about three hundred women boxing as amateurs (tapescript 13).
- So I got into the right-hand lane ready to turn off the motorway when this car screeched up next to me, pulled in front of me and slammed on his brakes (tapescript 14d).
- Suddenly I realized that I couldn't feel the bottom and the sea was pulling me further away from the beach (tapescript 14 e).
- I suppose I'd better ring the bus company (tapescript 24).
- Apart from clothes, maybe a painting from one of the artists who sell their stuff along the river (tapescript 41).
- My gran sent me a beautiful tartan cashmere scarf for Christmas and I love it (tapescript 49-8).

NEW HEADWAY

- Talking about keeping in touch with people a long way away, I'm still very impressed by planes, actually, especially jumbo jets. (tapescript 6 M).
- The ways we can communicate with each other, how much and how quickly (tapescript 6 P).
- My grandfather had two sons from his first marriage (tapescript 21-4).

2. Diphthong |eɪ| is realized as |aɪ|

So-called "broad diphthongs" are typical of Cockney accent and some phoneticians believe that it is also a feature of Estuary English, borrowed by it from Cockney. This peculiarity, just like the one described above, is mostly heard in conversations and interviews.

ENTERPRISE

- I was waiting for the number 6 when I noticed an old lady who started to cross the road (ex. 10 p. 48, a witness of the road accident questioned).
- I've never been to horse races before so when my uncle offered to take me I jumped at the chance (Listening task p. 64, a man talking about horse races).
- Because it was Susie's birthday (ex. 13a p. 73, a person commenting on the restaurant).
- All of their pizzas are home-made (ex. 13b p. 73, a person commenting on the cafe).

INSIDE OUT

- My mother hasn't spoken to me since the day we got married (tapescript 09-2).
- But we've moved away from our home town (tapescript 09-2).
- That's amazing! So, anyway, did you meet a girl in every country you visited? (tapescript 40).
- We're open from nine to five every day (tapescript 46-2).

NEW HEADWAY

- I play football, volleyball, tennis and table tennis but volleyball is my *favourite* game as it's a team game and you could play it with your friends and enjoy it as a team (tapescript 15-3).
- She does not have to work on Thursdays and Fridays but she has loads of different things to do in a day (tapescript 25).
- We won't be able to carry it all (tapescript 28 b).
- So anyway I said to him that I really didn't think it was right to change the arrangements without letting everyone know (tapescript 31-1).

3. Consonant |t| is glottalised

The glottal stop seems to extend into ever more phonetic environments. It was first recorded in the speech of educated people in the mid twentieth century when it started to replace the traditional t-sound at the ends of a word before a consonant [3]. Now the sound is encroaching RP and can be frequently found in all positions in a word: before a consonant (let

me), before a vowel (get over), before a pause (street), in the middle of a word between vowels (letter) [1]. In fact glottal stop has become so widespread that this peculiarity should be described as the one on the verge of RP, probably with the exception of using a glottal stop between two vowels, which is still considered to be Cockney pronunciation. In the analyzed recordings the glottal stop was used mostly at the end of a word before the consonant or before the vowel.

ENTERPRISE

- Yes, but the prices are a bit high, aren't they? (ex. 17 p76, a man talking about a restaurant).
- When I went into a shower block there was no hot water (ex. 11 p. 23, a man complaining about a hotel).
- It was awful. The flames were leaking at my feet. I was trapped (ex. 3 p. 47, a man describing a fire).

INSIDE OUT

- We met working there (tapescript 03).
- I drink two glasses of milk every morning and a cup of hot water in the evening (tapescript 06-3).
- I found it a bit difficult to get into, but now that I'm past the first few chapters, I can't put it down (tapescript 07)
- Do you believe in love at first sight or do I have to walk past you again? (tapescript 10).
- I saw a documentary about women boxing in America about six years ago, a little television programme (tapescript 13).
- Guess what (tapescript 24).
- What about you? (tapescript 44).
- My favourite thing is my black leather jacket which I bought in America (tapescript 49-3).
- I spend quite a lot of money on clothes (tapescript 49-5).
- My favourite outfit is my baggy trousers and check shirt (tapescript 49-7).
- You are not a teenager yet so you are not annoyed and uptight about everything (tapescript 54).

NEW HEADWAY

- Well, first of all you need the ball and the net and obviously the court but you may also need knee pads (tapescript 15-3).
- We just wanted to say thank you for putting us up before we caught the plane last week (tapescript 43).
- I promised to do it carefully (tapescript 44-2).
- Please don't forget to post my letter (tapescript 44-6).

Glottalisation of |t| provides a kind of transition from regional forms of pronunciation to those which are on the verge of RP. These are the most influencial tendencies which have practically occupied their place in the standard pronunciation, but educated speakers are still inconsistent about using them thus making two variants coexist. Among such changes the following seem to be most widespread:

4. Monophthongization of the diphthong |uə|

According to J.C. Wells, RP speakers have been inconsistent about monophthongization of this diphthong for a long time but now it has become rather obvious [3].

Still, even though this diphthong is not used as much as the others and many English speakers use |o:| instead, it is considered important for foreign learners to be aware of this diphthong because it distinguishes words like *moor* and *more*, *poor* and *paw* [2].

The range of words affected by this change in pronunciation seems to be rather narrow: (*sure, poor, moor, tour*). In all analyzed materials both diphthongized and monophthongized forms can be heard. Here are the most typical examples with |o:|:

ENTERPRISE

- On behalf of Boomerang tours welcome aboard (ex.1 p. 27, an announcement at the airport).

INSIDE OUT

- My mother's in New York and my father's often on tour (tapescript 01).
- Just think about those poor little animals (tapescript 35)

NEW HEADWAY

- Do you enjoy going on tour? (tapescript 14).
- You mentioned you played in tournaments (tapescript 15-3).
- Poor you! (tapescript 21-2).
- You must make sure you eat well (tapescript 27b).

5. Monophthongization of triphthongs

Most linguists agree that in present-day English the extent of the vowel movement is very small, except in very careful pronunciation. The middle part can hardly be heard in rapid speech and the resulting sound is difficult to distinguish from a diphthong or a long vowel. In fact there is much variation in the amount of vowel movement depending on how slow and careful the pronunciation is. It seems that triphthongs in RP pronunciation are in a rather unstable state and a change in the phonemic system of RP is in progress in this area [2]. In the analyzed examples the change affects mostly triphthong |auə|:

ENTERPRISE

- Today I'm going to talk about acid rain and a damage it is causing to our planet (ex. 8 p. 111, a professor talking about acid rain).
- These areas of the globe have the most cars and power stations which produce harmful chemicals (ex. 8 p. 111, a professor talking about acid rain).

INSIDE OUT

- I was just fascinated from there on, from like a half an hour programme (tapescript 13).
- We helped each other with our studying as well (tapescript 32).

NEW HEADWAY

- We take our dog on long walks (tapescript 12).
- The flight takes nine and a half hours (tapescript 16).
- There was no heating apart from one little electric fire for the whole place. In fact there was only one power point, so you had to make a choice (tapescript 23).
- We get some money from our parents but it's never enough (tapescript 25).
- And then there's our honeymoon (tapescript 34).
- We never had any problems ourselves at all (tapescript 46a).

6. Phoneme |æ| is realized as |a|

It has been observed by many phoneticians that the quality of |æ| becomes more open, making it more similar to |a|.

ENTERPRISE

- I mean, can you believe it? (ex. 11 p. 23, a woman complaining about bad hotel service).
- The ones on the upper floor have small balconies. Shall we go upstairs? (ex.1 p. 27, estate agent trying to sell the house).
- I was really happy when in the last chapter she found him (Listening task p.38, a woman speaking about the book she read)
- Where would you fancy going? (Listening task p.77, a man inviting his girlfriend out).
- As you know this morning's lecture is about acid rain and its effect on the environment (ex. 8 p. 111, a lecturer introducing a guest professor).
- A boat was reported to have hit the rocks after running into bad weather (ex. 5 p.48 a news report)

INSIDE OUT

- You are not supposed to run away from animals because they can sense your fear (tapescript 14b).
- Not too bad (tapescript 23).

NEW HEADWAY

- Oh, it' not too bad then (tapescript 41).
- She agreed to lend us her flat while she was away (tapescript 44-11).
- …A wedding invitation for next Saturday. … Without doubt some of the happiest times of my life have been spent in our kitchen (tapescript 47).

7. Sound |j| is omitted in the combination |j+u| after certain consonants

The list of consonants after which this elision takes place usually includes |n, l, s, z|. In the analyzed recordings there is indeed a strong tendency for yod dropping after |s| before |u:|, in words like *assume*, *suitcase, supermarket*. As far as the word *suit* is concerned the form with |j| seems to be on its way out.

At the same time omission of |j| can take place after such consonants as |d| (produce) and |t| (stupid, attitude), though typically these sounds are associated more with yod coalescence that yod dropping.

It is believed that such developments involving sound |j| have strong resemblance to American pronunciation. Some British phonetician argue howev-

er that |j| is lost under the influence of Estuary English, especially elision of |j| after |n| (new).

ENTERPRISE
- The novel I intended to read during the flight was packed at the bottom of my suitcase by mistake (Listening task p.38, a man talking about the book he read).
- The book itself was beautifully produced (Listening task p.38, a woman talking about the book she read).

INSIDE OUT
- He is mean, big-headed and stupid (tapescript 11).
- In some countries, for example, in Latin America, there's a more relaxed attitude (tapescript 36).
- The new computer game you ordered has come in (tapescript 46-3).
- I only wear it on very special occasions and I usually wear it with a suit (tapescript 49-2).

NEW HEADWAY
- Are you wearing a suit? (tapescript 5).
- We just looked at each other and knew that we couldn't stay (tapescript 23).
- We have to wear a stupid school uniform (tapescript 25).
- .Would you mind looking after newspapers? (tapescript 31-4).
- We met our friends, Bill and Sue, and they invited us to have a meal with them (tapescript 43).
- And some supermarkets are open twenty four hours a day (tapescript 46a).

8. Yod coalescence: |t+j+u| and |d+j+u| are pronounced as affricates

More and more words are pronounced with affricates. In our analysis this feature is especially noticeable in words like *situation, reduce, during*. Still, it is important to stress that different speakers pronounce such words differently. In fact three pronunciations coexist: the one where all three elements of a cluster are pronounced, the one with yod dropping and the one where the first two elements are blended into an affricate. Here are the most typical examples of the latter type:

ENTERPRISE
- The book was set in France during the First World War (Listening task p. 38, a woman talking about a book).
- Pilots need to be trained to deal with dangerous situations (Listening task p. 52, a radio interview about aircrashes).
- First class compartments are situated in the first two carriages (ex. 5 p.147, an announcement at the railway station).

INSIDE OUT
- I was born in London but I've got dual nationality because my mother's from Nicaragua (tapescript 01).
- I've never had a massage or a nap during the day (tapescript 06-1).
- Have you ever been in a dangerous situation? (tapescript 14).
- Congratulations! (tapescript 23).

NEW HEADWAY
- And then at the end of the day there's the evening, food and wine with friends and everybody talking about their excitements during the day (tapescript 15-1).
- I play at local sports centers more during the winter (tapescript 15-3).

These are but a few peculiarities of the speakers' pronunciation in the audio materials analyzed, but being aware of these peculiarities is extremely important for foreign learners of English. As our experience of using *Enterprise, Inside Out and New Headway* recordings in the classroom indicates, the kinds of pronunciation described above often cause difficulties in understanding speech. However it doesn't mean that they should be excluded from the teaching process. Just the other way round, we fully support the idea that exposure to different accents is useful for ear training in the receptive skill of listening comprehension. So the main implication on the stage of teaching speech perception is for the teacher to familiarize the learners with the existing variants of pronunciation.

As far as teaching productive speech is concerned it is important to avoid things that can most impede intelligibility while encouraging fluency and confidence [4]. Regional features can hardly be recommended for foreign learners of English to imitate, so substitution of |ʌ| with |u| or |eɪ| with |aɪ| is out of the question. Speaking about |t| glottalisation, is is rather difficult to say whether students should be encouraged to use it in the phonetic contexts in which it is frequently used by some RP speakers.

We'd rather leave it at the stage of recognition, the more so that it is not registered in most dictionaries yet as one of the possible variants of pronouncing words with |t| sound. The same is true about realisation of |æ| as |a|. Since it is not recorded in the dictionaries in most cases, there is no need for foreign learners of English to try and copy it. The other four peculiarities involving monophthongization and combinative changes with |j| sound seem to be more appropriate for active usage by more advanced learners in rapid casual speech. Such pronunciation can make their speech sound more native-like and modern.

Summing everything up it is important to stress, that even though the RP of recent years is characterized by a greater amount of permissible variants compared to the classical type of general RP, the notion of standard pronunciation still remains crucial for teaching pronunciation and incorporating new developments into teaching productive speech should be careful and selective.

REFERENCES

[1] Kerswill, P. 2000. Mobility, Meritocracy and Dialect Levelling: the Fading (and Phasing) out of Received Pronunciation. *British Studies in the New Millennium: Challenge of the Grassroots. Proceedings of the 3rd Tartu Conference on British Studies*. Estonia.

[2] Roach, P. 2004. *English phonetics and phonology*. Cambridge University press.

[3] Wells, J. C., 1997. Our changing pronunciation. Transactions of the Yorkshire Dialect Society, 42-48.

[4] Wells, John, 2005. Goals in teaching English pronunciation. In Dziubalska-Kołaczyk, K. & Przedlacka, J. (eds), *English pronunciation models: a changing scene*. Bern: Peter Lang. pronunciation models: a changing scene. Bern: Peter Lang.

VARIATION OF RHYTHMIC PATTERN IN AMERICAN SPONTANEOUS SPEECH

Yuri Artyushenko

Amur State University, Russia
Yorkman84@mail.ru

ABSTRACT

Prosodic features of spontaneous speech appear in the center of linguistic research quite rarely. Rhythm is no exception to that. The following report is concerned with the peculiarities of rhythmic composition of spontaneous monologue due to variations of intensity and duration. The results of acoustic study demonstrate a lot of violations of isochrony of rhythmic pattern in American spontaneous speech, which are much sharper compared to prepared monologue.

Keywords: rhythm, spontaneous, variation, intensity, duration.

1. INTRODUCTION

Linguists give many interpretations of such a phenomenon as rhythm. The most widespread among researchers is the point of view according to which the term "rhythm" is used as a synonym of a regularity or tendency to regularity. The speech rhythm is traditionally defined as repetition of stressed syllables in a more or less equal time intervals in a speech continuum. More detailed definition of rhythm treats it as regular alternation of acceleration and delay, a relaxation and strengthening, a longer and shorter durations, similarity and distinctions of elements in speech. The rhythm, from the point of view of linguistics, is realized by lexical, syntactic and prosodic means and their combinations. For example, such phenomena of speech as repetitions of certain sounds and words, intensification etc. are perceived at all these levels.

2. SPEECH RHYTHM

In speech, rhythm type depends on the kind of a language. Linguists allocate two kinds of languages depending on their rhythmic organization: syllable-timed, such as French, Spanish and other languages of Romance group, and stress-timed languages, like languages of the German group, for example English and German. In syllable-timed languages a speaker spends approximately the same time on each syllable, no matter if it's stressed or not.

In stress-timed languages the rhythm is based on a unit larger, rather than the Syllable. Though the quantity of time spent on each syllable differs considerably, the general time of utterance of each rhythmic unit is almost invariable. There is a tendency to pronounce stressed syllables with regular intervals not depending on the number of unstressed syllables between each two stressed ones. The regularity is provided by so-called accurate "beats".

It is necessary to mention that the speech rhythm has influence on vowel reduction and their elision. Such word forms as prepositions, conjunctions, and also auxiliary and modal verbs, personal and possessive pronouns are usually unstressed and are pronounced in their weak forms with a reduction or even elision of vowels or consonants for maintaining the equality of intervals between stressed syllables.

It is considered that the basic rhythmic element in English speech is the rhythmic group, a segment of speech which contains a stressed syllable with previous and-or subsequent unstressed syllables attached to it. The most widespread kind of rhythmic group includes 2-4 syllables, one of which is stressed, others - unstressed. The stressed syllable is a prosodic nucleus of a rhythmic group. The initial unstressed syllables preceding to a nucleus are called proclitics, and those which follow a nucleus are enclitics. There are two various approaches of phoneticians to qualification of unstressed syllables located between the stressed ones. According to the so-called semantic approach, unstressed syllables are drawn to stressed syllables of the same word or to a lexical unit, according to their semantic link.

According to another approach, unstressed syllables which are situated between stressed, are inclined to join the preceding stressed syllable. This is the so-called enclitic tendency.

The more the speech is prepared, the more rhythmical it appears to be. For example, the text uttered aloud in the form of a lecture is more rhythmical, than conversational speech. On the other hand, the

rhythm is individual – the speech of a person who speaks fluently, is more rhythmical, than the person searching for a proper word and correcting the speech in the process of speaking.

According to the data offered by Kohler [Kohler 2009: 32], rhythm should be studied from four various points of view for definition of its nature and functioning in speech communications: symbolical representation, reproduction, perception and communicative function. Besides, global temporal differentiation of a speech signal into parts which has repeating phonetic characteristics under and over the syntactic and semantic organization, but nevertheless cooperating with it, is essential to a rhythm in speech of any type and in any language.

While studying speech rhythm the most important question that arises is one of the functions which are carried out by rhythm in speech. The speech rhythm is multifunctional. Interpretation of speech rhythm completely depends on the function that is considered the basis of definition of certain speech phenomenon, proposed by this or that researcher. Rhythm serves for arrangement of elements of speech: smaller units will be organized in bigger ones, the bigger ones will include the smaller ones. On the one hand, rhythm unites certain segments into a coherent utterance. On the other hand, it enables to split a coherent utterance into smaller units. This integrative and delimitative rhythm functions reflect dialectic unity of its opposite displays.

At the linguistic level, the pragmatic value of a speech rhythm is realized in its function of expressing volition due to which the rhythm is capable to express various levels of emotional influence on the listener. It is necessary to stress that prosodic elements, together with lexical and syntactic means work as components of a rhythm. The rhythm itself functions as the structure of speech organization and is the effective means of speech expression. This is the function which we also are going to deal with using the material of spontaneous speech.

The speech rhythm is in many respects defined by biorhythms of a human body and the nature surrounding us. As a rule, many forms of speech rhythm are not realized by the speaker, but subconsciously everyone feels the rhythmic organization of a speech stream and aspires to recreate its each time, though sometimes in a slightly new, but, nevertheless, similar form.

We start to realize periodicity only when it is broken or when this periodicity gets a strict character in respect of similarity of the repeating phenomena, and in respect of their commensurability.

A systematic character of the phenomena assumes presence of interconnected units. Any speech segment can carry out functions of rhythmic units if these segments are similar, commensurable and periodically repeated in the text or a text part. Rhythm is a difficult language system and its elements are organized in a hierarchical order. They represent hierarchy of functional lines, or in other words, this system includes well organized elements of a different size in which smaller rhythmic units join more complicated ones: rhythmic group – syntagma.

Large rhythmic units possessing certain semantic completeness, represent a certain complete piece of speech in which integrity is transferred by lexico-syntactic and prosodic means. These units are a subject of the analysis of syntax, semantics and prosody. The same units are capable to carry out a rhythm - forming role. The multifunctionality of the listed above segments is shown in it.

The speech rhythm possesses all system characteristics, that represent a certain set of the interconnected elements (rhythmic units) which form a unity. The present unity is defined by unity of functions. Each element of a system (rhythmic unit), in turn, represents an independent system. In a rhythmic system dual quality of elements is also found: inherent in the element (rhythmic unit) and the system itself. Not all elements of rhythmic system are homogeneous in respect of revealing their independent qualities. Large units possess the properties inherent in semantic units which they never lose. At the same time they can carry out functions of rhythmic units equal to the units which do not belong to the category of semantic. Smaller rhythmic units (sound, syllable, rhythmic group) in a larger degree are "suppressed" by a rhythmic system meaning that in case of entering the system, rhythm-forming becomes the basic function for them.

As far as speech rhythm is largely based on semantics I hypothesize that in spontaneous speech prosodic isochronism that is said to be characteristic of rhythm will be sacrificed to a great extent.

2. EXPERIMENT

My pilot experiment deals with the rhythmic organization of spontaneous monologue of American English. Spontaneous speech, unlike prepared, has a number of the features which do not correspond to the standard concepts of the content of the rhythmic organization of speech, for example: the rhythm of spontaneous speech has a big dependence on the emotional state of the speaker at the moment of pronouncing, his being interested in the

topic of a discussion and can becharacterized by a considerable quantity of pauses and hesitations. The purpose of our experiment is to reveal timing of stressed syllables within the limits of every pausal group and correlation of the received results with the duration of every pausal group. Pausal groups have also been divided into two types long and short depending on their duration and the number of the uttered stressed syllables. To begin with we took a fragment of a spontaneous speech of one speaker. The total number of pausal groups is 20 . There is an opinion in the linguistic literature that spontaneous speech has no certain rhythmic outline, unlike a prepared one. So I decided to measure the duration from one stressed syllable to another. Having taken for the minimum unit a pausal group which duration is not less than 1 second and includes not less than 2 stressed syllables on a fragment of the text without hesitations, at first we, however, have revealed the following tendency (see Table 1).

Table 1: Rhythmic pattern of spontaneous speech.

Pausal groups (PG)	A number of stressed syllables in a PG	Average distance between stressed syllables
1	5	0,742
2	5	0,408
3	5	0,572
4	6	0,657
5	5	0,375
6	4	0,480
7	5	0,500
8	8	0,340
9	2	0.493
10	13	0.188
11	4	0.232
12	6	0.161
13	11	0.283
14	4	0.263
15	3	0.370
16	4	0.285
17	3	0.313
18	4	0.274
19	5	0.295
20	3	0.237

If we adopt the following symbolic notation

L - the maximum duration (Long) between stressed syllables in one pausal group

M - medium duration

S - the small duration between stressed syllables

then in the first three pausal groups it is possible to express the syllables duration pattern withing the pausal group as L – S – M.

Hence, the rhythm of a spontaneous speech, from the point of view of distance between stressed syllables, should be characterized by certain isochrony (repeatability through identical time intervals). However the research showed that one and the same rhythmic pattern is not exploited by the speaker all the time. There are constant shifts in rhythmic pattern. 1-3-2 pattern is exploited in evry block of 3 PGs following each other up to PG 10 where the shift occurs. From PG 11 and up to 20 inclusive we have to adopt another symbolic notation: just Larger (Lr) and Smaller (Sr) would be enough to express that rhythmic pattern – Lr-Sr – which is repeated 5 times up to PG 20 (Table 1).

3. CONCLUSION

The results of the measurements and comparison can not confirm the tendency for the length between the stressed syllables in different parts of a spontaneous text to be approximately equal. corresponding to certain parts of the text: at the beginning the average length between stressed syllables is rather large (~ 0,507), as the subject went on speaking it becomes shorter (~ 0.216). The more so, the number of stressed syllables in one pausal group has increased too. We are inclined to think that an average distance between stressed syllables here depends not only on the emotional side of the speech but also on the speed of ideas the speaker gets during the process of speaking.

4. REFERENCES

[1] Kohler K. J. 2009. Rhythm in speech and language. Phonetica, 66 (1-2), 29-45.

[2] http://grammar.about.com/od/essaysonstyle/a/rhythmofprose.htm

[3] http://content.karger.com/ProdukteDB/produkte.asp?Aktion=BackIssuesJG&searchParm=2009&ProduktNr=224275

[4] philol.msu.ru›~ruslang

PRAGMATIC EXPLOITATION OF PROSODIC FEATURES

Mark Campana

Cobe University, Japan
bugsofjapan@hotmail.com

ABSTRACT

The paper deals with the way speakers of different languages attach secondary, non-linguistic meaning to their utterances through the use of prosodic features, initially focusing on those encoded in the phonological system of the language. In English and Russian, where stress plays a distinctive role at word level, implied meaning is often signaled by variations of loudness. In Japanese, where segment length has been grammaticized, prolongation can have the same effect. Changes in pitch may also be exploited in Chinese for the same purpose, a reflection of the tones that serve to classify lexical meaning. Non-grammaticized prosodic features are exploited for pragmatic purposes too. Thus in English (and presumably Russian), pitch and/or lengthening can also serve mark affective stance (e.g. 'very í::nteresting'). The distribution and frequency of such features are discussed, as well as the different types of implied meaning they convey. These may pertain to the speaker's judgment, volition, perception or claims to knowledge--roughly states corresponding to the mental predicates Wierzbicka (1999) proposed for a Natural Semantic Metalanguage. The latter, have their correlates in sound, from which speakers chose to implement their intended meaning. The survey includes samples of recorded speech from different language types, including stress-timed (English, Russian), syllable-timed (French), mora-timed (Japanese) and tonal (Chinese).

Keywords: prosody, meaning, loudness, prolongation.

1. INTRODUCTION

First I will address the aims of this conference, specifically " ...sharing in the enrichment of knowledge [regarding] acoustic and perceptual analysis." First, while I am not a phonetician or phonologist *per se,* I began my linguistic studies in these fields, and apparently this conference is where they end. Allow me to explain...

1.1. History

From my first teacher, André Rigault (a phonetician) I learned that all one needs to do linguistics are a pencil and a piece of paper. Phonology taught me that puzzles could be solved by applying the right theory to a set of data, mimicking ostensibly mental constructs in the human mind. This viewpoint is the same as that espoused by Chomsky and his followers for syntax too, and for many years I was concerned with finding the best way to explain patterns of word- and sentence structure in non-Indo-European languages. This path led to an an encounter with discourse particles in a polysynthetic language, which ultimately convinced me that no matter how far-reaching a theory of syntax was, it wasn't going to capture the extemporaneous meaning of these elements. One got the feeling that, despite the insight gleaned from syntactic structures, they are merely distillations of a more dynamic processes than envisaged by generative linguists. Thus I came to study talk-in-interaction and pragmatics, where contextual meaning is applied, and things like discourse particles better understood.
Once touched by the simplicity and elegance of Gricean logic, it was only a matter of time before considering how exactly a listener is able to infer the very thing which the speaker had implied, when in fact other choices were available. As it happen, a proper explanation cannot exclude the sound of the speaker's voice in the execution of the utterance, along with other presentational details. And so, my fellow phoneticians, we have come full circle.

1.2. Background assumptions

Consider next how the acoustic bears on the perceptual. Put somewhat differently, How does prosody relate to pragmatics? There is no doubt that it does--a vocal display of emotion will quickly draw attention to the speaker regardless of the words that issue forth. The listener is then left to contemplate the implications of this display, scanning the signal for patterns that would account for the most plausible explanation. This is how the systems works; one of my goals today will be to recast it in the light of some new trends and observations.
Pragmatics became an established field through the confluence of several different approaches to language: logico-semantic; performative, politeness,

etc. (cf. Yule 1999). The literature is truly vast, but, inasmuch as I can tell, much of it is argumentative, i.e. what empirical data does exist consists of anecdotes and/or examples constructed by the practitioner to illustrate a point. In short, very little data is taken from naturally-occurring conversations, let alone the sound of them. Clearly, this omission should be remedied in the future (but cf. Reed 2006).

At the same time, great progress has been made in the field of acoustics, where digital technology has enabled researchers to study in detail the properties of sounds, as well as to construct new tools for analyzing them. In my experience, however, this progress has taken place with little or no awareness of what cognitive scientists (pragmaticians) have been up to. Research on human emotion is a prime example of how two camps can focus on the same phenomena, oblivious to the activities of the other. One camp has very little to say about the sounds that carry an emotion, and the other is seemingly uninterested in how cognition governs it. Thus, another goal of this talk is to bridge the gap.

1.3. Prosody defined

What exactly is meant by prosody? There are in fact two kinds, but the one mostly used by linguistics (linguistic prosody) is concerned with phonemic differences, e.g. stress or stress-accent (as in stressed syllables, but also sentence stress), tone (pitch), length (as in long vs. short vowels) and intonation. When it is phonemic, it means a language has incorporated the feature. But the wider definition includes all the vocal sounds that speakers make. The inventory is essentially the same, but the features may not be incorporated.

The fact that languages incorporate prosodic features at all means that we (humans) are aware of them--unconsciously in daily use, but also sometimes consciously in creative use or language play. Their significance is thus psychological. Moreover, that different languages incorporate different p-features means that all of them are universally available.

1.4. Language processing

The ability to produce and perceive (unincorporated) prosodic features is a complex task, but one which humans are apparently well-equipped to handle. This may be facilitated by the specialized neurons that mirror the actions of others. The speech-role of such "mirror neurons" is still ongoing, but their discovery has far-reaching implications. Although initially shown to mirror the motor skill of grasping (cf. Rizzolatti et al 2007), they are likely to be involved in speech processing too, not only when the word for e.g. *grasp* is used, but also in empathizing with (interpreting) the affective displays of others.

Mirror neurons inform our understanding of the relationship between planning, producing, and perceiving language too. They allow speakers to monitor their own speech, or act as 'first listeners'. They also enable hearers to imagine how *they* would feel if *they* were producing the stream of affect emanating from the speaker. In short, they provide the neurological template for processing prosody.

One of the core assumptions of this talk is that "talk" consists of two streams of information, but only two. Most researchers would acknowledge the two: one, which we may call 'lexical', is made up of words and their meanings, syntax and *its* meaning (relevant to speech acts and information structure), and logical forms. This type of information is mainly processed in the left hemisphere of the brain. The second stream contains affective information and its concomitant prosodic features, processed in the right hemisphere. The reason that we process only two streams of information is that we only have two hemispheres (cf. *Science,* v. 328).

**It could be argued that at least one other mode of communication is available to the speaker, namely gesture. First, the reference cited doesn't hold that humans can't perform more than two tasks at once, only that their efficiency is compromised (they're just not as good at it). Second (and more importantly) prosody is inherently gestural, and could plausibly have evolved from meaningful displays of motor skills.*

Human cognition is such that, given two of anything, interlocutors will attempt to find a relationship between them. Perhaps this follows from the need of brain cells to better support themselves (through synapses), hence ensure survival. In any case, the simultaneous production of two streams of information—one lexical, the other affective—forces the listener to come up with a third, derived set of meaning that accommodates the juxtaposition of the first two.

2. DATA AND METHODS

In this section, we introduce specific research questions, as well as some of the assumptions that lead up to them. First, we consider the notion of a pragmatic moment. This involves both the speaker, who intends to impart bits of (mainly) non-lexical

meaning, and the hearer, whose job it is to understand them. Various typed of speaker-meaning can characterize a pragmatic moment—deictic, presuppositional, polite—but for this talk we will concentrate on implicature in the Gricean sense.

2.1. Pragmatic moments

A pragmatic moment would be a point in a conversation where one speaker responds to something that may not follow directly from the propositional content of the previous speaker's turn. The assumption is that the instigator of this event—Speaker A—is somewhat aware of the workload that is necessary to 'unpack' the pragmatic meaning of the utterance, and cues Speaker B to be on the lookout for something special. These cues are acoustic in nature (based on intensity, frequency, volume and/or timing), but they are not the only possible triggers (gesture, and of course lexis also serve this function).

How does one identify a cue? While not 100% reliable, some cues can be identified through an acoustic analysis of the recorded message and cross-checked during playback by the interlocutors themselves. Strictly speaking, a cue entails some deviation from a norm, which is itself a subjective construct. Every speaker has a normal speaking rate for the time and circumstance in which they're speaking. Still, participants usually display confidence in their ability to gauge the normal rate of others, and they do take note of deviations from it. Presumably this is an unconscious evaluation, but when their attention is directed to the record they can articulate their thought processes. This type of knowledge (what is normal, or unmarked in any particular environment) applies to other acoustic features.

The following are some examples pragmatic moments taken from the literature (cf. Yule 1999).

(1) A: Smith doesn't seem to have a girlfriend these days.
B: He is spending a lot of time in New York...

(2) A: Coming to the wild party tonight?
B: My parents are in town.

(3) A: How's your hamburger?
B: A hamburger is a hamburger.

(4) A: Whoa—has your boss gone crazy?
B: Let's have a cup of coffee, shall we?

From these exchanges we are to suppose that Smith does indeed have a girlfriend (1); that B won't be attending the party (2); that all hamburgers taste the same (3), and that the boss's behavior should be discussed somewhere else (4). What kind of sounds (or features) do the speakers use to help the hearer arrive at these conclusions?

2.2. Framing the questions

Of course, the many sounds (and gestures) that speakers can exploit to signal their intended meaning is not a homogeneous set. Some of them are decidedly paralinguistic, not unrelated to the calls of Vervet monkeys. Others are prosodic in the general sense, i.e. related to the 'tone-of-voice' or speech quality that the speaker may assume. Then there is prosody in the narrow (linguistic) sense, pertaining to features incorporated in the lexicon of different languages. English (and Russian) words carry stress (or pitch-accent), for instance, whereas words in Chinese specify tone. Japanese speakers, on the other hand, make use of various timing patterns: long & short vowels and consonants, full- or partial reduplication.

While the array of vocal sounds is quite large, the discussion here will concentrate on the ones that are better known to linguists. The research questions take the following form:

(5) To what extent are incorporated prosodic features used to convey pragmatic information?

For English (or Russian) the question would be about stress, or pitch-accent, since this represents the incorporated prosodic feature. For Chinese it would be about tone/melody, and for Japanese it pertains to timing units (phonological and/or morphological).

At the same time we may also ask whether (and to what extent) paralinguistic, or affective prosody is used independently of incorporated prosodic features.

(6) To what extent do unincorporated p-features convey pragmatic information?

Although we may ask the questions, some of them must remain unanswered for the time being. This has as much to do with the state of our tracking technology as it does with our knowledge of how the tracking process works.

3. ANALYSIS

Good data is something that can be quantified. One way is comparing the number of times a specific

sound is adjudged as 'highly affective' with other sounds which are not. In the case of Japanese, we focus on those moments where rhythmic perturbations in the speaker's utterance frequently trigger a pragmatic response in the listener, that is, a reply that does not completely follow from the previously-spoken words.

Natural conversation provides the best data, but what makes a conversation natural? Those involving just two speakers are representative. The degree of familiarity between the speakers is extremely important, assuming that shared knowledge plays a large part in the recovery of intended meaning. For the purposes of our study, it was not necessary to document the exact nature of relationships, since the participants themselves were consulted as to what (they thought) transpired during the exchanges.

3.1. Pilot survey

Some of the data consisted of recorded conversations between native speakers of Japanese. The recordings were made in MD format and examined using PRAAT. Key exchanges were selected as examples of perceived implied meaning, where one speaker might reply in a way that is inconsistent with preferred responses to lexical material. The following English conversation is typical.

(7) Speaker A (distracted): Now where did I put my keys?

Speaker B (mildly offended): How should I know?

Speaker A (somewhat chastised): Take it easy. I'll find them.

Far from subtle, this exchange is a good example of how prosodic features bear on affect, which in turn colors the response. Something in or about A's utterance causes B to take offense ("Why are you asking me to listen?"). Something prompts A to discern offense (sharp rising intonation in B's voice). These data contain specific audible cues to which the following speech-act is addressed, as in exemplified in Wichmann (2000).

3.2. Analyzing the sounds

For this project, the voice samplings of speakers were separated into tracks that follow rhythm, pitch and other prosodic units, and a base mean was established for each span. Any deviation from a norm provided an opportunity to impart pragmatic meaning, and the record often shows this. Following the procedures of Tannen (2005) and others, participants were then asked to hear their recorded utterances and answer questions like "As a listener, why did you respond *that* way?" (e.g. drawing attention to a prosodic peak) or "As a speaker, why did you modulate your voice *here*?" Answers were typically of the form "Because I thought [the speaker] meant X" (where X=pragmatic meaning), or "Because I wanted [the listener] to know Y." Over 40 such exchanges were collected where the audio record was analyzed and replayed for participant feedback.

Elicited samples of lengthened sounds in Japanese were also collected, analyzed and presented to native speakers for further evaluation. The length of certain segments was controlled (e.g. the delayed release of geminate stop consonants), and listeners were asked to judge the degree of perceived affect (generally speaking, the longer the release, the more affective).

3.3. Japanese

If acoustic/prosodic features are in part responsible for communicating affect, which ones do Japanese speakers use? Here we argue that speakers use duration to signal pragmatic information to their listeners. This property is firmly entrenched in Japanese phonology in the form of vowel and consonant length (geminates).

(8) Vowel & consonant length

a) i**s**ho 'a will'; i**ss**ho 'together'

b) ka**t**a 'how to'; ka**tt**a 'bought

c) ha**k**a 'grave'; ha**kk**a 'ignition'

d) k**o**t**o** 'thing'; k**o:**t**o** 'coat'; k**o**t**o:** 'desert island'; k**o:**t**o:** 'high, advanced'

Vowel and consonant length are thus central to the phonology of Japanese. And yet, these features are just convenient points from which speakers choose to depart in the display of affect and other types of intended meaning. From an articulatory point of view, it's a fairly simple matter to prolong a vowel or nasal consonant. Stops, on the other hand, have a built-in a pause before release. What easier way to insert a timing unit for pragmatic effect than to hold it longer?

As outlined above, we assume that sound (like meaning) can be judged by speakers as 'normal' or 'remarkable', along Gricean lines of reasoning or the 'politic conduct' as articulated by Locher (2004), Watts (2003). In terms of language processing, neural networks monitor certain aspects of the human voice, including emphasis/intensity, pitch/frequency, timing units and intonation (thus it comes as no surprise that they would be grammaticized as prosodic features in one language or

another). When an unexpected variation is perceived for any given property, the listener will assign some kind of value to it from his/her own 'repertoire of reasons'. In the case of Japanese, these perturbations are often rhythmic.

Anecdotal evidence suggests that rhythmic variation is also used to project an end-of-turn in Japanese. This makes sense from a language design point of view, but deserves a separate investigation that cannot be offered here.

Prolongation is also used extensively to signal pragmatic information (mainly affective), as the following examples show:

(9) Sonorant lengthening

A: So**nn:**a ni watashi no koto ga kirai na no ('I don't go in for *that* sort of thing').

Compare: So**nn**a ni watashi no koto ga kirai na no

B: S**o::**yu imi janakute ('That's *not exactly* what I had in mind')

Compare: S**o:**yu imi janakute

In A's utterance, the nasal consonant of the first word is extended, resulting in a deictic element invested with the speaker's (attitudinal) perspective. B replies with another lengthened segment, this time a vowel. The pragmatic effect is similar, but crucially absent when the vowel is not prolonged. Consider the following exchange involving obstruants (glottal stop inserted):

(10) Obstruant lengthening

A: Yamada-san ni mail o okuttemo henji ga konai-n dakedo

'I keep sending (e)mail to Yamada, but still no reply'

B: Sotsuron no shimekiri ga chikai kara to**'t**emo isogashi mita daiyo

'The graduation-thesis deadline is getting close so he's *really* busy'

In B's answer, the release of the stop consonant internal to *to**t**emo* 'really' has been delayed. Our claim is that B does this to convey some added information, such as "Don't even bother trying to get ahold of him." This of course remains for A to work out.

From these examples it is apparent that Japanese uses timing mechanisms such as segment extension (along with pitch and intonation) to impart various types of pragmatic information. This cannot be seen as accidental, since Japanese is a mora-based (timing) language.

At the same time, reduplication—seen here as a lengthening process—occurs in such a way as to induce pragmatic, non-predictable meaning. *Mita-mita* (see+Pst) is emphatic, as in "I really did see it" whereas *kirei-kirei* (beautiful) often has an opposite meaning, as in "I acknowledge that *you* think it's beautiful, but … ."

Finally, we present an example of pragmatic exploitation in Japanese which shows how p-features of duration are conserved. One of the more colorful intensifiers in the standard dialect has the structure of a semi-reduplicated form: *mucha-kucha* ('very'), as in *Hito-ga mucha-kucha oi* 'very many people'. Already *mucha-kucha* is long enough to qualify as a self-contained rhythmic element which will stand out from surrounding material, hence draw attention to the speaker. To this s/he may then add various tones or emphatic stress (unincorporated p-features in Japanese) to convey their affective stance. There is a shorter version of this word in colloquial usage, however: *mut'cha* (apostrophe indicates a geminate consonant). Here, although (partial) reduplication has been lost, the potential for prolongation has not. What was once a single consonant onset (second syllable) has become a double one, again allowing the speaker to withhold its release for as long as s/he desires. Thus duration (and role of adding pragmatic information) is retained for this word.

3.4. Caveats

Although we can ask the questions (5-6), the answers may not be immediately forthcoming. For one thing, the technology needed to track deviations in the acoustic record is still being developed (cf. Lichtenstein 2011). To be honest, we don't know what all the relevant variables are. Fundamental frequency (F0) is certainly one of them, but even basic vowel formants (F1, F2) are also subject to speaker manipulation (within certain limits). Still higher frequencies with attenuating ranges of intensity (cf. timbre) are difficult to pinpoint, let alone analyze and match with intended meaning. A similar host of challenges accompanies changes in rhythm, speech rate, syllable length, onset, etc. — any of one which fluctuates from one moment to the next.

Which brings us to a second hurdle that stands before a proper understanding of the questions launched above. The value (and potential significance) of whatever feature we wish to track is only relative to that which came before it—and possibly to the value of other features too. Imagine that a speaker (for whatever reason) releases a stop consonant with the force of X, easily measured in terms of pitch, duration and/or intensity on a spectrogram. X is only remarkable if similar sounds in

working memory have a higher or lower value. Likewise, X may (or may not) be judged as remarkable (hence significant) only in relation to the status of the word as a piece of new or unexpected information. How many variables are there, and which algorithms determine whether (and to what degree) X is remarkable? How deep is a working memory? Clearly, more work must be done to build a proper model.

4. DISCUSSION

At some point, the data must be organized, or fitted to a system that interfaces with the mental representations of the speaker/hearer. No-one knows the full extent of what these are, but clearly affect in its many guises has an important role to play, as well as other psychological states pertaining to knowledge or volition. To this end, we will adopt a cognitive approach based largely on the work of Wierzbicka (1999) called Natural Semantics Metalanguage, or NSM.

4.1. Affect

First we note that affect can refer to basic emotional states such as *anger, fear, joy,* etc., but also to more nuanced ones like *disappointed, miffed,* or *pleased.* Displays of affect—howsoever fleeting—will be limited only by the terms of the Metalanguage and their arrangement into propositions and cognitive scenarios. Since the latter may contain contextual variables, affect properly includes attitudes, where e.g. the speaker signals (elements of) emotion "about something". Affect also subsumes the concept of mood, which can be understood as repeated displays of similar states over an extended period of time (thus it is that a person's mood can be detected not just through the sound signal, but from the repeated use of similar words).

4.2. Emotion and its parts

The NSM was invented as a way of defining emotions across languages and cultures. As Wierzbicka (1999) points out, researchers in fields that deal with emotion (psychology, linguistics) often assume that speakers of other languages experience the same mental states that characterize the researcher's language--when in fact they don't. Rather, their experience is framed in terms of their own lexical expressions. The result is a natural problem for translation, which Wierzbicka proposes to resolve by using only 58 terms, the vocabulary of the metalanguage.

The terms are selected on the basis of their universality. Every language has a word for *want,* for instance, hence is one of the 58. The words are then assembled into simple propositions like "I don't want this to happen" that together make up a 'cognitive scenario'. The following scenario defines the English emotion word *miffed:*

(11) *miffed* ([having been] put into a bad humor)

"X felt miffed" (non first-person arguments bracketed and/or underlined)

a) X felt something [because X thought something]

b) sometimes a person thinks

c) I know now: [someone did something bad to me]

d) I don't want [things like this to happen]

e) I want [other people to know that I feel like this]

f) when this person thinks this, this person feels something

g) X felt [something like this]

h) because X thought [something like this]

The so-called mental predicates of the NSM are of interest because they are marked by prosody; leaving aside *see* and *hear,* we propose that *wanting/not wanting, knowing/not knowing, feeling good/bad* and *thinking* can be detected through their acoustic features.

4.3. Intoning mental predicates

One of the reasons for thinking that these mental predicates have an important role to play in the conveyance of emotion (hence affect, hence pragmatic meaning) is that they can be 'intoned', or stripped away of lexical material, and voiced only by using prosody. Examples of this phenomenon can be easily produced as simple responses to appropriate prompts (declarations or questions) made by the researcher. Consider the following procedure:

(12) The prompter explains to the subject that all responses will involve non-verbal vocalization, as when communicating with a familiar person. (Imagine a scenario in which the subject is lying in bed not fully awake when his/her spouse initiates a conversation. Not inclined to begin speaking, s/he will intone their responses.) The best prompts probably sometimes contain the target predicates themselves. The responses should be "pure tones", i.e. NOT superimposed over actual words or phrases like "yes" or "uh-uh".

- *want/not want*: For this series, the prompter asks easy-to-answer questions e.g. "Do you want to stay in bed?" Food is also a good topic for *want,* e.g. "You want cake?" For *not want,*

something undesirable works best: “Do you want to see my scar?” or “You want to talk about XX?” They can be neutral too, such as “Do you want to watch TV?” Who knows what the subject will say at any given moment?

- *know/not know*: For *know*, make an announcement about something which is already known to the subject, e.g. “It's starting to rain” or “We have to call my mother today”. For *don't know*, the targeted (intoned) response is equivalent to “How about that?” or “Who would have thought?”--a clear acknowledgement of new (and impersonal) information. The best prompt could be a question like “Did you know that …?” There may be a slight tone indicating interest in the response.
- *think* (level of engagement): Note that the contrast here is not *think/not think*, but rather the level of interest or engagement. The subject can be prompted to intone his/her evaluation of a new topic, e.g. “I hear Bush is coming back”. The idea is to get the subject to show interest, e.g. “Oh really” (without too much surprise) or disinterest “Who cares? (without too much negativity). The researcher may have to 'feel out' a topic with their subject first.
- *feel good/feel bad*: The best way to prompt these mental predicates is through the recollection of past states. The researcher asks the subject to tell a story about a happy event (could be something trivial, like what happened at the end of the meeting). S/he then prompts the subject with the question “How do you feel about that now?” Here the answer is potentially complex because there are different types of good feelings. Use the same strategy for *feel bad*.

4.4. Tone-of-voice

Suppose that a unique pattern of tone, rhythm, and timbre is established for each the mental predicates *wanting, knowing, feeling good* and *thinking* and their negative counterparts: *not wanting, not knowing, feeling bad* and null (i.e. *not thinking*). To some extent, such 'signature patterns' would be different for each speaker, but others would almost certainly be shared by in-group members, if not by speakers of different dialects.

This is the genesis of tone-of-voice, whereby the speaker imparts (mainly) affective information relating to the co-/context. As the name suggests, it is essentially paralinguistic in nature, but no less important in the establishment of positioning or stance, hence of politeness.

Informally, tone-of-voice is often associated with negative attitudes such as ‘condescending’, ‘disrespectful’ or ‘sardonic’, but may also project positive ones like ‘confident’, ‘sexy’ or ‘upbeat’. The fact that it reflects such a wide range of (imputed) mental states suggests amenable to the NSM approach. Given the local (lexical) semantics and/or speech situation, a specific array of sounds will be interpreted as signaling a specific attitude. Crucially, the acoustic properties of one mental predicate does not preclude that of another. This then paves the way for the possibility of a ‘grammar of politeness displays’, the specific nature of which will be determined by each community of speakers.

4.5. Summary and conclusion

In this paper, we have addressed several questions that would underlie any study of pragmatic phenomena. First, how do speakers signal pragmatic (non-lexical) meaning related to politeness? Our conclusion--corroborated by evidence from clinical studies--is that prosodic features play a key role in the display of affect, that is, whenever they stand out from otherwise normal patterns. Specifically, intensity is exploited in languages where stress has been incorporated (English and Russian), and duration is exploited in where vowel- and consonant length are phonemic (Japanese). This anticipates the question, How are listeners (including the speaker) able to interpret signals so quickly and efficiently? We have proposed that components of emotion are tracked separately in a way that listeners easily simulate. By recognizing the signature sounds of mental predicates, listeners learn to categorize a range of internal affective states.

5. REFERENCES

[1] Arbib, Michael A. (ed.). 2003. *The Handbook of Brain Theory and Neural Networks, 2nd edition.* MIT Press, Cambridge, MA.

[2] Armstrong, D., W. Stokoe & S. Wilcox. 1995. *Gesture and the Nature of Language*. Cambridge University Press.

[3] Blakemore, Diane. ‘Discourse markers’. In *The Handbook of Pragmatics*. Horn, Laurence, & Gregory Ward (eds.). John Benjamins, Amsterdam.

[4] Brown, P. & S. Levinson 1987. *Politeness: Some Universals in Language Use*. Cambridge University Press.

[5] Charron, S., & E. Koechlin (2011). 'Divided representation of concurrent goals in thehuman frontal lobes.' In *Science.,* Vol. 3288, No. 5976, pp. 360-3.

[6] Couper-Kuhlen, Elizabeth, & Cecelia E. Ford. 2004. *Sound Patterns in Interaction: Cross-linguistics Studies from Conversation.* Typological
Studies in Language 62. John Benjamins.

[7] Couper-Kuhlen, E., & M. Selting (1996). Prosody in Conversation. *Studies in Interactional Sociolinguistics* 12. Cambridge University Press.

[8] Fukuda, H. 2003. *Jazz Up Your Japanese with Onomatopoeia.* Kodansha, Tokyo.
[9] Goffman, E. 1967. On face work. In *Interactional Ritual.* Pantheon Books, NY.
[10] Grice, P. (1975). 'Logic and Conversation', in *The Logic of Grammar*, D. Davidson and G. Harman (eds), Encino, CA: Dickenson, 64-75.
[11] Grundy, P. 2000. *Doing Pragmatics (2nd edition).* Oxford University Press.
[12] Hamano, S. 1998. *The Sound-Symbolic System of Japanese.* Stanford and Kuroshio.
[13] Hirschberg, Julia (2002). *'The pragmatics of intonational meaning'.* Ms. AT&T Labs Research Unit.
[14] Hirschberg, J. 2004. 'Pragmatics and intonation'. In *The Handbook of Pragmatics*. Horn, Laurence, & Ward (eds.). Jon Benjamins, Amsterdam.
[15] Kamiya, T. 2002. *Handbook of Japanese Adjectives and Adverbs*. Kodansha, Tokyo.
[16] Kekahi, H., I. Tamori, & L. Shourup (eds.). Dictionary of Iconic Expressions in Japanese. *Trends in Linguistic Documentation 12*. Mouton de Gruyter, 1996.
[17] Levinson, S. C. 2000. *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. MIT Press, Cambridge, MA.
[18] Lichtenstein, J. 2011. 'Next Generation Scientists'. In *New York Times,* March 25, 2011.
[19] Locher, M. 2004. *Power and Politeness in Action: Disagreements in Oral Communication.* Berlin: Mouton de Gruyter.
[20] Locke, J., & B. Bogin. 2005. 'Language and life history: A new perspective on the development and evolution of human language.' *Behavioral and Brain Sciences;* Cambridge University Press.
[21] Maynard, S. 1997. *Japanese Communication: Language and Thought in Context.* University of Hawaii Press.
[22] Millington, S. 1993. *Nihongo Pera Pera.* Charles Tuttle Publishing Co., Rutland, VT.
[23] Noveck, I., & D. Sperber (eds.). 2004. Experimental Pragmatics. Palgrave (Macmillen).
[24] Ofuka, E., McKeown, J. D., Waterman, M.G., and Roach, P.J. 2000. *Prosodic cues for rated politeness in Japanese speech.* Speech Communication, 32, 199-217.
[25] Ono, T. 2006. 'An emotively motivated post-predicate constituent order in a 'strict predicate final' language: Emotion and grammar meet in Japanese everyday talk'. In *Emotive Communication in Japanese*; Suzuki, S. (ed.)
[26] Reed, B.S. 2006. *Prosodic Orientation in English Conversation.* Palgrave MacMillan, UK.
[27] Rizzolatti G., C. Sinigaglia, & F. Anderson (2007). *Mirrors in the Brain.* Oxford University Press.
[28] Schlüter, J. 2005. *Rhythmic Grammar: The Influence of Rhythm on Grammatical Variation and Change in English.* Mouton de Gruyter, Berlin.
[29] Sperber, D. & D. Wilson. 1995. *Relevance: Communication and Cognition.* Oxford: Basil Blackwell.
[30] Suzuki, S. 2006. *Emotive Communication in Japanese*. John Benjamins, Amsterdam.
[31] Tannen, D. 2005. *Conversational style: Talk among friends* (new edition). Oxford University Press.
[32] Watts, R. 2003. *Politeness.* Cambridge University Press.
[33] Watts, R., S. Ide, & K. Ehlich (eds.). 2005. *Politeness in Language, 2nd edition.* Mouton de Gruyter.
[34] Wichmann, A. 2000a. *Intonation in Text and Discourse: Beginnings, Middles and Ends*. Longman, New York.
[35 Wichmann, A. 2000b. 'The attitudinal effects of prosody, and how they relate to emotion'. Department of Cultural Studies, ms. University of Central Lancashire.
[36] Wierzbicka, A. 1999. Emotions Across Languages and Cultures. Cambridge University Press.
[37] Yule, G. 1999. *Pragmatics.* Oxford University Press.

THE PECULIARITIES OF "SCHWA" PERCEPTION IN AMERICAN ENGLISH

Svetlana Derkach

Amur State University, Russia
svetich_d2000@mail.ru

ABSTRACT

Being widespread and actively promoted, American English (AE) is in great demand in the Russian Far East; it is, therefore, important for Russian learners to improve their skills in AE oral speech perception.

The phonological concept "schwa" presents difficulties for auditory processing. If "schwa" is viewed as a mid-open mixed vowel that really occurs in the initial and final positions of words like "Alaska", then it is much less common in natural speech than dictionaries would make us believe. For instance, the acoustic picture of the first vowel in the words "society" (which is supposed to be "schwa") points to centralized *i*-like sound that occurs due to great influence of the neighboring forelingual consonants. Another example would be the second vowel in the word "university" (which is also supposed to be "schwa"). Due to distant regressive assimilation initiated by the primarily stressed vowel, the vowel in the second syllable is a shorter allophone of the stressed one. It is important to have good understanding of the phonetic conditions determining the quality of unstressed vowels.

The paper describes the methods and experimental data obtained in the study of the perception of "schwa", and compares its perceptual and acoustic features. The percepts were obtained from 23 synthetic vowels in different phonetic conditions. According to Clark and Yallop [5], many works on word recognition are usually considered to be research in psychology rather than phonetics. Less central to phonetics and phonology and of more significance in cognitive psychology is work on the cognitive processes involved in the recognition of words – how listeners process phonological structure sequentially and how they access lexical information from memory.

Keywords: perception, schwa sound, acoustic characteristics, articulation, synthetic stimuli.

1. INTRODUCTION

The knowledge about the material realization of the language system obtained by computer systems requires linguistic description; any phonological theory makes sense when it coincides with the experimental data [4]. Such approach enables us to have full understanding of a language ideal system of category structures having material representation.

At present studies of sound perception – including speech sounds – are given much attention in the physiology of sensor systems, in the recognition theory. The subsequent distinction of acoustic and perceptive characteristics of sounds is connected with great achievements in the acoustics and psychophysiology areas, however earlier specialists distinguished between two different approaches to sounds – acoustic, targeted to study the objective physical properties of sounds, and psychoacoustic related to the study of the way these physical properties are estimated by hearing.

The idea of perceptive aspect importance is inherent in the first works in phonology saying about the ability of sounds to distinguish morphemes and words: as phonological theories are related to the level of sounds, distinguishing significant objects should be based on certain acoustic characteristics resulting from articulation and are necessary for perception.

The ability to perceive speech having limited acoustic information lends every support to the fact that perception is not simply an analysis of acoustic features segments in chain but a more complex procedure including other levels of utterance – grammar, syntax etc.

To solve all these problems various studies are conducted. The results of the study of Native speakers perception depend greatly on the way these studies are conducted i.e. what tasks the researcher puts, what material is used, what instruction subjects, participating in the experiment get.

Synthetic stimuli are widely used by the scientists researching perception. Their peculiarity is that experimentator creating such stimuli defines not

only acoustic characteristics of synthetic sounds but he|she can change certain acoustic properties in the course of the experiment.
The current experiment aims at the clarification of the «schwa» status by studying its perceptional properties and comparing them to the acoustic ones. The hypothesis was that «schwa» is not a phoneme but allophone. The consideration of «schwa» perception by participants in different phonetic conditions was very important.

2. MATERIAL

23 synthetic stimuli were taken for the experiment. The stimuli were syllables containing different reduced vowels classified by the pronunciation dictionary as schwa sound. Different types of syllables were used as stimuli. The most stable period of the reduced vowel was multiplied in order to increase the vowel duration making it more suitable for perception.

3. SUBJECTS

A series of 23 stimuli were played and used as the basis of "schwa" perception to a group of 10 adult English-speaking listeners with normal hearing and 6 Russian-speaking phoneticians. They had no knowledge of the purpose of this experiment.

4. PROCEDURE

The instructions were different to different groups of participants. To English-speaking listeners the instruction was given in the following form "In this test block you will be listening to the stimuli, listen to the vowel and write down what you hear using any form you like (transcription or orthography)". Russian-speaking phoneticians who teach English were asked to transcribe vowels. English-speaking participants preferred orthography. The written form was orthographically to a presented item in varying degrees. It could, for example, be an English spelling of the item itself, or it could be a spelling of a word that contains the same vowel as the presented item. Sometimes orthography was combined with transcription that caused some difficulty related to the interpretation of the material obtained. To transcribe orthography we used rules of reading typical for American English.
The instruction and test materials were presented over headphones with the volume adjusted to a comfortable listening level. Subjects were tested in a sound-attenuated room and were allowed to ask questions of clarification about all the aspects of the procedure. The test was begun only after all subjects expressed confidence that they completely understood the task.
The experiment aims at answering the following questions

- Will these vowels be identified as «schwa»?
- Which phonemes will the listeners recognize instead?
- Will the perceptional and acoustic properties of these segments coincide?

In the graph «stimulus» (table 1, 2) the transription corresponds to the acoustic properties of the segments obtained after sonagram analysis. The vowels formants were compared to the mean values of vowels formants.

We were planning to get 368 responses in the course of the experiment (138 – Russian-speaking participants, 230 – American-speaking participants). But in fact we got 360 responses as participants had difficulty in identifying 8 stimuli (3 – Russian participants, 5 – American participants).

5. RESULTS

Tables 1, 2 show the responses of 15 subjects to the vowel identification test. The results show that participants were able to identify the sound «schwa» as belonging to different phonemes. The participants identified 10 stimuli out of 360 as «schwa» which makes 2,8%. Some might argue that modification of duration exerted a certain influence on a bad vowel perception. But maximum duration of natural «schwa» is less 50 msec. It is known that vowels with such duration can't be percieved adequately by a man [1], that's why the increase of duration could improve the perception.
Let's see cases of the most frequent coincidence of the participants' responses. We have 7 cases of such coincidence. The highest percent of coincidence has stimulus # 23 [ɛnd] – 93,8 %. This form of the conjunction "and" was realised between two pauses with hesitation duration that's why the duration of the stimulus wasn't changed.
To use words for the perceptual experiment is considered undesirable because participants guess easily what was pronounced due to the context [2]. In this case we have a pair of the quazihomonyms "end" – "and", which have different vowels /ɛ/ – /æ/. The majority of participants identified this stimulus as the word "end", though in fact this phonetic form is the phonetic manifestation of the word "and". The results of this stimulus perception fully coincide with its acoustic properties.

Next comes segment from the word famous. The participants in 87,5 % cases identified vowel /ɛ/. The participants were supposed to choose between two quazihomonyms “mess” – “mass”. Only in 2 cases the participants identified this sound comlex as “mass”. No one identified vowel as «schwa» though its formants correspond to «schwa». So we can say that acoustic and perceptual properties of this vowel don't coincide.

Retroflex unstressed vowels are also identified well (stimuli № 6, 9, 14, 17, 21) – 75 % of cases. These stimuli were segmented from the words bakery, burglar, from, perhaps, or. Usually participants associate these stimuli with /ɜ/ phoneme. Russian-speaking participants identify retroflex vowels much worse than American-speaking participants.

The analysis of the perceptual properties of the unstressed retroflex vowel lends every support to the idea that it is necessary to exclude it's use from the schwa distribution. Native speakers of American English identify unstressed synthetic retroflex vowel as combination of the phonemes /ɜ/ + /r/. The participants gave the spelling of the words that contained the same vowel as the presented item per/purr, word/world, curtain, stir, bird, her, church, were. The native speakers always identify this stimulus as the stressed one.

Stimulus # 13 the suffix from the word “freedom” has 68,8 % of coincidence. The majority of the participants identified this vowel as /ɛ/. In 93,8 % of cases this stimulus was identified as the front vowel.

Then comes stimulus #3 – unstressed syllable in the word “also”(62,5 %). It is close to /ɪ/ in its acoustic properties. The majority of the participants identified this segment as front vowel.

In stimuli № 10, 11, 22 which were segmented from the words Christmas, dominant, was correspondingly the vowels were identified as close in the degree of openness.

The vowels from stimuli № 1 и 2 which were segmented from the words acclaim and adult the Russian - speaking participants identify as close in the degree of openness. The English-speaking participants percieve the same segments as close in the degree of retraction|advancement.

In the stimulus № 20 segmented from the word society the vowel is close to /ɪ/ in its acoustic properties. Nevertheless it was identified as ɪ-like only in 6 cases out of 16 (3 cases for every group of the participants) which makes 37,5 %.

The vowel in the stimulus № 15 (from the word opinion) was identified as ɛ-like (4 responses out of 6). It has similar acoustic properties. As for the native speakers they identified this vowel as /ɪ/ -- 8 responses out of 10. The majority of the participants (75%) identified vowels close in the degree of advancement|retraction.

The vowel in the stimulus # 4 (from the word a person) is close to «schwa» in its acoustic properties. The participants in most cases identified mid central vowel /ʌ/ or low central vowel /a/. 62, 5 % of the participants identified this stimulus containing central vowel.

In the stimulus № 18 segmented from the word silence the vowel is close to /ʊ/ in its acoustic properties. However the majority of the native speakers identified it as /ʌ/ (6 out of 10 or 60 %). Only one Russian -speaking participant identified it as ʊ-like segment.

The last group of stimuli № 5, 7, 8, 16, 19 (from the words babble, beautiful, beautifully, people, social correspondingly). These are vowels with l in the post-position. All the vowels are ʊ-like segments. However in the stimulus # 5 the vowel was identified as /ʊ/ only in 5 cases. In 5 more cases the vowel was identified as /ɔ/. The majority of the participants identified back vowel (62,5%).

The stimulus № 7 has the same perceptual properties. Russian-speaking participants identified this vowel as /ɔ/. Three native speakers identified this segment as /ɔ/, another three as /ʊ/. In general 81,3% of the participants identified back vowel.

The vowel in the stimulus № 8 was identified as back by 75% of the participants.

6. CONCLUSION

As we can see from the tables the perceptual properties of the vowels classified by the pronunciation dictionary as «schwa» sound lend every support to the hypothesis that «schwa» is not the phoneme because it is not practically identified as schwa. The authors of this hypothesis competely share the point of view of Zinder [3] that to percieve the sound of speech means to identify it with the definite phonemes. As the participants in many cases don't identify what they hear with schwa there are serious doubts in the existence of the «schwa phoneme». Instead of «schwa» native speakers and Russian -speaking participants identify allophones of other phonemes.

7. REFERENCES

[1] Бондарко Л. В. 1981. *Фонетическое описание языка и фонологическое описание речи*. Л.

[2] Бондарко Л. В. 1998. *Фонетика современного русского языка.* Санкт-Петербург.
[3] Зиндер Л. Р. 1979. *Общая фонетика*. М.
Маслов Ю. С. 1997. *Введение в языкознание*. М.
[4] Clark, J., Yallop C. 1995. *An Introduction to Phonetics and Phonology*. Blackwell Publishers.

Table 1: The perception of the synthetic «schwa» by Russian-speaking participants (phoneticians).

№	stimulus	Participants' responses					
		R1	R2	R3	R4	R5	R6
1	ək	ɜ:	ɜ	9	æ	ɜ	ɜ:
2	əd	ɜ	ʌ	9	æ	ʌ	ə
3	sɪ	ɨ^{9}	ɪ	ɨ	ə	ɨ	e
4	əp	9	ɐ̟	ɑ̃	a	ʌ	ʊ
5	bʊl	–	ɞ	ʉ	ɔ	ʊ	ɔ:
6	kɜr	^{w}œ	ɜ̃	ʉ	^{w}œ	ɜr	ɜ:
7	fʊl	9	ɔ^{r}	o	ɔ	ʊ	ɒ
8	fʊlɪ	ɵ	œ	^{w}ɻ	^{w}ɔ^{ə}	ʊ	ʊ̯e
9	lɜr	ɵ^{l}ə	ɛə^{ᵛ}	er	ə^{ᵛ}	ɜr	ɛə^{r}
10	mɛ	–	ø	ɞ	ə	e	e
11	nən	ɛ	ʊ	ɛ̞	ə	ɛ	e
12	məs	ɛ	ɛ	ɛ̞	ɛ	ɛ	e
13	dəm	œ̃	ɪ	ø	ɛ	ɛ	e
14	ɜr	ʌ̃	ã^{ʊ}	ɑ	ɑ̃	ɑ̃	ʌ
15	ɛn	ĩ	ɪ	e	ɛ	e	e
16	pʊl	ʊ	ɞ	ɞ	ʊ	ʊ	ɔ:
17	pɜr	ɛ	ɜ̃	ɸ	ʀ	ɜɾ	ɜ:
18	ʊn	ɜ̃	ʊ	ɵ	ɑ̃	9	ʌ
19	ʃʊl	ə:	ʒ̃	œ	ə	ɜ	e
20	sɪ	9	u	ɨ	ə	ɨ	ɪ
21	ɜr	ɜr	ɜ̃	ə^{ᵛ}	ɛr	ɜr	e^{r}
22	wɜz	–	ɪ	ʏ	ɛ	ɜ	ɪ
23	ɛnd	ɛ̃	ɛ / e	e	ẽ	e:	e

Table 2: The perception of the synthetic «schwa» by American -speaking participants.

№	stimulus	Participants' responses				
		A1	A2	A3	A4	A5
1	ək	ʌk	ʌk	ak	ɛ:	ɛ:k
2	əd	ʌ	ʌt	at	ʌ	a
3	sɪ	sut	ts	sɛp	ɪ	st
4	əp	ʌp	ʌp	ɛʊk	ʌ	aʊk
5	bʊl	vɔl	brr	dʒɜr	–	o
6	kɜr	kɜr	dʒɜrt	kɜr	kɜr	kr
7	fʊl	florəl	ʌfɔrl	ɛsɛʊ	ɛʃɛʊ	askl
8	fʊlɪ	pʊi	fʊi	tɛʊi	tɔi	trɔi
9	lɜr	lɜr	blɛər	glɑr	blɜr	blɜr
10	mɛ	mm	nɛt	mɛ	n	me:s
11	nən	rʌf	n: ɛt	ɛə	nʌt	gn
12	məs	mɛs:	mæs	mɛs	mæs	mɛ:s:
13	dəm	lɛm	dɛm	ɪn	m	ɛ:m
14	ɜr	bɜrm	bɜrm	ɪ	bon	ɪ
15	ɛn	in	in	ɛn	n	ɪ:n
16	pʊl	fo	pbr:	ɜr	–	o
17	pɜr	pɜr	pɜr	ɑrt	pɜr	hlɜr
18	ʊn	mon	ʌnh	ad	bʌd	ʌt
19	ʃʊl	ʃʌl	ʃɜr	ʃɔl	pʃɛ	kʃal
20	sɪ	sʌt	sɪt	sʊt	sʌt	tsʌ:t
21	ɜr	ɜr	dɜr	ɜr	ɜr	ɜr
22	wɜz	blus	luis	lɪst	blɪs	lɪz
23	ɛnd	ɛnd	ɛnd	ɛn	ɛ:nd	ænd

№	The stimulus	Participants' responses				
		A6	A7	A8	A9	A10
1	ək	ək	ək	ʌk	ɜʊk	ʌk
2	əd	ʌp	ʌgh	ʌp	ɜr	ʌp
3	sɪ	sɪ	sɪ	sɪ	sɪ	sɪ
4	əp	ʊ	a	ak	ɔrk	ak
5	bʊl	ʌl	fʊl	ʊh	–	ʊl
6	kɜr	kɜr	kɜr	tʃɜr	kɜr	kɜr
7	fʊl	fʊl	fʊl	bɜʊl	slɔ	fʊl
8	fʊlɪ	toi	toi	toi	tʊwɪ	toi
9	lɜr	wɜr	wɜr	bɜr	lɜr	lɜr
10	mɛ	mʌ	nɛ	nʌ	lɪ	nɛ
11	nən	nɜr	mɛn	hɜr	nɜr	mɛn
12	məs	mɛs:	mɛs:	mɛs:	mɛs	mɛs
13	dəm	ðɛm	dɪm	tɪm	ɛlm	dɛm
14	ɜr	ɜr	bɔl	pæl	pɑm	ɜr
15	ɛn	ɪn	ɪn	ɪn	ɪn	ɪn
16	pʊl	fʊl	fɔl	fʊl	ɔr	fʊl
17	pɜr	bɜr	–	bɜr	–	pɜr
18	ʊn	ɜr	ʌn	ʌn	ad	ʌn
19	ʃʊl	ʃɛl	ʃɛl	ʃɛl	kʃɛl	ʃɛl
20	sɪ	sʊt	sɛ	sʊt	sɪt	sɪ
21	ɜr	ɜr	ɜr	ɜr	ɜr	ɜr
22	wɜz	bluz	ðɪs	ðɪs	lɪvɪst	luz
23	ɛnd	ɛnd	ɛnd	ɛnd	ɛnd	ɛnd

SOUTH AMERICAN ENGLISH. DO YOU NEED A PEN OR A PIN?

Ekaterina Gnatiuk

Amur State University, Russia
Kate_blaga@mail.ru

ABSTRACT

The article presents a complex, based on linguistic and experimental data analysis of phonetic features in Older and Newer white Southern American English (SAE).
Part one of the article gives a description of Older SAE phonetic features. The data on Older SAE were taken from academic publishing and internet resources on the subject.
Part two focuses on phonetic characteristics of Newer SAE, which to the author's knowledge haven't been fully investigated to date. An acoustic experiment was conducted in the Amur State University Phonetic Laboratory in order to illustrate tendencies in realization of some phonemes in Newer SAE. Separate analyses show that some trends in Newer SAE are already becoming apparent, though extent to which the results of that evolution yield something that is recognizably "Southern" remains to be seen.

Keywords: Southern American English, dialect, phonetic features, Southern Drawl, Southern Vowel Shift.

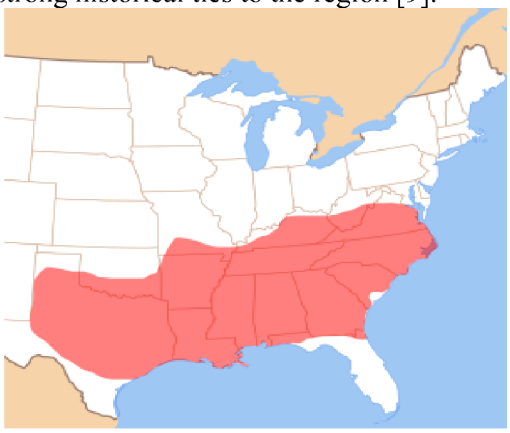
1. INTRODUCTION

Southern American English (SAE) is the most widely recognized regional dialect of American English, but as most of its speakers know, widespread recognition is a mixed blessing. SAE is also the regional dialect that is most negatively evaluated. In a study of folk beliefs about American dialects, Dennis Preston found that 90% of his respondents from Michigan and Indiana and 96% of those from South Carolina recognized SAE as a distinct variety of American English [8]. Both the Michigan and Indiana respondents, however, also evaluated SAE as the most "incorrect" variety of American English (New York City speech was the only serious competitor), and the South Carolina respondents were ambivalent about its correctness as well [1].
The widespread recognition and negative evaluation of SAE can have practical consequences for its users that in some cases include negative stereotyping and linguistic discrimination, just as with African American Vernacular English (AAVE), or Ebonics. While SAE almost never generates the extreme reactions and extensive prejudice that AAVE often does, its users can anticipate at least polite (and often not polite) condescension to their speech by non-Southerners. In spite of its low status outside of the South and of standardizing forces such as interregional migration and universal education that threatens many minority languages and dialects, SAE continues to persist [1].
Misunderstandings about what comprises SAE are almost as wide-spread as the recognition of its distinctiveness. These misunderstandings in large part have been fueled by media portrayals in movies such as *Gone With The Wind* and in television shows such as *The Dukes of Hazard* that presented exaggerated and inaccurate stereotypes of SAE. More recent portrayals in television shows are more accurate.

2. SPREAD OF SAE AND ITS FEATURES

Southern American English can be defined as a group of dialects of the English language spoken throughout the states which seceded to form the Confederate States of America during the American Civil War (War of Northern Aggression – to those Southerners who preserve historical heritage), and into those bordering them. This linguistic region includes Alabama, Georgia, Tennessee, Mississippi, North Carolina, Louisiana, and Arkansas, as well as most of Texas, Virginia, and Kentucky. It also takes in northern Florida, southern West Virginia, and southern and eastern Oklahoma. Parts of Maryland, and the far southeastern areas and south-central areas of Missouri, in the Ozarks region, are also included [5]. Outside of the South and Border States, are also small and/or isolated places in Kansas, New Mexico, Colorado, Arizona, Montana, and the San Joaquin Valley of California where the prevailing dialect is Southern in character or heavily Southern-influenced, due to historical settlement by Southerners. Also, the speech patterns of most of the southernmost counties of Ohio, Indiana, and Illinois – settled by Southerners and Southern Appalachians - have a predominately Southern

influence rather than Midwestern. The Southern dialects make up the largest accent group in the United States. Southern American English can be divided into different sub-dialects, with speech differing between regions. African American Vernacular English (AAVE) shares similarities with Southern dialect due to African Americans' strong historical ties to the region [9].



It's widely known that Southern dialects substantially originated from immigrants from the British Isles who moved to the South in the 17th and 18th centuries. The South was predominantly settled by immigrants from the West Country in the southwest of England, the dialects of which have similarities to the Southern US dialects. Settlement also included large numbers of Protestants from Ulster and Scotland [5].

Traditionally, SAE differed from other varieties of American English in some lexical, grammatical and phonetic features, but many of the lexical differences, which were rooted in an agrarian economy and a traditional society, have begun to disappear. For instance, most young Southerners are more likely to use *dragon fly* than either *snake doctor* or *mosquito hawk*. Many of the distinctive phonetic features of SAE still persist, however.

Few generalizations can be made about Southern pronunciation as a whole, as there is great variation between regions in the South and between older and younger people. For that reason scientists differentiate Older and Newer SAE.

The following features are characteristic of older SAE.

- Like Australian English and British English, the English of the coastal Deep South is historically non-rhotic: it drops the sound of final /r/ before a consonant or a word boundary, so that *guard* sounds similar to *God* and *sore* like *saw*. Intrusive /r/ where a /r/ sound is inserted at a word break between two vowel sounds ("*lawr and order*") is not a feature of coastal SAE, as it is in many other non-rhotic accents. Today only some areas like New Orleans, Mobile, Savannah, and Norfolk have non-rhotic speakers [5]. Non-rhoticity is rapidly disappearing from almost all Southern accents, to a greater degree than it has been lost in the other traditionally non-rhotic dialects of the East Coast such as New York and Boston. The remaining non-rhotic SAE speakers also use intrusive r, like New England and New York City.
- The distinction between the vowel sounds of words like *caught* and *cot* or *talk* and *tock* is mainly preserved. In much of the Deep South, the vowel found in words like *talk* and *caught* has developed into a diphthong, so that it sounds like the diphthong used in the word *loud* in the Northern United States.
- The distinction between /ɔr / and /or/, as in *horse* and *hoarse*, *for* and *four* etc., is preserved.
- The wine-whine merger has not occurred, and these two words are pronounced with /w/ and /hw/ respectively.
- Lack of yod-dropping, thus pairs like *do*/*due* and *loot*/*lute* are distinct. Historically, words like *due*, *lute*, and *new* contained /ju/ (as RP does), but Labov, Ash, and Boberg [5] report that the only Southern speakers today who make a distinction use a diphthong /ɪu/ in such words. They further report that speakers with the distinction are found primarily in North Carolina and northwest South Carolina, and in a corridor extending from Jackson to Tallahassee.
- The distinction between /ær/, /ɛr/, and /er/ in *marry*, *merry*, and *Mary* may be preserved by older speakers, but fewer young people make a distinction. The *r*-sound becomes almost a vowel, and may be elided after a long vowel, as it often is in AAVE [10].

The following phenomena are relatively wide spread in Newer SAE, though degree of features may differ between different regions and between rural and urban areas. The older the speaker the less likely he or she is to have these features:

1. Mergers of [ɛ] and [ɪ] before nasal consonants, so that *pen* and *pin* are pronounced the same, but the pin-pen merger is not found in New Orleans, Savannah, or Miami (which does not fall within the Southern dialect region). This sound change has spread beyond the South in recent decades and is now found in parts of the Midwest and West as well. Some scientists believe that the merger of [ɪ] and [ɛ] before nasals has its sources in British

English [3,4]; Lucia Morgan states that "the substitution of the vowel [ɪ] for[ɛ] as in *pin* for *pen*" is derived from the British Isles and is common in the speech of North Carolina [6]. Kurath and McDavid state the following: "In English folk speech again as th[ɛ]e of *pen* in the central counties...[and] the[ɪ] of pin in the eastern counties...The variants existed in Middle English, survive in English fold speech of today, and had a social rating in London English of the eighteenth century not unlike that in present-day American English." [4]. Krapp [3] also maintains that this phenomenon (which he finds in southwestern Virginia) is a relic of seventeenth-century colonial English.

2. Monophthongization of the diphthong /aɪ/ to [a] before obstruents. Some speakers exhibit this feature at the ends of words and before voiced consonants, so that *ride* is [rad] and *wide* is [wad]; others monophthongize /aɪ/. Throughout most of the region, this [a] tends to be more front so that word pairs like *rod* (SAE [rad], normally pronounced without any noticeable rounding) and *ride* (SAE [ræd]) are never confused [5]. In British English, J. C. Wells believes that it is widely agreed that the "mouth" vowel is a "touchstone for distinguishing between "true Cockney" and popular London" and other more standard accents. Cockney usage would include monophthongization of the word mouth. Example: *mouth* = [ma:f] (maaf) rather than [mae:f] mouth [7];

3. The Southern Drawl, or the diphthongization or triphthongization of the historically short front vowels as in the words *pat*, *pet*, and *pit*: these develop a glide up from their original starting position to [j], and then in some cases back down to schwa.

/æ/ → [æjə]
/ɛ/ → [ɛjə]
/ɪ/ → [ɪjə].

4. The Southern (Vowel) Shift**,** a chain shift of vowels which is described by William Labov [4] as: a result of the "drawl" described above, [ɪ] moves to become a high front vowel, and [ɛ] to become a mid front vowel. In a parallel shift, the nuclei of [i] and [e] relax and become less front. People can hear its most important feature simply by listening to Bill Clinton's pronunciation of the vowel in *way* or *stayed.*

5. The shift of the back vowels /u/ in *boon* and ***/o/*** in *code* considerably forward.

6. The shift of the open back unrounded vowel /ɑr/ *card* upward towards /ɔ/ *board*, which in turn moves up towards the old location of in *boon.*

7. The preserved distinction between /ɝ/ and /ʌr/ in *furry* and *hurry.*

7. In some regions of the south a merger of [ɔr] and [ɑr], that makes *cord* and *card*, *for* and *far*, *form* and *farm* etc. homonyms.

8. No distinction between /ɪr/ and /ir/ in *mirror* and *nearer*, *Sirius* and *serious* etc.

9. Substitution of /i/ by /ɛ/ at the end of a word, so that *furry* is pronounced as /fɝɪɛ/ ("furreh").

10. No distinction between /ʊr/ and /ɔr/ in *pour* and *poor*, *moor* and *more.*

11. Occasional pronunciation of the phoneme /l/ in the words *walk* and *talk*, causing the words *talk* and *walk* to be pronounced /wɑlk/ and /tɑlk/ by some southerners [11].

12. Merger of lax and tense vowels before /l/, making pairs like *feel*/*fill* and *fail*/*fell* homophones for speakers in some areas of the South. Some speakers may distinguish between the two sets of words by reversing the normal vowel sound, e.g., *feel* in SAE may sound like *fill*, and vice versa [5].

13. Omission of final t, d after another consonant: an(d), tol(d) [12].

14. Replacement of /z/ by [d] before /n/, for example [wʌdn̩t] *wasn't*, [bɪdnɪs] *business*, but *hasn't* is sometimes still pronounced [hæzənt] because there already exists a word *hadn't* pronounced [hædənt].

15. Accentuation of first syllable (rather than the second) in such words as: *guitar, insurance, July, police*, etc. [12].

It should be mentioned that much of the research on SAE has focused on its relationship to British regional dialects – on what many linguists see as its roots. This focus is primarily a result of the assumptions that American regional dialects are a reflex of settlement history and that they were formed during the colonial period. Recent research on SAE, though, suggests that both assumptions are inadequate. A case in point is the *pen/pin* merger. This merger occurred in the American South at least as early as the second quarter of the 19th century [3], but it occurred in only a relatively small segment of the population. During the last quarter of the 19th century, however, the *pen/pin* merger began to spread rapidly throughout the South until by World War II virtually all Southerners had the merger. This same 50 year period also saw the emergence and spread of the lost offglide in /ai/ and of the distinctive vowel pronunciation in words like *way*.

The diffusion of these features after 1875, after the initial settlement of the South, may seem odd, but demographic and socioeconomic developments of

this era suggest why these features may have begun to spread when they did. In *The Promise of the New South*, Edward L. Ayers points out that during the last quarter of the 19th century the emergence of stores, villages, and towns and a dramatic expansion of the rail system set in motion a process of urbanization that would ultimately reshape the region. In 1860 less than one in ten Southerners lived in urban areas (communities with populations of 2500 or more), and only 21 towns from Virginia westward through Texas had populations of 5000 or more. By 1900 the urban population of the South had doubled, and it doubled again by the onset of World War II. What seems to have happened linguistically is that migration to towns and cities created contact among dialects that were formerly local and insular, and as a result, features that were relatively restricted in occurrence began either to spread out or disappear. The parallel processes of diffusion and extinction eliminated many local vernaculars but at the same time gave rise to the larger regional dialect known today as SAE. Vestiges of some local vernaculars still persist among older residents of insular communities, as the work of Wolfram and his associates shows, but among younger Southerners they have all but disappeared.

Demographic developments since World War II raise some interesting questions about future prospects for SAE. The urbanization that began before World War II expanded dramatically during and after the war, but with some significant differences. Before World War II people in Southern towns and cities came from the surrounding countryside, and most industry involved low wage, manual labor operations such as cotton mills and petroleum processing plants. After the war, and especially after 1970, migration to Southern cities was as likely to come from the North as the South, and new industries often included such things as the corporate headquarters of J.C. Penny and the Dell computer production facilities. In addition, in Texas, Florida, Virginia, and in large cities throughout the South, migration from outside the United States is now occurring at an astonishing rate.

The linguistic impact that the new arrivals from outside the South will have is not yet clear, but some trends are already becoming apparent. In Texas and Oklahoma and in many metropolitan areas around the South, some national linguistic trends such as the merger of the vowels in *caught* and *cot* (both sound like the latter) are emerging, and in several of the larger metropolitan areas (e.g., Dallas-Fort Worth and Memphis) some traditional Southern vowel features such as the distinctive pronunciation of the vowel in words like *way* are beginning to wane. While the long-term linguistic consequences of the new developments are impossible to predict, it is apparent that SAE is continuing to evolve – just as it has over the last century and a half. The extent to which the results of that evolution yield something that is recognizably "Southern" remains to be seen.

3. REFERENCES

[1] Bailey G., Tillery J. 2006. American varieties. Sounds of the South. *Language Magazine*. p.1.

[2] Brown, A. 1991. *Teaching English pronunciation: a book of readings*. London: Routledge.

[3] Krapp, G.P. 1925. *The English Language in America*.

[4] Kurath, H., McDavid, R. 1961. *The Pronunciation of English in the Atlantic States*. University of Michigan Press.

[5] Labov W., Ash S., Boberg C. 2006. *The Atlas of North American English: Phonetics, Phonology and Sound Change*. Berlin.

[6] Morgan, Lucia C. 1969. North Carolina Accents. *Southern Speech Journal* 34, 223-229.

[7] Wells, J.C. 1982. *Accents of English 2 The British Isles*. Cambridge University Press.

[8] Preston D. Some plain facts about Americans and their language. 2000. *American Speech* 75.

[9] American Phonology.*www.bookrags.com/wiki/Southern*

[10] *www.ic.arizona.edu.*

[11]Linguistics.*http://www.utexas.edu/courses/linguistics/resources/socioling/talkmap/talk-nc.html.*

[12] Linguistics 201: The Dialects of American English.*www.edu/*

PHONETIC MANIFESTATION OF RHEMES HIERARCHY IN SPONTANEOUS SPEECH

Svetlana Guseva, Maria Etenko, Tatjana Nezhivykh

Amur State University, Russia
s_guseva@mail.ru; taelka@rambler.ru

ABSTRACT

The paper presents the results of the research of segmental and prosodic peculiarities of repeated words caused by their position in the utterance informative structure. Discourse informative structure is determined on the basis of perception experiment results. Differences in phonetic manifestation of nuclear and peripheral elements in spontaneous monologue are studied. The results confirm the hypothesis about the potential link of a phoneme with the utterance informative structure and enable to consider the location of a phoneme in prosodically prominent rhematic parts as a factor providing stability of phoneme characteristics.

Keywords: allophonic variation, informative structure, Rhemes hierarchy, prominence.

1. INTRODUCTION

1.1. The Phoneme and Sense

Different types of discourse have been the focus of linguists attention for a number of years. The aim of the current research is to study the potential link between the Phoneme and Sense through its participation in Rhemes hierarchy phonetic expression in different types of verbal spontaneous discourse.
The hypothesis of the research is that the Phoneme, having no meaning of its own and thus not being a typical symbol-like unit of Language, is the smallest constituent of verbal spontaneous discourse. That is why phonemes phonetic correlates largely depend on prosodic characteristics of the utterance which in their turn are determined by speakers' communicative intentions. Thus phonemes as specific autonomous language units participate in forming these intentions. Because of that phonemes realization pattern can never be fully understood without addressing semantic and prosodic modality of the utterance that is without taking into account the potential link between the Phoneme and Sense.

1.2. Information distribution in the Utterance

It is quite an old observation that information is distributed irregularly within the scope of verbal spontaneous utterance. It forms a certain hierarchy that is described in varying terminology by different scholars: old vs new words, or new vs given information, or unusual/low-probability vs high-probability words, or informative vs uninformative words [Fowler, Housum, 1987: 489-490; Bolinger, 1963, 1981; Chafe, 2009].
However important information can also have certain gradations – Rhemes hierarchy. In present study the Rhemes hierarchy is defined as information gradation within the utterance, which is conditioned by speaker's communicative strategy and realized in prosodic Rheme characteristics.
The report presents the research results of Rhemes hierarchy in different types of verbal spontaneous discourse of German native speakers. The report also presents phonetic analysis of rhemes with different informative potential.

1.3. Rheme and theme

The definition of the term rheme, which is a key term in our study is traced back to the famous Prague Linguistic School representative Vilem Mathesius [Матезиус, 1967] and his so-called functional sentence perspective according to which the sentence is split into Rheme and Theme. The appearance of the terms was caused by obvious differences in information distribution during linear development of language units in speech process. In other words linearity means that some utterance parts contain more significant information and the other less significant one.
There is a considerable number of Rheme definitions in modern linguistics (new information/words, unusual/low probability words, key words, most important information/words, communicative center). All of them have a common idea that rheme segments are somehow or other prosodically prominent, non redundant. The major prosodic cue of Rheme parts has long been considered overall lengthening of words [Bolinger, 1963: 7, 1981; Fowler, Housum, 1987: 492]. Other

cues suggested: vowel peak amplitude increase and mean F0 increase [Fowler, Housum, 1987: 492].
The Theme as old/given/high-probability words/information is the communicative periphery of an utterance. It is not prosodically prominent and is characterized by certain features of redundancy: pitch and vowel peak amplitude lowering and distressing vs stressing in the Rheme part [Chafe, 2009: 139; Fowler, Housum, 1987: 492].
Russian scientist T.M. Nikolayeva [Николаева, 1981] performs the analysis of functional sentence perspective and prosodic prominence correlation. In her study she notes that there is a main-stressed element in every sentence which is normally at the end of the utterance. This element is the Rheme. However Nikolayeva underlines that the Rheme can include several words whereas sentence stress is focused on the accented syllable of the final word only. As a result there is a discrepancy between prosodic prominence and verbal content of the Rheme: The Rheme part of an utterance can be rather wide, but sentence stress effects spread is limited.
Thus functional sentence perspective determining might turn out problematic for some reasons. The first one is the mentioned above discrepancy between prosodic prominence and verbal Rheme content. The second is the absence of uniform structural-semantic organization of different forms and styles of speech. Spontaneous discourse can be a perfect example. Spontaneous discourse is less structured than a prepared speech, so quite often it is impossible to perform syntagma division using existing syntagm boundary signals such as pause, slight breaks in timing, a shift to a new pitch baseline and the like (see more about boundary signals in W. Chafe [Chafe, 2009: 137]).
Illegibility of Rheme boundaries and the absence of universal acoustic criteria for its segmentation in spontaneous discourse make it difficult to single out the Rheme in the utterance. In order to make it possible we turned native speakers' perception. We believe that such approach correlates with anthropocentric ideas of modern linguistics, according to which an individual is treated as a representative of language community whose perceptual base is the typical reflection of the whole community and who is the source and consumer of language information.
At the same time we took into account an obvious assumption that informative important or Rheme segments of a discourse are marked by certain phonetic parameters, which are the signals for Rheme identification. Obvious connection between Rheme and pitch accent leads to iconic encoding of semantically important, actual information by prosodic marking; that makes it possible for representatives of the same language society to determine communicatively important or Rheme parts of an utterance.
In the psychophysiological experiment devoted to the perception of segments with different informative prominence listeners compare the information that they hear with models of communicative structure of utterance they keep in their memory. The knowledge of communicative strategies and tactics is a part of all cognitive areas types of language personality. They are characterized by dual nature being collective and individual at the same time. The communication is impossible without "shared knowledge": shared language system and rules of speech interaction.

2. EXPERIMENT

2.1. Material and procedure

The material for the research is different discourse types: spontaneous, prepared political and prepared religious presented in the form of monologues of German native speakers (male and female).
Political discourse was represented by two speeches of Angela Merkel. Religious discourse was represented by a German priest's sermon.
Experimental procedure included several stages.

2.1.1. The perceptual study

The subjects were German native speakers (21 for spontaneous, 15 for political, and 19 for religious discourse types). The subjects were to listen to the recordings and underline informatively prominent parts in each phrase. The results were quantitatively processed. Informatively significant words and word combinations, underlined by the majority of auditors (60 – 100 %), made up the main body of the experimental material. These words and word combinations were called D*iscourse Informative Core (DIC).*
According to the frequency of the words in DIC they were given indexes: 1,0 to words written out by 100% of listeners, 0,6 to the ones written out by 60% of listeners etc.
According to the research the words indexed from 1,0 to 0,8 were seen as the nuclear parts of DIC, the words indexed from 0,6 to 0,7 – as the peripheral ones. The analysis of DIC words distribution enables to define its informative saturation.

2.1.2. Transcription

The spontaneous discourse transcription is an uneasy task. A lot of decisions should be made in order to turn the speech flow into the graphic discrete record. The way the transcription is made depends on the goals of the research. In this study, we use intonation transcription based on the perceptual study of pitch, pauses, stress and tempo. Here we present some characteristics of pitch.
In the present research we also made a segmental transcription of DIC words. Identical words having different informative prominence were selected for that purpose. Such choice was determined by the necessity to exclude coarticulation effects as an independent variable and concentrate only on allophonic variation as a result of location in the Rheme or the Theme parts of the utterance.

2.1.3. Acoustic study

The next stage of the research was the acoustic analysis of the identical words having different informative prominence and the identical words, having DIC index contrast. The analysis of colloquial spontaneous discourse was performed with the help of EDS program of acoustic analysis. PRAAT was used for analyzing political and religious discourse. Similar words, taken from the different parts of discourse informative structure – the ones belonging to DIC and the ones not belonging to DIC. Using similar words enabled to concentrate on acoustic differences determined mostly by discourse informative structure and not to mistake them for pure coarticulatory effects.

3. RESULTS AND DISCUSSION

The results of perceptual and acoustic study of various discourse types provide us with the material for comparing the characteristics of spontaneous speech, political and religious discourses.

3.1. Spontaneous Discourse

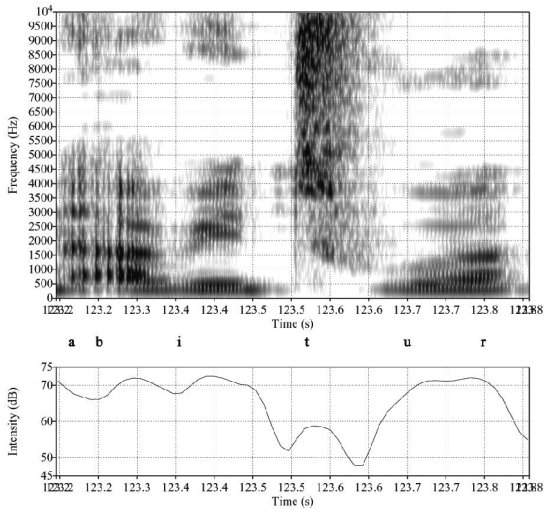
3.1.1. Intonation peculiarities

Perceptual study of spontaneous monologues covered 3395 words. 708 of the (21%) were chosen by the majority of listeners as informatively prominent. These words were viewed as the Reme parts or DIC. Among them 367 words were considered as nuclear parts the Rheme or DIC nucleus (words indexed from 1 to 0,8) and 341 were treated as the peripheral parts of the Rheme or DIC periphery (words indexed from 0,6 to 0,7).
The acoustic analysis of DIC words intonation pattern demonstrated that the majority of them (about 60%) were characterized by rising or level pitch pattern. The other 40% of the words were located in the terminal parts of syntagma or phrase taking falling pitch pattern.
Traditionally in prepared speech (e.g. prepared reading) the falling pitch pattern is considered most typical for the nuclear words whereas the peripheral words are pronounced with rising or level pitch pattern three times more often than the nuclear ones. The result of the current study enables to speak about considerable discrepancy of pitch patterns used for important words in spontaneous and prepared types of discourse.

3.1.2. Phonemes realization peculiarities

In general the results of acoustic study enable to speak about some weakening and shortening of consonants and differences in coarticulatory effect namely increasing coarticulation from DIC nucleus to DIC periphery and finally to words not belonging to DIC. Typical examples are given in fig. 1-2.
Figure 1 shows two spectrograms of the word "Abitur" with contrasting DIC indexes one taken from DIC nucleus, another – from DIC periphery. Both realizations of the word are from spontaneous monologue devoted to the topic "Ausbildungssystem" (Fig. 1 a, b).

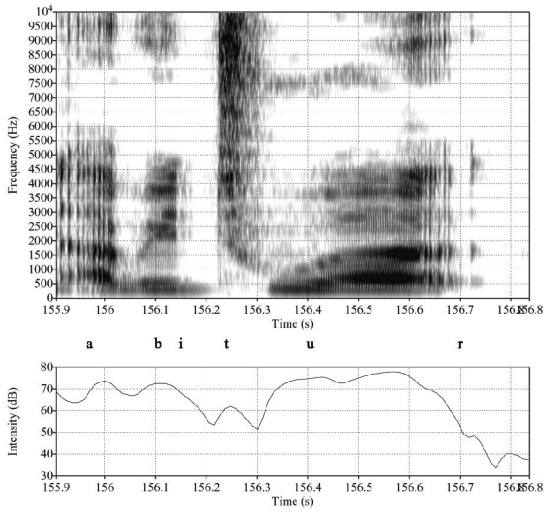
Figure 1: Sonogram of the word "Abitur" with contrasting DIC indexes.
a) Abitur – 0,9



As we can see, the differences of of the stops /b/ and /d/ manifestation are not significant as far as noise configuration is concerned. However, word indexed 0,6 has some weakening of its occlusive phase of /t/. The study of consonant /t/ in DIC nuclear words shows that the duration of its occlusive and burst phase is a little bit longer, than in periph-

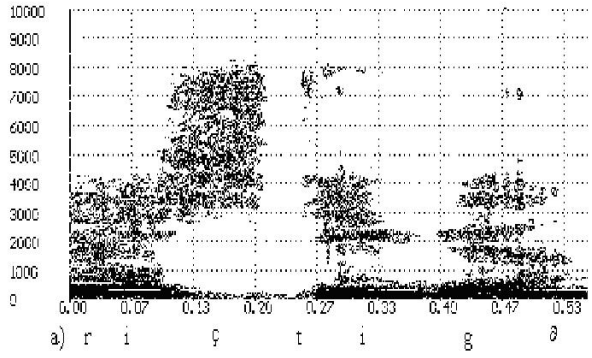
ery words: 83 ms vs 77 ms for the occlusive phase, and 122 ms vs 109 ms for the burst phase correspondingly.

b) Abitur – 0,6

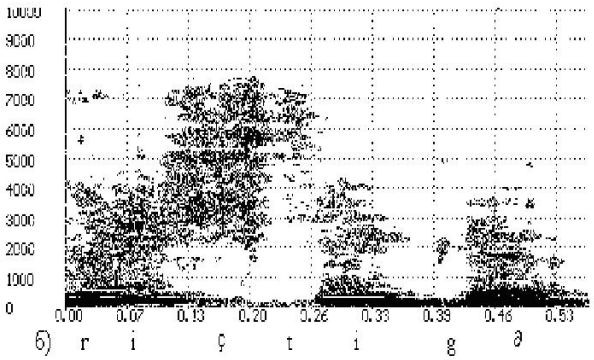


The similar comparison is presented here for word "Richtige" with indexes 1,0 and 0,6 (Fig. 2 a, b).

Figure 2: Sonogram of the word "Richtige" with contrasting DIC indexes:
a) Richtige – 1,0



б) Richtige – 0,6.



Spectrograms comparing shows stronger coarticulation of vowels with the neighboring consonants in word indexed 0,6, which in turn also results in the occlusive phase weakening.

3.2. Political discourse

The material for perceptual study included 3646 lexical units. 716 words (about 21%) words composed DIC (were marked by the listeners as informatively prominent).

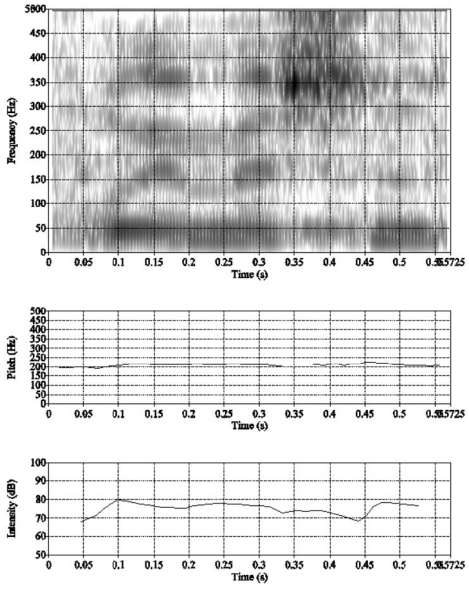
3.2.1. Intonation peculiarities

The acoustic analysis showed, that in political speeches informative prominent segments of utterance have falling or rising pitch patterns (more than 60% in two speeches). Unlike in spontaneous discourse no level tone was used for important words. Comparing the frequency of DIC words marked by falling pitch pattern with the ones marked by rising pitch pattern we can say that the former outnumber the latter.

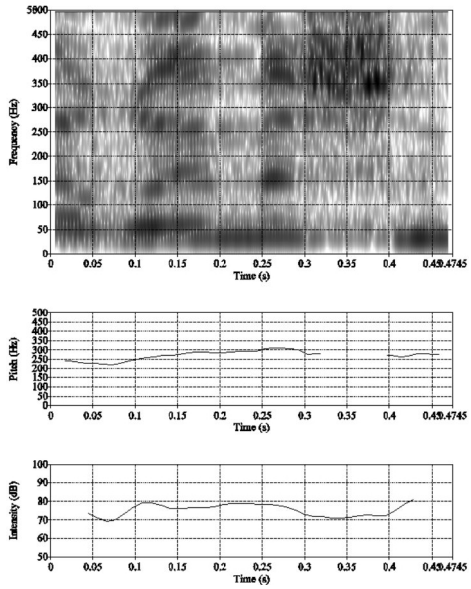
3.2.2. Phonemes realization peculiarities

Two words with DIC contrasting indexes were taken from Angela Merkel speeches to examplify the results of acoustic study of the political discourse: "Römischen" and "Europa" (Fig. 3, 4).

Figure 3: Sonogram of the word "Römischen" with contrasting DIC indexes.
a) Römischen01 – 0,9

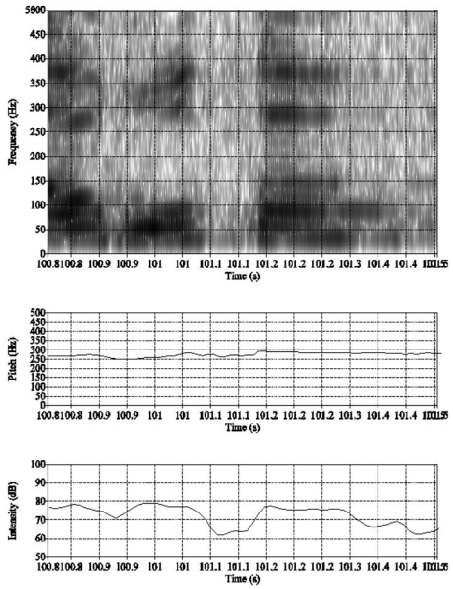


b) Römischen03 – 0,6



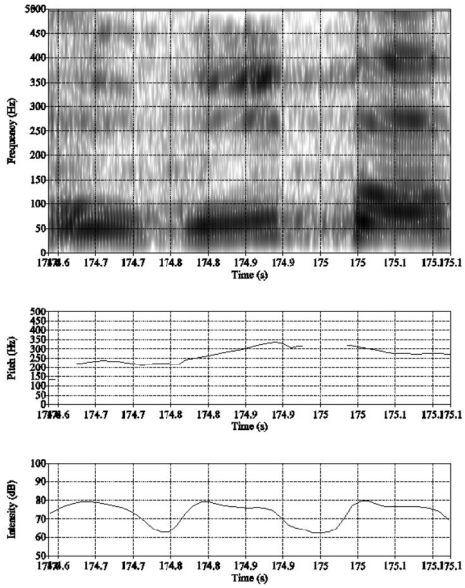
As in the examples from spontaneous discourse overall duration of the word Römischen indexed 0,9 is longer than the one indexed 0,7. Neither pitch patterns not intensity seem to differ much. However, the word with the lower index seems to have bigger gestural overlap of the segments (the stressed vowel and the following nasal — fig. 3b). The overall duration and intensity patterns of two realizations of the word "Europa" does not differ much.

Figure 4: Sonogram of the word "Europa" with contrasting DIC indexes.
a) Europa01 – 0,7



However the one from DIC nucleus with the higher index is characterized by rising pitch (40 Hz F0 increase during the stressed vowel) and better formants expression (fig. 4b) while the one with the lower index from DIC periphery – by level tone and less localized energy increases in the spectrum which is especially clearly seen for the second stressed vowel (fig. 4a).

b) Europa03 – 0,9



3.3. Religious discourse

The material was a 6 minutes sermon of German priest. The listeners were 19 German Christians. The results of experiment show that 7% of the words were marked as informatively prominent. It is considerably less compared to 21% for spontaneous and political discourse samples. 36% out of those 7% were regarded by the listeners as nuclear parts of discourse (the words indexed 1,0-0,8) and 64% as peripheral ones.

3.3.1. Intonation peculiarities

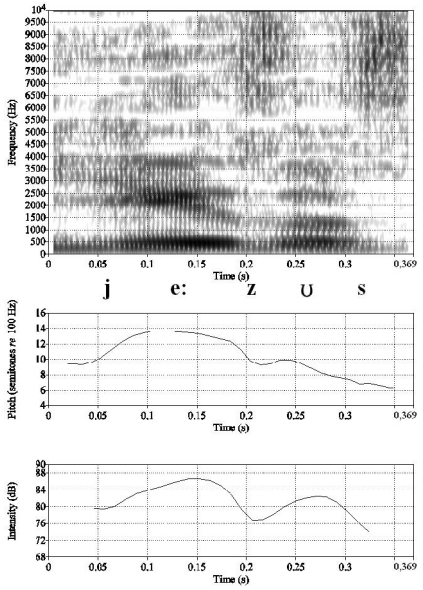
Acoustic analysis shows that the 61% of DIC words have rising tone, and other 39% have falling tone. Falling tone is typical both for DIC nucleus and DIC periphery. DIC periphery words are marked by level or rising tone two times more often than DIC nucleus ones.

3.3.2. Phonemes realization peculiarities

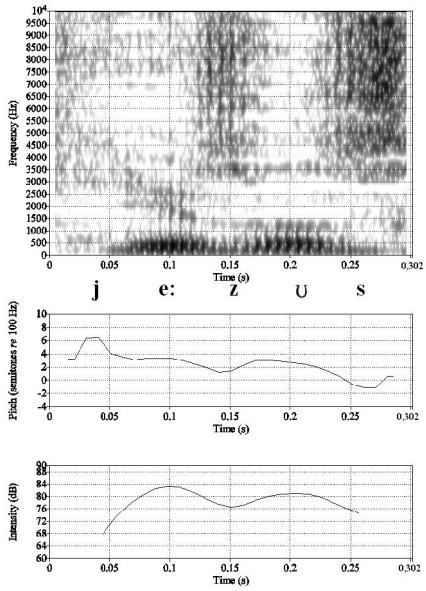
In the present paper we will look at two most frequent words in our samples of religious discourse – "Jesus" and "Christ" (Fig. 5, 6).
The analysis of spectrograms shows, that the intensity of /e:/ formants in word "Jesus04" (1,0) is higher than in word "Jesus09" (<0,6). The overall

duration of the word “Jesus04” is 67 ms longer that the one of the word “Jesus09”. The word “Jesus09” not being a DIC part is characterized by level tone, whereas in the word “Jesus04” – by the falling one (65,1 Hz declination) (Fig. 5 a, b).

Figure 5: Sonogram of the word “Jesus” with contrasting DIC indexes
a) Jesus04 – 1,0

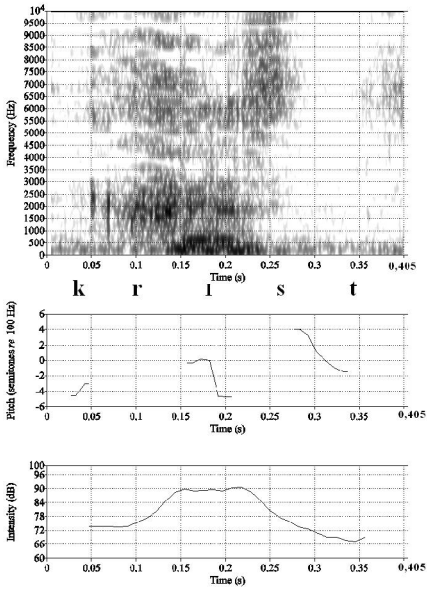


b) Jesus09 – <0,6

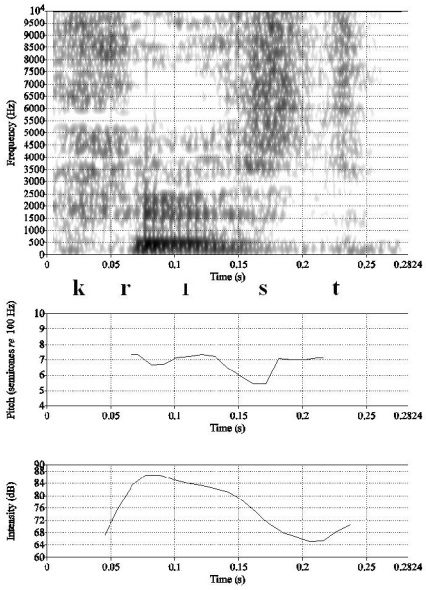


Level and rising tone for DIC nucleus, however, are not excluded. Thus, both realizations of the word “Christ” are characterized by level tone. The word “Christ01” from DIC nucleus is 123 ms longer than the one from DIC periphery (Fig. 6 a, b).

Figure 6: Sonogram of the word “Christ” with contrasting DIC indexes:
a) Christ01 – 1,0



b) Christ04 – <0,6



4. CONCLUSION

The results of perceptual and acoustic study of phonetic manifestation of informational structure of various types of discourse – spontaneous colloquial, prepared political and prepared religious, – produced by German speakers, partly replicate the results for the English language as far as lengthening as an indicator of the Rheme (new

words). Our finding in this aspect is that duration is also consistently used for information gradation within the Rheme itself (DIC nucleus and DIC periphery).

Another finding is obvious preference for rising and level pitch patterns for DIC in spontaneous discourse and for falling pitch pattern for DIC in prepared discourse. As far as differentiating between DIC nucleus and periphery there is a greater number of falling pitch pattern occurrences in the former compared to the latter.

Comparing the spectral characteristics and knowing about the articulatory gestures and their acoustic output we can say that DIC nucleus is marked by more careful realization of the articulatory phases of speech sounds. It is obvious that informatively prominent parts are the "area" of more careful pronunciation. Such "areas" guarantee stability of the phonological system keeping the necessary for the listener amount of phonetic characteristics from being vague.

5. REFERENCES

[1] Bolinger, D. 1963. Length, vowel, juncture. Linguistics. An international review. Hague: mouton and Co., 5-29.

[2] Bolinger, D. 1981. *Two kinds of vowels, two kinds of rhythm*. Bloomington IN: Indiana University Linguistics club.

[3] Chafe, W. 2009. Consciousness and language. In *Cognition and pragmatics.* A handbook of Pragmatics highlights (HoPH), vol. 3. Amsterdam/Philadelphia: John Benjamins Publishing Company.

[4] Fowler C. A., Housum J. 1987. Talkers' signaling of "New" and "Old" words in speech and listeners perception and use of the distinction. *J. of Memory and Language*, 26, 489-504.

[5] Матезиус, В. 1967. *О так называемом актуальном членении предложения*. Москва.

[6] Николаева, Т. М., 1982. *Семантика акцентного выделения.* Москва.

STRESSED VOWELS IN BRITISH POLITICAL DISCOURSE

Yulia Ivanashko

Amur State University, Russia
polia-80@mail.ru

ABSTRACT

The article presents the results of instrumental analysis of stressed vowels in British political discourse. The article gives the review of acoustic features of vowels under nuclear (syntagmatic) stress and nonnuclear stress, located in different parts of syntagma (initial, middle, final). The results show the dependence of vowels' quantity (length) and quality on stress type and their position in syntagma.

Keywords: political discourse, stress, length, melody, intensity, acoustics.

1. INTRODUCTION

Probably the two most problematic and widely discussed areas of phonology of English language are those which relate to word stress and vowel length [1].

Prosodic stress is an integral component of spoken language, particularly for languages, such as English, that so heavily depend on it for lexical, syntactic and semantic disambiguation. Prosody also provides important information about the focus of the speaker's attention, highlighting for the listener what is "new" and "important" information, thus serving to facilitate processing via parsing the utterance into delimited "chunks" for reliable understanding. In other words, certain vocalic segments (whether they are somehow connected (or not) with the pragmatic or semantic aims of the speaker) are far more likely to be accented (or not) than others. Such stress-related information is derived from a complex constellation of acoustic cues associated with the duration, amplitude and fundamental frequency of syllabic sequence within an utterance. Stress and duration have the closest interdependence than the rest of the mentioned factors, as duration is the primary acoustic correlate of stress accent [2].

The current study focused on the relation between different types of stress accent and the degree of vowels (monophthongs) duration.

In this study it is demonstrated that nuclear (syntagmatic) stress is characterized with longest vowel duration within the syntagma.

2. MATERIAL AND METHOD

2.1. Corpus

Our material consists of the recordings obtained directly from the official website of British government (www.pm.gov.uk).

The speakers are three British former and current Prime Ministers. These are the abbreviations to be used throughout: D1 – David Cameron, 44 years old; D2 – Tony Blair, 64 years old; D3 – Gordon Brown, 66 years old. The recordings took place in 2010, 2006 and 2007 correspondingly. Three quasi-spontaneous speeches pertain to political topics and present the orthoepic norms of British English.

In this paper, we only report male data on the grounds that gender is known to be an important factor of acoustic variations (especially vocalic system variation); thus including data from two genders would have necessitated separate tables and comments, which would have considerably increased the length of this paper.

The total length of the speeches is 1 hour 30 minutes.

2.2. Method

At the beginning of the experiment the present speech materials were segmented into the syntagmas.

Syntagmatic segmentation is still somewhat problematic as there aren't any formal criteria of determining the boarders and signs of syntagmas [3].

In the present paper we will name the features of a syntagma to be used in further syntagmatic segmentation.

L.V. Scherba was the first to describe a syntagma as "a kind of a phonetic unity, expressing a semantic entity in the speech-thinking process" or "a simple syntactic unit". According to the author, the sentence may be divided into the segments, characterized by the stress intensification of the final

word and expressing one concept. A syntagma may consist of one or several words [4].

O. S. Akhmanova determines syntagma as a two-term structure the parts of which are related as a determinate and a determinative [5].

L. V. Bondarko presents syntagma as a minimal linear unit of a phonetic structure associated with the meaning of the utterance and representing only its part. The author underlines that the syntagmatic segmentation doesn't have a close relation to the syntactic one as the same syntactic pattern can be implemented in speech in different ways. Thus, the phonetic integrity of a syntagma is determined by its intonation pattern including melody, intensity, tempo, the presence or absence of pauses [6].

As for the melody pattern, it isn't the same in a syntagma. Maximum melody pitch can be the symbol of a nuclear (syntagmatic) stress [7, 8].

The fall of intensity from the beginning of a syntagma to its end is a general phonological law [6].

Pauses usually determine the syntagmas boarders. Experiments show that the pauses appear between the phrases including 7-9 words. It is explained by some physiological factors such as breathing and short-term memory peculiarities. But in some cases the segmentation can be done in non-pause environment or, in contrary, pauses can appear inside a syntagma to indicate a focal unit [8].

Our investigation shows that the syntagmas in present materials include from 1 to 7 words. The speech of D3 is characterized by shorter syntagmas (1-5 words).

It is known that the phonetic integrity of a syntagma is determined by the fact that all the words in it excluding the final one have word stress. As for the final word, it has stronger stress than the others in a phrase. Such kind of stress is called syntagmatic or nuclear [6]. But in real communication this is an informative structure of an utterance which determines the position of syntagmatic stress. Due to its influence syntagmatic stress can be the feature of any word in a syntagma. According to B. K. Murzalina, the placement of syntagmatic stress depends on the semantic relations of an utterance [9].

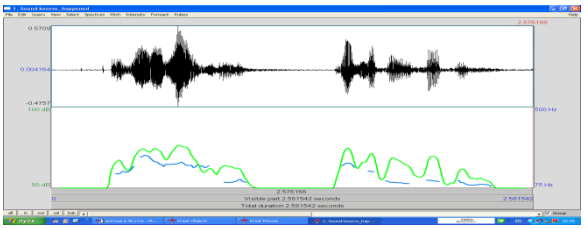
Foreign authors label the semantic stress as "accent" and "focus". Russian scientists distinguish "emphatic stress", "logical stress" and "accent".

In the current study emphatic and syntagmatic types of stress are seen as equal and termed syntagmatic stress.

All of the present material was segmented into the syntagmas taking all the above mentioned intonation components into account.

In Fugure 1 the intonation pattern of the phrase *A year later Kosovo happened* is shown. According to the pitch movement and intensity level it is clearly seen that the phrase consists of 2 syntagmas. Nuclear accent peak of the first syntagma *A year later* is [jɪə] syllable, the second (*Kosovo happened*) – [kɒ] syllable as they are the highest stressed ones in the phrase. But maximum pitch is not always the indicator of the prosodic stress. The pitch raise can be the result of some segmental characteristics. Thus, voiceless stop consonants can cause the pitch raise and voiced ones – pitch fall.

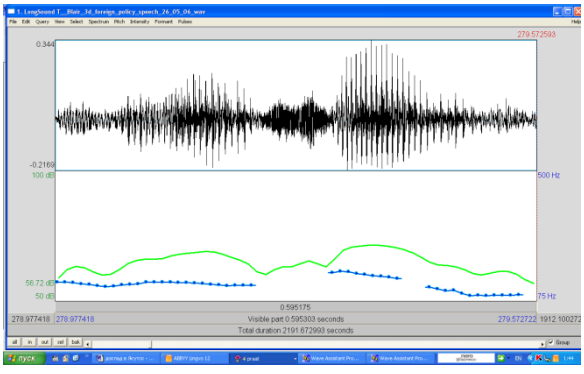
Figure 1: Intonation pattern of the phrase *A year later Kosovo happened.*



The second stage of the study was the acoustic analysis of the speeches (in Praat and Wave Assistant computer programs) aiming at labeling all the vowels under word and syntagmatic stress.

Figure 2 shows the waveform of the word "the USA". We can see that the sound [ei] has the highest intensity and pitch rate than the others, so it has the main stress.

Figure 2: The waveform of the word "the USA".



The syntagmatic stress was determined according to the study of the intonation pattern of each syntagma.

At the next stage of the investigation we examined vowels duration under syntagmatic stress and word stress.

While examining the vowels length some factors were considered:

- acoustic parameters of a sound depend on its

position in the word or syntagma [10];
- vowels have greater length in final position of a syntagma [11].

In Fugures 3 and 4 there are the examples of the words segmented out of different parts of syntagmas.

Figure 3: The sonogram of the word "argued" (the beginning of the syntagma, word stress).
[a:] duration is 72 ms.

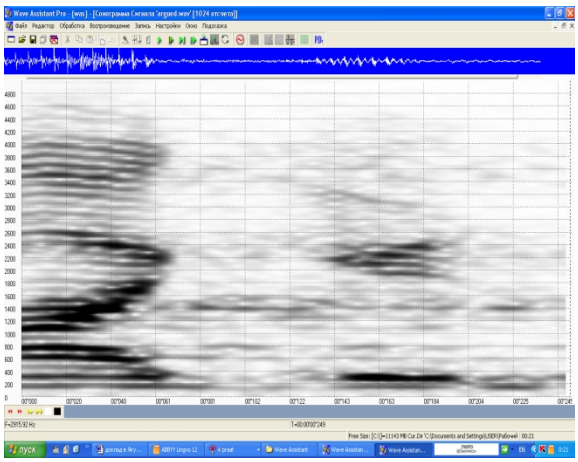
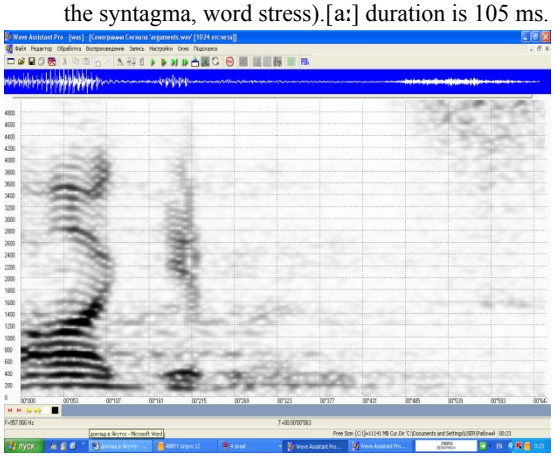


Figure 4: The sonogram of the word "arguments" (the end of the syntagma, word stress).[a:] duration is 105 ms.



Another intonation component influencing the vowels duration is tempo (speed of utterance). Low speed leads to the most careful pronunciation of sounds while high speed causes somewhat of their reduction [10].

So, one of the aims of our experiment is to find out tempo averages of the speeches in order to be sure in an adequate comparison of vowels duration in different speakers.

In the present study speed is measured as the number of sounds a second, as sounds are the most appropriate units than syllables (as a syllable can be a unit consisting of different number of sounds which can lead to the measurement errors). Thus, the speed of utterance in our study is the total sound output divided into the total time of the speech segment:

(1) $T = L\tau\Delta t$

T – tempo

L – total number of sounds

τ- time scale

Δt- total time of the speech segment

Table 1 presents the data of speakers' tempo averages.

Table 1: Tempo averages .

Speaker	Average tempo (sounds per sec.)
D1	13
D2	13
D3	11

The data shows a slight difference in speakers' tempo. It lets us to compare the speeches in an adequate way.

For studying the variety of vowels duration it's also important to consider some other factors:

(1) inherent duration of vowels (low vowels are longer than high ones) [12];
(2) neighboring sounds and the position in a word (the same vowel is longer if it precedes the voiced consonant and shorter if it goes before the voiceless one) [11];
(3) the number of unstressed syllables following the stressed one (more the syllables, the shorter the sounds).

Thus, in the present investigation the following vowels were studied:

(1) in the open syllable before the pause (i.e. see, far, three, car, be);
(2) preceding a voiced consonant (i.e. speed, carbon, these);
(3) preceding a nasal sound (i.e. term, long, none);
(4) preceding a fricative (i.e. speeches, large).

We consider these positions to be relatively equivalent as they have almost the same effect on vowels duration.

The experiment doesn't include some English monophthong (see Tables 2, 3, 4 in the Appendix) as they are not frequent in present speeches and can't be processed statistically according to their position in a syntagma.

3. RESULTS

Tables 2, 3 and 4 (see the Appendix) show the duration of vowels in this study.

In the tables above vowels' duration is broken down by type of the stress and the position of a sound in a syntagma (the initial position, the middle position, the final position).

As can be seen in Tables 2, 3 and 4 the vowels under syntagmatic stress are longer than those which are under word stress. The difference of vowels' durations under these types of stress fluctuates from 9 to 45 ms.

The position of a vowel in a syntagma also has a significant impact on the vowel duration. Stressed vowels in the final position in the syntagma are longer than stressed vowels in the initial position due to preboundary lengthening [13]. This is true with the respect to the following vowels: [æ] (D1, D2, D3), [e] (D1, D2, D3), [ɪ] (D1, D2, D3), [iː] (D1, D2, D3), [ʌ] (D1, D2, D3), [ɔː] (D2). The difference of vowels' durations under these types of stress is 5 – 49 ms.

The cases of vowels lengthening in the middle position of a syntagma can be explained by the presence of emphatic or logical stress (due to the peculiarities of political speeches studied in the paper). Thus, vowels [iː], [æ], [aː] (D2) and [iː] (D1) in this position are longer than those in other positions in syntagma. This phenomenon is going to be studied in further works.

4. REFERENCES

[1] Brend, R.M. *From Phonology to Discourse Language data. Amerindian Series,* 9.

[2] Hitchcock L., Greenberg S. 2001. Vowel Height is Intimately Associated with Stress accent in Spontaneous American English Discourse. *The 7th European Conference on speech Communication and Technology* (Eurospech-2001).

[3] Вольская, Н.Б. 2004. О паузе и не только о ней. *Фонетические чтения в честь 100-летия со дня рождения Л.Р. Зиндера.* СПб.: Филологический факультет СПбГУ.

[4] Щерба, Л.В. 2008. *Языковая система и речевая деятельность*. Под ред. Л. Р. Зиндера и М. И. Матусевича. Изд. 4. – М.: Издательство ЛКИ.

[5] Ахманова, О. С. Словарь лингвистических терминов. Изд. 4-е, стереотипное. – М.: КомКнига, 2007.

[6] Бондарко Л. В., Вербицкая Л. А., Гордина М. В. 2004. *Основы общей фонетики*. 4-е изд. СПб.: Филологический факультет СПбГУ.

[7] Брантов, С. А. 2004. *Просодическая составляющая аргументации в публичной речи*. Диссерт. на соискание учен. степ. к.ф.н. – Москва.

[8] Вольская, Н. Б. 2007. Вариативность интонационных моделей в спонтанной речи и чтении: интонационный центр. *Материалы 36 Международной филол. конф*. Вып. 17: Фонетика. СПб.: Факультет филологии и искусств СПбГУ.

[9] Мурзалина, Б. К. 2004. О контекстуальной обусловленности интонационного маркирования слова в тексте. *Вопросы русского языкознания*. Вып. 11. М.: Изд-во МГУ.

[10] Зиндер Л. Р. 2007. Общая фонетика и избранные статьи. 2-е изд., исправл. и дополн. СПб.: филологический факультет СПбГУ.

[11] Торсуев Г. П. Проблемы теоретической фонетики и фонологии. Изд. 2-е доп. – М.: издательство ЛКИ, 2008.

[12] Language files: Materials for an Introduction to Language and Linguistics. 7th Edition. The Ohio State University, 1998.

[13] Sluijter A. M. C., Heuven V. J. van. 1996. Spectral balance as an acoustic correlate of linguistic stress. JASA.

[14] netfiles.uiuc.edu/cls/www/course/ealc550/readings/sluijter_vanheuven.pdf

APPENDIX

Table 2: Vowels mean duration in D1 speech.

Vowels	Average vowel duration (syntagmatic stress)			Average vowel duration (word stress)		
	The initial position in a syntagma (ms)	The middle position in a syntagma (ms)	The final position in a syntagma (ms)	The initial position in a syntagma (ms)	The middle position in a syntagma (ms)	The final position in a syntagma (ms)
æ	85	87	92	64	65	70
e	70	73	73	55	57	68
ɪ	61	59	66	50	52	72
iː	90	100	100	69	71	80
ʌ	65	61	82	50	60	68
ɑː	100	105	118	106	96	113
ɒ	-	74	125	-	57	69
ɔː	-	125	129	79	72	115
əː	98	97	-	-	60	77

Table 3: Vowels mean duration in D2 speech.

Vowels	Average vowel duration (syntagmatic stress)			Average vowel duration (word stress)		
	The initial position in a syntagma (ms)	The middle position in a syntagma (ms)	The final position in a syntagma (ms)	The initial position in a syntagma (ms)	The middle position in a syntagma (ms)	The final position in a syntagma (ms)
æ	73	86	90	72	86	84
e	60	64	77	54	57	68
ɪ	63	69	54	50	46	55
iː	-	157	118	65	81	102
ʌ	53	55	70	60	55	64
ɑː	123	150	137	92	102	110
ɒ	65	58	55	51	49	55
ɔː	82	120	144	99	114	117
əː	104	141	132	124	100	110

Table 4: Vowels mean duration in D3 speech.

Vowels	Average vowel duration (syntagmatic stress)			Average vowel duration (word stress)		
	The initial position in a syntagma (ms)	The middle position in a syntagma (ms)	The final position in a syntagma (ms)	The inintial position in a syntagma (ms)	The middle position in a syntagma (ms)	The final position in a syntagma (ms)
æ	90	90	120	55	63	104
e	75	64	70	57	58	70
ɪ	63	56	-	47	53	76
iː	93	128	98	45	72	82
ʌ	69	58	98	52	60	70
ɑː	-	103	120	110	96	111
ɒ	-	74	125	-	57	69
ɔː	-	126	130	76	66	120
əː	-	97	-	-	56	99

SOME PROSODIC CHARACTERISTICS OF THE INTONATION GROUPS NOT SEPARATED BY PAUSES

Valery V. Kisseliov

Amur State University, Russia
kvv888@list.ru

ABSTRACT

Normally, any intonation group is bounded on both sides with pauses, or breaks in phonation. However, in some cases the adjacent intonation groups in a phrase are not separated with a pause, supposedly due to lower semantic value and relatively increased tempo of speech. This paper studies prosodic parameters of such intonation groups as well as supposed semantic factors responsible for formation of this type of intonation groups.

Keywords: intonation group, pitch, tempo, intensity, pause.

1. BACKGROUND

The purpose of this paper is to draw attention to the intonation group, as one of the major phonetic units, and to those intonation groups, which are not marked with pauses on one or either sides. As is known, phonetics investigates speech from the positions of unity of its material structure and semantic meaning, and possesses two own strata, or tiers, of objects – segmental and suprasegmental. Within the latter one may distinguish suprasegmental units and suprasegmental means. In our opinion the hierarchical organisation of suprasegmental units is as following:

1. Phonetic (phonological) text
2. Phrase (phonological phrase)
3. Intonation group aka syntagma
4. Phonetic word
5. Syllable

The smallest of the suprasegmental units is the syllable, the unit of pronouncing. The phenomenon of accent, i.e. making prominent by phonetic means of one syllable in a speech continuity from the adjacent ones, serves for discrimination of larger language units and brings us to the level of the phonetic word, which combines a basic meaningful word with its unaccented environment.

The next phonetic level is presented by the intonation group, in which semantic and phonetic elements are also combined. The intonation group is a minimum unit in which intonation patterns of a certain language are realized. Alongside the syllable the intonation group is a unit of pronouncing, though not only segments, but also suprasegmental means, collectively called intonation, are realized in it.

One or several intonation groups form the phrase, or phonological phrase, which is defined as the basic unit of speech expressing a finished thought, a semantic unity, the integrity of which is created by intonational means as well as by a certain syntactic structure. An obligatory attribute of a phrase is that it is considered to be marked with pauses on both sides whereas boundaries of an intonation group are not always marked. A stretch of a speech continuity, longer and possessing a more complex structure than a phrase, should be referred to as the phonetic text.

In Russian linguistic tradition there have been suggested some other terms for stretches of speech between the phrase and the text, like a “complex syntactic whole” or a “superphrase unity”. But unlike the “phrase” and the “text”, none of these terms has a direct bearing to phonetics (as they are used in the text linguistics), and do not correlate with any phonetic units equal to them in meaning or size.

This ranking of phonetic units is given to define the place of the intonation group within it and to stress its importance as a constructive unit for longer utterances. It is given in the Russian linguistic terminology, which may not strictly coincide with the possible classifications of phonetic objects existing in other linguistic traditions.

Some recent American and British classifications show both a similar ranking of phonetic objects and high diversity of terms labelling the same phenomena. Here are some examples in chronological order:

J. Pierrehumbert and M. E. Beckman (1988):

1. Utterance
2. Intermediate phrase
3. Accented phrase
4. Word
5. Syllable

Keating et al. (2003):

1. Utterance

2. Intonational phrase
3. Smaller phrase
4. Word
5. Foot
6. Syllable

Gussenhoven et al. (2004)

1. Utterance
2. Intonational phrase
3. Phonological phrase
4. Phonological word
5. Foot
6. Syllable

From the above classifications we can see that the greatest ambiguity lies in the intermediate level, i.e. at the level of the phonetic unit we call after A. Cruttenden the "intonation group", which in Russian we call "синтагма". As for A. Cruttenden, he chooses this term from a range of alternatives like: "intonational phrase", "phonological phrase", "phonological clause", "tone unit", "tone group", "sense group", "word group", etc.

According to M. Grice, there is no common structure at this intermediate level. What there is, is only a universal agreement that every intonation group possesses a tonal marking of some kind - a nucleus, or a focus, at least one prominent element or accented part. As for any other possible criteria to distinguish an intonation group in the process of segmentation they are still to be defined.

2. CURRENT STUDY

In order to contribute in searching a mechanism for segmentation we have undertaken measurements and analysis of two spontaneous texts produced by native English speakers, an American and a Canadian. The studied oral texts are narrations, discussing different subjects - sports and peculiarities of national cuisine. Both texts contain no obvious emphatic constructions, though speech of both speakers is not devoid of emotional colouring. The aim of this investigation is to compare prosodic characteristics of the intonation groups marked with pauses on both sides and those not separated with a pause and to find possible compensation for the lacking pause. Since the volume of the investigated material is very limited we could expect only preliminary conclusions.

The measuring procedure employed is based on the segmentation of the textual material into syllables and intonation groups. The syllable is used as a counting unit, and an intonation group represents a stretch of speech in which a pitch contour is realized. We consider a model intonation group as containing a full pitch contour and separated from its neighbours by pauses.

Segmentation was carried out with the use of the PRAAT program for the analysis of a sound signal. Procedure of measurements consisted of several stages: 1) acoustic analysis and segmentation, delimitation of syllables, intonation groups and pauses; 2) measurements of quantitative parameters of the pitch movements; 3) calculations of temporal characteristics on the basis of the measurements.

The analytical procedure is brought to the comparison of results of the measurements according to a few parameters. In the intonation groups studied the number of syllables, and thus presence or absence of a scale, the character and range of the pitch movement, conformity to the rule of declination (each subsequent syllable shows smaller pitch reading) as well as the pattern of the pitch movement in the nuclear syllable (including tail) have been defined. The length and speed of pronouncing of each intonation group and phrase have been measured as well as acceleration and slowing down of speech tempo, the place and duration of pauses have been correlated with the position of each intonation group within a phrase and text. The intonation groups marked with pauses on both sides or only on one side have been singled out.

Phrases have been singled out basing on the semantic criterion, while boundaries of intonation groups have been defined predominantly on the basis of the presence of delimitation pauses as well as accomplished pitch contour.

The received material can be characterized as follows:

Speaker 1. The text, or utterance, is 2 minutes 14 seconds long. It comprises 22 phrases and 83 intonation groups. The phrases are mainly partitioned, the duration of intonation groups being from 374 ms up to 2363 ms, with mean value of 1106 ms. There are 15 intonation groups consisting of only one syllable. Four phrases consist of one intonation group, the largest number of intonation groups in one phrase is 13, while five phrases contain four intonation groups each. The mean tempo is 8.3 segments per second. There are 14 intonation groups in 10 phrases, which are not separated by pauses from the adjacent ones, these are mainly the first or second intonation groups in a phrase.

The length of pauses is from 105 ms to 1929 ms, only 14 of 69 pauses are longer than 1 second, and only half of them are located on the borders between phrases. There are 24 filled pauses, but only four of them are found on the borders between phrases. The duration of pauses and the pattern of the filled pauses distribution are most likely to de-

pend on the process of speech planning, sometimes there is also a wish to lay special emphasis on a certain single word. For example, the longest pause of 1923 ms separates the word *toast* forming the last intonation group of the text.

And now those intonation groups, which are not marked with a pause. What suggests that there is a separate intonation group? Five of 14 such intonation groups show a sharp change in speech tempo in addition to accomplished pitch contour, in other cases there is only an appreciable change of pitch: HR - 5, HF - 2, FR - 1, while only one such intonation group is characterized by LF.

Speaker 2. The text is 1 minute and 28 seconds long. It is divided in 12 phrases with 29 intonation groups. There are 18 pauses, only three of which are filled. The length of intonation groups is from 504 ms to 3248 ms, mean length being 1618 ms. Only two phrases consist of one intonation group, the largest number of intonation groups in a phrase is seven. Six phrases in this text consist of two intonation groups. The mean tempo is 10.4 segments per second. There are 10 intonation groups in five phrases, which are not separated by pauses from the adjacent ones, including three successive intonation groups in one phrase expressing enumeration.

The pauses separating phrases are basically longer, than those on the boundaries of intonation groups within the phrases – only one of eight such pauses is longer than 1second, whereas on the borders of phrases there are eight out of ten pauses longer than one second or very close to this value. There are three filled pauses, of which only one is located within a phrase and shows the boundary between two intonation groups, while two others separate not only phrases, but also mark the places of partial change of a topic. The first speaker never marks any change of topic by a pause of appreciable length.

In the intonation groups not separated by pauses there is usually a sharp change of speech tempo: six cases out of ten such intonation groups. Always, except for cases of enumeration, there is a sharp change of the pitch movement: mainly HR, in one case HFR is observed. The phrase with enumeration contains three successive. The first of them is marked with the sharp change of speech tempo with its insignificant increase to the last intonation group. The pitch contour in the first two intonation groups is LF, and HF in the last one.

3. CONCLUSION

There is some arbitrariness in terminology concerning a suprasegmental unit of intermediate level because of the diversity of terms used by different phonological schools and researchers.

In the material studied we failed to find any bright features distinguishing intonation groups not separated by pauses from those marked by pauses on both sides. Possible means of compensation like sharp change of tempo are not fully consistent and the question of compensation needs further investigation.

On the whole the investigation undertaken seems to fail, which urges us for continuation of search of other segmentation mechanisms and tools.

4. REFERENCES

[1] Cruttenden, A. 1990. Nucleus placement and three classes of exception'. In *Studies in the pronunciation of English: a commemorative volume in honour of A.C. Gimson*. London: Routledge, 9 – 18.

[2] Grice, M. 2005. Intonation. In *Encyclopedia of Language and Linguistics*, second edition. London - Amsterdam : Elsevier-Pergamon, 778 – 788.

[3] Gussenhoven, C. 2004. *The phonology of tone and intonation.* Cambridge: Cambridge University Press.

[4] Keating P., Cho T., Fougeron C. & Hsu C. 2003. 'Domain initial articulatory strengthening in four languages.' In Local J, Ogden R & Temple R (eds.) *Phonetic interpretation. (Papers in Laboratory Phonology 6)* Cambridge. University Press, 143–161.

[5] Pierrehumbert J. & Beckman M. E. 1988. *Japanese tone structure*. Cambridge, MA: MIT Press.

TRENDS IN THE SYLLABLES VARIATION OF MANDARIN CHINESE IN SPONTANEOUS SPEECH

Marina Lobacheva; Zhang Li; Li Xin

Amur State University, Russia
lomarinaa@mail.ru

ABSTRACT

The greatest difficulty students face when studying Chinese is understanding native spontaneous speech. The reason is that canonical "phonemic" patterns of Chinese syllables are often violated. Segmental peculiarities are not explored to a necessary extent. Neither they are reflected in language teaching books. The paper aims at describing the variations and "phonemic" structure changes with the attempt to make a descriptive grouping of the similar patterns of modifications. The results of acoustic and perceptual study enable to demonstrate modifications unexpected from the canonical point of view. As far The obtained results also enable to assume that syllabic variations of spontaneous Chinese speech samples studied may be a reflection of typical structural variations and should be regarded as a normal effect of language reality.

Keywords: phonemic, spontaneous, syllable, modification, grouping

1. INTRODUCTION

In Western European and Slavic languages, the phonetic segmentation of words into syllables does not coincide with the morphological boundaries. If we turn to a number of Asian languages like Chinese, Thai, Burmese and Vietnamese, and especially to the Far Eastern tonal languages, there is a special role of the Syllable. In European languages, minimal unit of the phonological system is the Phoneme, whereas in Chinese and other languages of the similar language family the minimal speech unit is the Syllable, words are made up of syllables and minimally can be the size of one syllable, morphemes can not be represented by a unit smaller than a syllable. Semantic divisibility in these languages does not go beyond the syllable and the syllable in this context is always associated with a particular meaning. The term syllable is considered a key term in any language research of the languages mentioned above.

Some authors describe the following relative features of Chinese syllables, using the theory of its hierarchical structure (initials, medials, finals).

Initials can be expressed by one consonant only. No consonant clusters are allowed. There are totally 21 initials in the Chinese language, e.g. /l/, /n/, /s/, /zh/, /ch/, /r/, etc. There is opposition of aspirated/non-aspirated consonants /b/-/p/, /g/-/k/, /d/-/t/. There is no palatalization of consonants in Chinese.

Finals can be expressed by monophthongs /a/, /e/, /u/, the diphthongs /ou/, /uo/, /ai/ or triphthongs /iao/, /uai/, /ing/. Thus, the Chinese Syllable forbidding any consonant clusters, allows combinations of vowels. Out the 35 finals there are 24 diphthongs and triphthong. The compound vowels and diphthongs or triphthong are combined according to certain rules, displayed in a table of Chinese phonetic syllables that can be found in any book for Chinese as a foreign language.

In the typical Chinese syllable structure the first sound is a consonant, the second – non-syllabic narrow vowel, the third – syllabic vowel, and the fourth – non-syllabic vowel or a final sonorant (which in the Chinese language serves as the second element of the final) [Задоенко, Шуин, 2010].

The originality of the phonological system of Chinese language is perfectly displayed by the theory of A. and E. Dragunovs, who offered to treat all Chinese finals in the form of three series. According to this theory, the finals of the Chinese language can be represented by the parallel five-membered series of /a/ and /ə/. The center of each group contains a vowel sound which is the kernel of the series. Further, the final core set the system as well as "zero" final may be presented in four types depending on whether they are attached to the initials directly or through the media /i, u, y /.

Thus, we consider two possible ways of division of Chinese syllable – the phonetic and phonological. If we consider the Chinese syllable phonetically, it is possible to produce a linear segment of the syllable division into components, if we consider the syllable phonologically, the articulation of a syllable leads to a certain hierarchy of phonologi-

cal units of different levels, with real sound combinations, including the finals, which can be represented as a coherent logic-based schemes.
An integrated approach to the Chinese syllable largely solves the problem of complex relations between the sound units of syllabic languages, which arise as a result of deep fusion of components of a syllable, their interpenetration. Such close relationships are especially clearly manifested in native speakers' spontaneous speech, where the fusion of initials and finals often leads to an inaccurate perception.
There are studies on tone variation in Manrarin Chinese [Xu, 1994, 1998, 2009]. However there are few studies if any concerning syllable segments variation.

2. MATERIAL AND METHOD

The current study is a pilot experiment to see some glimpses of syllable segments variation in connected natural Chinese Mandarin speech.
The subject was a native male speaker of Mandarin Chinese (Speaker 1), aged 28, a teacher of Chinese calligraphy. The speaker reported of no speech or hearing disorders and felt free with the microphone. The subject was seated in a quiet room to answer a number of questions. The questions were not given to the subject ahead of time and he was answering them spontaneously producing a number of monologues 1-3 minutes duration each (15 minutes total duration). He expressed his point of view on the topics (family life, social problems, arts, economics etc.) He was requested to speak the way he usually speaks to make his speech sound as natural as possible.
The speech was digitized at a sampling rate of 44 kHz and 16 bit quantization using Audacity.

2.1. Experiment 1 – acoustic study

PRAAT was used for acoustic analysis of syllables' constituents. The first step was to find the structurally modified syllables (target syllables), which pronunciation was different from the canonical variant, fixed in the dictionary. The total number of extracted modified units (tokens) was 30. No computer modeling was involved.
During the second stage the list of target syllables was given to the same speaker with the task to read them carefully.
The third step was to compare the same units segmented from spontaneous speech isolated syllables reading.
The comparison clearly demonstrated that none of the target syllables was the result "wrong" dialect. The phonetic manifestations of the syllables were the perfect canonical samples of Mandarin Chinese. The the target modifications might be viewed as spontaneous speech phenomena.

2.2 Experiment 2 – perception study

10 native speakers of Mandarin Chinese were involved in perception study as listeners (3 male and 7 female, aged 24-28, teachers of different aspects of Chinese).
All 30 tokens from the spontaneous speech of Speaker 1 were randomized and designed into a questionnaire to be given to the listeners. The task of perception experiment was to choose only 1 correct variant out of the 3 given in the questionnaire. Out of the three variants one was the reflection of the acoustic transcription of the target syllable, the other – the canonical variant of the target syllable, and the third – a syllable that was the closest alternative to the target one (with a different vowel or consonant depending on the target modification). The tokens were given to the listeners through the headphones. The segmented units were repeated as many times as the listener needed for completion of the task. All the 10 participants were worked with individually. The listeners understood the task clearly.

3. RESULTS OF PERCEPTUAL STUDY

According to an analysis of the perception experiment results, it is possible to make the selection of certain blocks of modification. Analysis of the data was based on the quantity (in percent) counting of recognizable syllables.
The first block consists of syllables, the correct perception of which was 100% for all listeners. In particular, such syllables were /rang/, /mian/, /jian/, /zhong/, /you/, /wo/. The obviously have not caused any difficulties in perception. However they present the smallest group of the tokens involved: only 6 out of of 30 syllables.
The second block included all the modifications of vowels in the final and medial. One of the trends here was monophtongization of historical diphthongs in spontaneous speech. The most probable reasons seem to be quick tempo and destressing. A typical example would be recognition of the final syllable /pai/ as /pa/ 50% of cases, while the rest 50% were perceived as /pai/. Another typical example would be the final syllable /zui/ which canonical model was correctly recognaized by 60% of the listeners. The other 40% of listeners misin-

terpreted it as /zi/. There also exists vice versa example of vowel change trend, connected with diphtongization of vowel finals.

Cases when the canonical models of target syllables were recognized by the minority of the listeners, definitely, present a special interest for a researcher. One of the examples would be the syllable /qu/ with the reduced vowel that recognized as /que/, where the final /e/ was recognized by 50% of the listeners, though it's historically absent in this syllable. Only 30% of the listeners recognized the canonical model of that syllable. Such result may be explained by quicker tempo of articulation and the syllable destressing in that particular part of the utterance. Another vowel modification occurred in the syllable /de/ which was perceived as /da/ by 60% of the listeners. Another 30% recognized the canonical model of the syllable successfully and only 10% did not recognize it at all. Such result my be explained by considerable weakening of the syllable that was located in the final position of the syntagma. Another explanation might be that the final position is usually characterized by the falling pitch pattern and slowdown of speech rate due to which a more opened a-shaped sound was used by the speaker.

The most numerous block was the modification of consonants, which in its turn was divided into 3 subgroups, modifications associated with aspirated initials, sibilant and fricatives initials and nasal finals.

The first subgroup consists of 3 examples of aspirated initials modifications, including /h/ in /hao/ detected as /p/ in 50% of cases, another 50 % of listeners correctly recognized the canonical model. The initial /t/ in /tai/ was perceived as /p/ or a /k/ in 50% and 10% of cases respectively. Of special interest there was the recognition /x/ in /xian/ as /p/ by 70% of the listeners. All of these are initially phonologically aspirated, perhaps this fact was the basis of their phonetic similarity, and homonymous perception, as is well known that aspiration of Chinese consonants initials is considered one of the most striking feature of the language. In our case it might have masked the consonant's locus.

The second subgroup includes sibilant and fricatives modifications of the initials. Thus, fricative /ch/ (in cheng) was recognized as /r/ by 30% of the listeners, /sh/ (in shi) was perceived as /r/ by 100% of the listeners, /sh/ (in shuo) was mistaken for /zh/ 80% of the listeners responses. The probable explanation might be that the listeners deprived of wider context can not ignore the modifications mentioned as they usually do perceiving the flow of speech and choose the closest match to what they hear from the same group of forelingual fricatives and sibilants.

Modifications and omissions of the nasal finals formed the third subgroup. In particular the presence of back lingual /ng / (deng) was not recognized by 70% of the listeners who chose the syllable /de/ instead of /deng/ from the questionnaire.

The syllable /hen/ was also perceived successfully only by 30% of listeners. It was recognized by 70% of the listeners as having backlingual final /ŋ/ – the syllable /heng/ was chosen from the questionnaire. Such cases of omission and mistaking forelingual nasal for the back lingual have been described in scientific literature and may be explained from the points of positional changes, in particular, N. Speshnev wrote that the "final nasal of Chinese language can lose the individual characteristics"[Спешнев, 2006].

The forth block included syllables of so-called drastically reduced syllables, recognition of which took some odd direction for the majority of the listeners. E.g. syllable /zhong/, which canonical form was recognized only by 30% of the listeners while other interpreted it as /jiu/. Such a change may be due to the weakening of articulation in areas with a rapid rate of speech, the assimilation of u-shaped in the direction of i-shaped, which caused particular difficulties of perception.

Assimilation of the place of articulation may be an explanation for the modification of the syllable /xue/ to /yue/, since it has been restored by 100% of the listeners. To those we assign, the perception of the lingual sibilant /x/ as a glide /y/ and the recognition of forelingual unaspirated /z/ as forelingual nasal /n/ by 100% of listeners in both cases.

Such results only seem odd. They rather reflect real phonetic changes that the syllables underwent in spontaneous speech.

4. CONCLUSION

Turning to the fact, that the Syllable is the minimal speech and language unit in Chinese, the received data should be viewed from the angle of syllable constituents (initials, finals) as inseparable parts of the whole – the Syllable – but not as sounds viewed separately. As phonological units all the Chinese syllables are given by the listed number in the dictionary. As any invariant phonological unit they are represented in speech by a number of variants that are the result of positional and coarticulatory changes that spontaneous speech might turn out rather "extreme" compared to the canonical models.

In general, the received data shows that:
1) trends to diphtongization, monophtongization and positional modification of vowel initials and finals at the end of the phrase or are the result of increased speech tempo and destressing of certain syllables;
2) consonants initials and finals modifications deal mostly with locus masking, voicing/devoicing and omissions being the result of the same processes mentioned for vowels – increased tempo and destressing.
The received data should be regarded with the certain skepticism, because only 1 speaker participated and the experimental base should be enlarged to find other type of modifications. Results presented may be considered preliminary.

5. REFERENCES

[1] Задоенко Т. П., Хуан Шуин. 2010. *Начальный курс китайского языка.* 5-е изд. Москва: Восточная книга.
[2] Касевич, В. Б. 2006. *Труды по языкознанию*. Санкт-Петербург: СПбГУ.
[3] Спешнев, Н. А. 2006. *Введение в китайский язык*. Санкт-Петербург: Каро.
[4] Odden, D. 2008. *Introducing Phonology*. Beijing.
[5] Xu, Y. 1994. Assymetry in conraxtual tonal variation in Mandarin. *Phonetica*, 1994, 1-14.
[6] Xu, Y. 1998. Consistency of tone-Syllable alignment across different syllable structure and speaking rates. *Phonetica*, 1998, 170-203.
[7] Xu, Y. 2009. Organizing syllables into groups – Evidence from F0 and duration patterns in Mandarin. *J. Phonetics*, 37, 502-520.

THE INFLUENCE OF F0 AND F1 ON VOWEL PERCEPTION

Tatiana Lobanovskaya

St. Petersburg, Russia
tatiana0531@yahoo.com

ABSTRACT

According to the linguistic and psycholinguistic studies vowel phonemic interpretation depends on its formant structure. This research is devoted to the influence of Fundamental frequency (F0) on the First Formant (F1) of American English vowels pronounced by female voice and F0 and F1 influence on vowels' perception. 37 words containing monophthongs (except for /ə/ and /ɜ/) were analyzed. Stationary parts of vowels were segmented out. Threshold duration of each vowel was 30-50 ms. Two perceptual experiments were carried out. Natural vowels were given in the first experiment while the second one was composed of modified vowels with (i) F1 removed, (ii) F0 removed. Factorial analysis was performed to specify the influence of F0 and F1 on both natural and modified vowel perception. The results demonstrate higher influence of F0 on closed vowels perception compared to that of open and mid-open vowels. F1 removing had a considerable influence on the mid-open vowels perception; F1 modification had "clearing-up effect" on the open /ɔ/ and /ɑ/ vowels having more than 1 intensification in their F1 spectral part. F1 removing produced much greater influence on close vowels than on mid-open and open ones.

Keywords: prosody, perception, female voice, fundamental frequency.

1. INTRODUCTION

There are contradictory opinions concerning the role of Fundamental frequency (further F0) in vowel identification. The presence of F0 in vowel spectrum is obligatory (except for whispered vowels). However, in accordance with one point of view, while perceiving vowels listeners do not take the value of the F0 into consideration (see e.g. the work of R. O Coleman [Coleman, 1971].

It's a well known fact that vowel quality depends on configuration of epiglottis cavities, which perform the function of a resonator. Gunar Fant [Фант, 1964] stated the supposition that if F0 is high enough, acoustic pattern of vowels will be altered and as a result it will lead to changes in perception.

The study of synthesized vowels, where the F0 underwent different manipulations, demonstrated the dependence of perception on F0 movement. I. Lehiste and D Meltzer [Lehister and Meltzer, 1963] synthesized vowels with F0 distinctive for male and female voices, they based their experiments on work of G. E. Peterson and H. L. Barny [Peterson and Barny, 1952]. It turned out that vowels were identified better if there were both F0 and F1 in their spectrum.

Linguistic and psychophysiological researches also prove that vowel phonemic interpretation depends on higher linguistic levels [Peterson and Barny, 1952; Штерн, 1992; Бондарко, 1998; Picket, 1999] that are possible for a native speaker to reach any time due to his/her knowledge of the language.

The main goal of this research is to study the influence of F0 and the F1 of American English vowels pronounced by female voice on vowels perception.

Female voice was intentionally chosen because of naturally higher F0 values compared to the one of males voice reducing the acoustic distance between F0 and F1 giving the possibility of their overlap in close vowels. Formant values (F0, F1, F2) characteristic for American English are presented in Table 1.

In general, such overlap for male voice is excluded even for close vowels with the lowest F1 values (not mentioned in the phonetic literature at all) not to speak about mid-open and open vowels which F1 values are considerably higher. Mid-open and open vowels realized by female voice also do not present any difficulties as far as F0 and F1 (see fig.1-2).

On the one hand we can see that mean values of F0 and F1 do not coincide for close vowels produced by women (see Table 1). On the other hand, it is a well-known fact that F0 may vary considerably depending on the intonation pattern. There is a question in this connection: if a female speaker uses Rising intonation will F1 rise together with F0 or will they overlap?

Table 1: American vowels produced by man's (M) and woman's (W) voice (after Hillebrand et al. [Hillebrand et al., 1995: 3103]).

Formant	F0 (Hz)		F1 (Hz)		F2 (Hz)	
Phoneme	M	W	M	W	M	W
i	138	227	342	437	2322	2761
ɪ	135	224	427	483	2034	2365
ɛ	127	214	580	730	1799	2058
æ	123	215	588	669	1952	2349
ɑ	123	215	768	936	1333	1551
ɔ	121	210	652	781	997	1136
ʊ	133	230	469	519	1122	1225
u	143	235	378	459	997	1105
ʌ	133	218	623	753	1200	1426

Figure 1: /æ/ from “cat”**.**

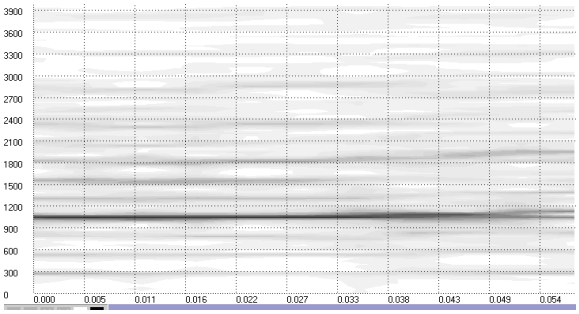


Figure 2: /ʌ/ from “cut”.

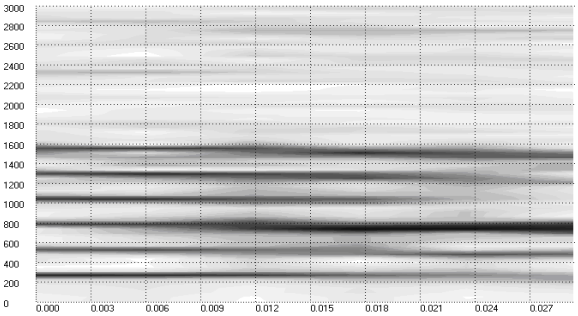
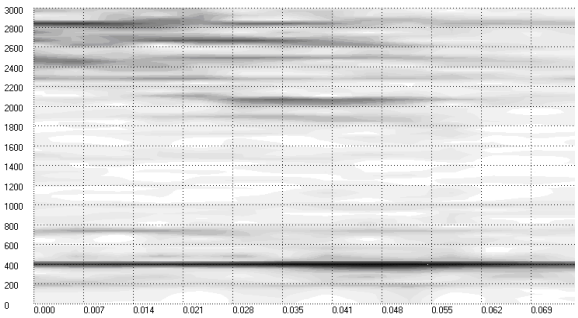


Figure 3: /i/ from “deeds”.



In order to answer these questions the following tasks were necessary to be solved:

1. To make detailed research of acoustic characteristics of high female voice, such as:
   - to get data about F0;
   - to define the magnitude of F0 effect on vowel perception;
   - to measure F1 mean value of all vowels within the speakers studied paying special attention to close front and back vowels.

2.To define dependence of vowel discriminability on F0 interaction with F1.

2. CORPUS

37 words containing monophthongs (except /ə/ and /ɜ/) were put in carrier sentences with different intonation pattern – narrative and interrogative (general questions). In narrative sentences the words were put into different positions: at the beginning, in the middle and at the end. In the interrogative sentences the words were put into the end position. According to the experiment every of the 37 words was pronounced with different pitch pattern (rise, fall, and level).

Sentences were read out by three female native speakers of American English aged 20-22. The analysis of F0 movement graph showed that 42 words were pronounced with level pitch, 22 words – with rising pitch and 47 words – with falling pitch.

After recording and digitizing all the sentences the target words were segmented out. Segmentation was carried out with the help of the visual and perceptual control according to the principles stated in the works of L.V. Bondarko [Бондарко, 1998] and P.A. Skrelin [Скрелин, 1999]. The boundaries between sounds were put on the waveform.

3. ACOUSTIC ANALYSIS

To make acoustic analysis stationary parts of the vowels were segmented out. Threshold duration of each vowel was not less then 50 mls. All received data were put into the table. Table 1 is the example of data after 1 female speaker (date received after two other speakers are not demonstrated here but taken into consideration in the present paper).

In the given table there are 3 full couplings of F0 and F1 (half-close front vowel /**ɪ**/ – 310 Hz, close front vowel /**i**/ – 231 Hz and close back vowel /**u**/ – 280 Hz), and there is one case of partial coupling of F0 and F1 (close front vowel /**i**/ – the meanings

of F0 and F1 – 251 Hz and 293 Hz correspondingly).

Table 2: Acoustic characteristics of F0 and F1 of vowels used by speaker M. in words with flat, rise and fall tone.

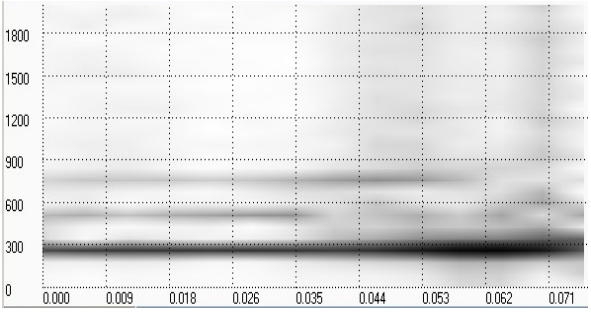
speaker	Vowel	Flat		Rise		Fall	
		F0	F1	F0	F1	F0	F1
M	ɪ	248	393	310	310	218	436
	i	231	231	251	293	230	395
	u			280	280	253	361
	ʊ	177	366	262	382		
	ɑ	236	710	231	769	215	613
	ɔ	290	581	190	550	234	532
	ʌ	183	602	247	684	183	695
	æ	181	684			204	700
	ɛ	269	527	215	505		

All these couplings are in the vowels pronounced by flat and rise tone (full coupling is marked by light grey color and partial is dark grey). In vowels pronounced by the fall tone there was no couplings fixed.

These couplings constituted the corpus for further perceptual experiment.

Figure 4 represents the vowel /ʊ/, segmented out from the word put. We can see 3 intensifications of frequencies. The first intensification may be considered F0 (260 Hz), the second one, which is on the level of 516 Hz, theoretically corresponds to the F1 (see Table 1). But it is very high value for the particular female speaker (see Table 2).

Figure 4: /ʊ/ from put.



It rather corresponds to F1 value characteristic for the central close-mid /e/ or back mid-open /ɔ/ then to the vowel /ʊ/. In other words if we consider 516 Hz value to be characteristic for F1 of the vowel /ʊ/, we should speak about qualitative changes of this sound or consider it harmonic although it is very intensive to be harmonic. In this case we can suppose that F0 and F1 are either very close to each other or merge.

4. PERCEPTION

The problem of perception of high female voice is one of the most difficult in phonetics. In acoustic point of view vowels are characterized by distinct formant structure and by a large number of harmonics approximately of the same intensity, situated on the whole spectra of the sound and correspondingly they can be found close to the inner frequencies of the vocal tract. That is why it is difficult to define formants' values. It is especially difficult for the close vowels /i/ and /u/ because their F0 and the F1 are located very close to each other or even merge.

Two perceptual experiments were carried out.

4.1. Experiment 1

94 vowels, segmented out of the words, used in the carrier sentences with level, rising and falling tone were given to the listeners for a discrimination task. They were to write down sounds they hear and then example of the word, where this sound can be used.

There were two groups of volunteer listeners – native speakers of English (group of Americans who came to Saint Petersburg to study Russian), the second group was the group of the First year students of the department of Linguistics of Saint Petersburg State University. Both groups were not told about the language which sounds they were going to listen to. The results of the discrimination task are presented in Table 3.

Table 3: Discriminability of vowels by the two groups (%): Americans (Group A) and Russians (Group R).

Phoneme Group	i	u	e	æ	ɑ	ɔ	ʌ
A	83	87	71	63	52	48	42
R	85	64	76		81		10

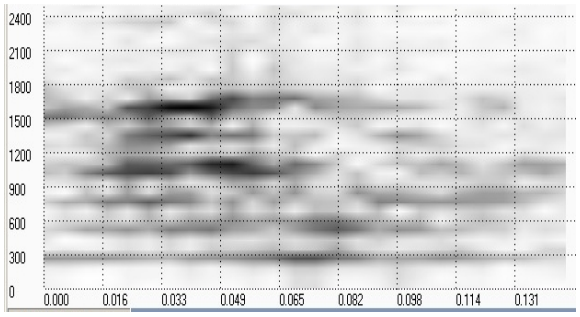
The data of Table 3 enables to say that both groups identified close vowels /i/ and /u/ quite well. Group A was better at identifying close-mid /e/, open-mid /æ/ and /ɔ/, open /a/ and open-mid /ʌ/, while Group R was good at identifying /i/. However

Group R confused /u/ and /ɔ/ (33%), /a/ and /ɔ/ (29%) and /ɔ/ and /ʌ/.

4.2. Experiment 2: Perception of so called "obscure" vowels

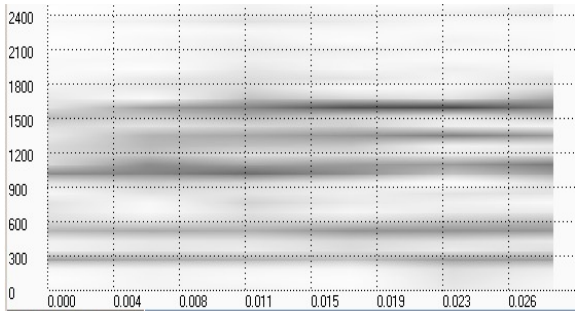
The results of the last experiment provided implications for the second experiment. Because there were problems in identification of vowels: listeners confused them or didn't hear at all.
Bearing in mind our main goal to define the influence of F1 and F0 on vowel perception we decided to remove, first, F1 and, second, F0 and see what effect the removing would produce on the listeners. For this purpose we used reject filter. This "operation" was done with all sounds, used in the first experiment.
For example, it was difficult to define vowels /ɑ/ -/ɔ/. On the sonogram on the figures 5 presenting /ɔ/ realization we can see two intensifications which can both refer to F1 – 800Hz and 554Hz.

Figure 5: sonogram of / ɔ / with two intensifications.



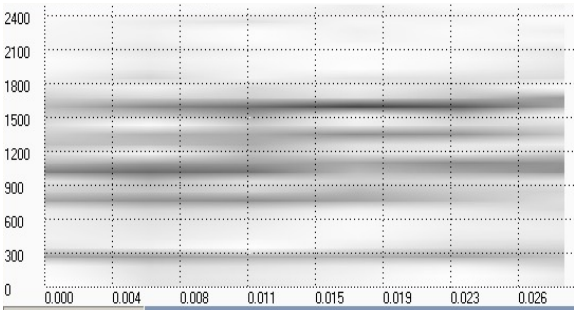
When the intensification on the level of 800 Hz was filtered out (see fig. 6), we had a "clearing-up effect": the listeners were more confident defining the sound as /ɔ/ - 69 % (see Table 3).

Figure 6: / ɔ / after modification**.**



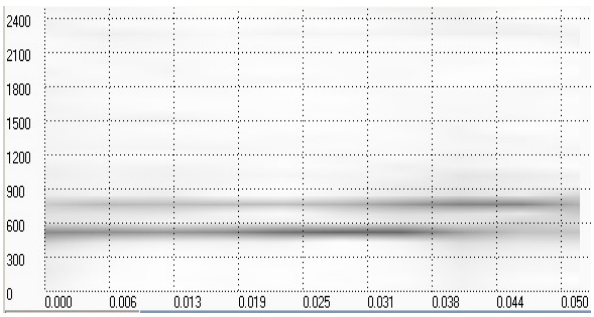
Thus we can speak about considerable increase of discriminability in comparison with the first experiment – 48% (21% increase, which is above chance). Such result was bewildering because 800 Hz is more characteristic of /ɔ/ for female voice (see Table 1).
When we filtered out intensification on the level of 554Hz in /ɑ/ realization preserving the intensification on 800 Hz level (see fig. 7), again there was a "clearing-up" effect observed: the listeners were more confident defining the sound as /ɑ/ – 77% in comparison with 52% in the first experiment (25% increase, which is above chance).

Figure 7: /ɑ/ after modification.



The same manipulations were done with close vowels /u/ which was mixed with the sound /ɔ/. As a result 92 % of the listeners defined this sound as artificial, or nonhuman, or did not hear any vowel at all. Similar effect was obtained for /i/ with F1 removed.
Obviously F1 removing had much more devastating effect on close vowels discriminability than that on mid-open and open vowels.
The next step was F0 removing and F1 preserving for the same "obscure" originally segmented vowels. The intensification in the spectral part corresponding to F0 was removed up to 300 Hz. Fig. 8 represents the example of close /ʊ/ from *put* with the F0 removed.

Figure 8: /ʊ/ from "put" (F0 removed).



These modified sounds were prepared for the same groups of listeners after 3 months. For both

groups perception of these sounds was unusual. They called them technical and nonhuman.

So, in general a person can identify sounds by the first two formants and this fact was proved by Picket when he made experiments with synthesized vowels.

According to the table 4, the recognition of these modified vowels is much lower in comparison with the first experiment.

As in experiment 2 removing of F0 had more devastating effect on close vowels, while discriminability deterioration of other vowels was not that drastic.

Table 4: Perception of vowels without F0 by Americans and Russians (%).

	i	u	e	æ	ɑ	ɔ	ʌ
Americans	9	21	63	41	28	33	32
Russians	10	8	44		39		9

5. CONCLUSION

- Though F1 and F2 are enough for describing vowel quality, F0 is an integral part of a vowel F-picture. Its absence produces the effect of general deterioration of vowel perception.
- F0 removing effect is considerably weaker for open-mid and open vowels discriminability than for close vowels.
- F0 removing influences the perception of close vowels in an extremely devastating way, which may be considered a proof for the possibility of F0 and F1 merger for these vowels when pronounced by high-level female voice. This merger performs the function of F1. Its removing together with F0 drastically decreases close vowels discriminability.
- 800 Hz intensification removing from /ɔ/ spectrum has an interesting implication about very small acoustic and, as a result, small perceptual distance between natural American English /ɔ/ and /ɑ/.
- Another implication connected with the previous one is for further research: it would be interesting to do a reverse procedure – 554 Hz intensification removing from /ɔ/ and 800 Hz intensification removing from /ɑ/ – and seeing if the perceptual effect would be reverse too.

6. REFERENCES

[1] Бондарко, Л. В. 1998. *Фонетика современного русского языка*. Санкт-Петербург: СПбГУ.

[2] Венцов А. Б., Касевич В. Б. 1994. *Проблемы восприятии речи*. Санкт-Петербург: СПбГУ.

[3] Скрелин, П. А. 1999. Сегментация и транскрипция. Санкт-Петербург: Санк-Петербургский государственный университет.

[4] Чистович Л. А., Кожевников В. А. и др. 1965. *Речь. Артикуляция и восприятие*. Москва-Ленинград.

[5] Фант, Г. 1964. *Акустическая теория речеобразования*. Москва.

[6] Штерн, А. С. 1992. *Перцептивный аспект речевой деятельности*. Санкт-Петербург: СПбГУ.

[7] Coleman, R. O. 1971. Male and female voice quality and its relationship to vowel formant frequencies. J. *Speech and Hearing*, vol. 14, 565-577.

[8] Hillebrand J., Getty L. A., Clark M. J., Wheeler K. 1995. Acoustic characteristics of American English vowels. JASA, 97 (5), 3099-3111.

[9] Lehiste I., & Meltzer D. 1963. Vowel identification in natural and synthetic speech. *Language and speech*, 16, 356-364.

[10] Pickett, J. M. 1999. *The acoustics of speech communication*. Boston, London…: Allyn and Bacon.

[11] Peterson G. E., & Barny H. L. 1952. Control methods used in the study of vowels. *JASA*, 24, 175-184.

REDUCED VOWELS IN AMERICAN ENGLISH

Yulia Lugantseva

Amur State University, Russia
julialbest@mail.ru

ABSTRACT

There are many concepts of speech perception: from motor theories to models of Lexical Access, none presenting a universal model but being important parts of the whole picture. Perception of a foreign language differs a lot from one of the native language. Phonetic peculiarities that a native speaker deals with automatically may present a serious obstacle to a foreign listener. One of these is the "schwa" sound. Acoustic analysis demonstrated that there are several types of "schwa" depending on the surrounding consonants. Another finding was that certain allophones of the neutral sound acoustically (by FI and FII values) coincide with the allophones of other phonemes namely /ɛ/, /ɪ/ and some others. We hypothesize that this peculiarity might be reflected in native speakers' perception in limited context and partly prove the hypothesis. Such result can be explained by differences between traditional and dynamic phonology the last being important for training the perception of learners of English.

Keywords: schwa sound, vowel reduction, perceptual, acoustic.

1. INTRODUCTION

Today spontaneous speech is in the focus of linguists' interest. It can be easily explained because only in spontaneous speech we can observe that many modifications of phonemes and tone units. Many of them are not noticed by the native speakers' perception who have context at their disposal and recognize words without complete identification of each speech sound with the corresponding phoneme and construction words out of these phonemes.

There are many concepts of speech perception: from motor theories to models of Lexical Access, none presenting a universal model but being important parts of the whole picture. Perception of a foreign language differs a lot from one of the native language. Phonetic peculiarities that a native speaker deals with automatically may present a serious obstacle to a foreign listener. One of these is the "schwa" sound.

The aim of this research is to study acoustic and perceptual properties of American unstressed vowels in spontaneous speech. We hypothesize that (i) the schwa sound is a rarer product of vowel reduction than it is traditionally considered; (ii) the quality of reduced vowels depends very much on the surrounding consonants. To prove the hypothesis acoustic analysis and three perceptual experiments were carried out.

2. ACOUSTIC ANALISYS

2.1. Method

2.1.1. Participants

The participants were three volunteer male native speakers of General American English aged 27-35 without any speech or hearing disorders. All of them felt comfortable with the microphone. All the three speakers reported on their educational background as having Bachelor of Arts degree.

2.1.2. Procedure

Each speaker produced spontaneous utterances answering a list of questions they have not seen before. No time for preparation was given. The task was to speak on a question as much as they can. The speakers could ignore a question if they for some reasons did not like it.

The recordings were made directly to the computer using Audacity at a sampling rate of 44 kHz and 16 bit quantization. The utterances were segmented using Wave Assistant. 395 reduced unstressed vowels were selected for acoustic analysis. The values of the FI and FII were measured from spectrograms. The vowels were divided into several sound types based on FI and FII. The values of FI and FII were compared to the values published in current phonetic literature.

2.2. Results

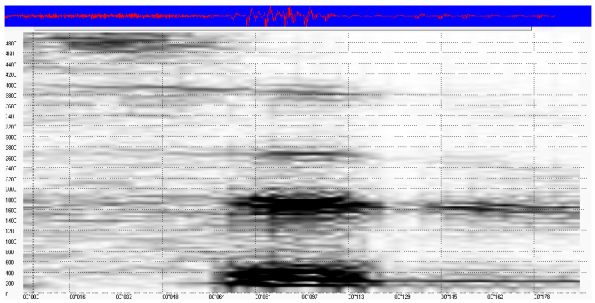
The results of the acoustic study are demonstrated in Table 1.

Table 1: Acoustic study of reduced vowels in American English.

Sound type	The mean values of FI and FII (Hz)	Frequency of usage (%)
ɪ-like (Group 1)	FI — 388, FII — 1736	61
ɛ-like (Group 2)	FI -529, FII-1554	20
ʊ/ɔ-like (Group 3)	FI -511, FII – 1004 The mean value of FI is closer than /ɔ/ (FI – 657-678) but lower than /ʊ/ (FI - 401-423)	11
ɜ-like (Group 4)	FI -472, FII -1473	4
schwa sound (Group 5)	FI - 579-775 FII — 1257-1533	4

The results demonstrate that the most frequently used unstressed vowel is ɪ-like vowel. This vowel usually appears as a result of coarticulation with the surrounding forelingual consonants except for lateral /l/ and retroflex /r/. Figure 1 shows one of the typical examples.

Figure 1: The spectrum of the [sɪd] part from the word "disadvantage".



The formant values of the reduced vowel taken from the word "disadvantage" are FI – 410 Hz, FII – 1720 Hz. Such a realization was influenced by forelingual consonant context and distant assimilation with the /ɪ/ vowel in the preceding syllable, which receives secondary stress. Compare the two vowels given in fig. 2.

Another example, the word "significant", shows ɪ-like vowel (FI- 410 Hz, FII slope 2000-1775 Hz) that occurs between velar /k/ and forelingual nasal /n/ (see fig. 3). It is another typical context for ɪ-like vowel.

The ɛ-like vowel occurs between forelingual and nasal consonants. The study shows, that this reduced vowel is frequently found in suffixes like – **ment, -ous**. Fig. 4 gives a typical example.

Figure 2: The spectrum of the [dɪsɪd] part from the word "disadvantage".

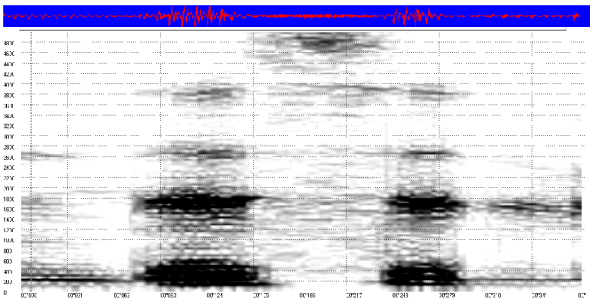


Figure 3: The spectrum of the [kɪn] part from the word "significant".

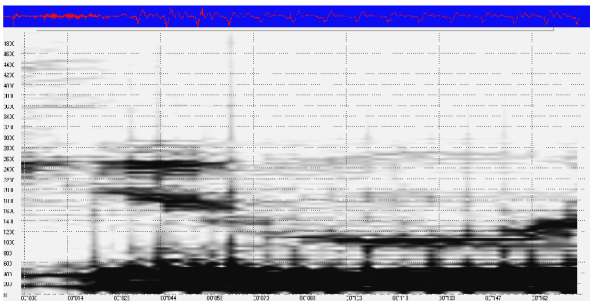
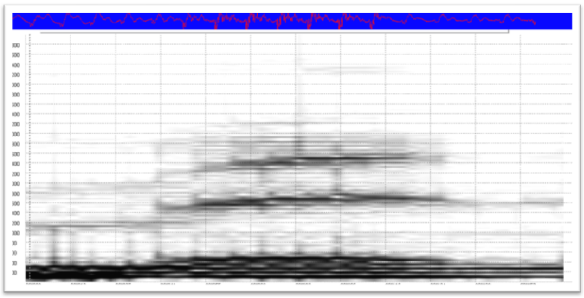
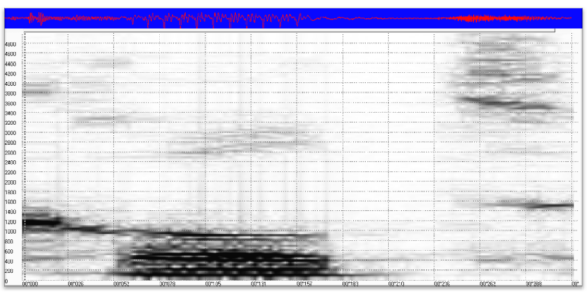


Figure 4: The spectrum of the [mɛn] part from the word "government".



The presence of the post vocal lateral /l/ determines the use of the ʊ-like reduced vowels. Like [ʊ] in the word "difficult" (FI – 520 Hz; FII – 927Hz).

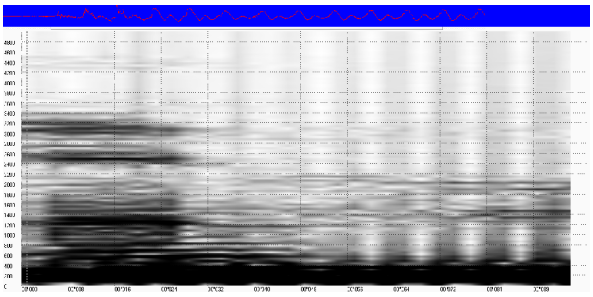
Figure 5: The spectrum of the [kʊlt] part from the word "difficult".



The next group of vowels is ɜ-like vowels with the mean values of FI -472 Hz, FII- 1473Hz. Such vowel occurs between fricative alveolar /z/ and labio-dental /v/. It should be mentioned, that in American English such vowel is rotocized, that is why we supposed that perception of such segments might cause a lot of difficulties.

According to current phonetic literature [2] the mean formants of the schwa sound are FI -500 Hz, FII – 1500 Hz. The measurements from spectrograms of isolated words such as *above*, *pizza*, where the schwa is surely present, show that the mean formant values are FI- 737 Hz, FII – 1327 Hz. These values are very close to those of /ʌ/ with the mean values for FI -749 Hz, and FII – 1150 Hz. However, it might be more fronted and a little close in other isolated words in the same positions. The FI value varies from 579-775 Hz, FII — 1257-1533 Hz. For example, the schwa sound at the beginning of the word "American".

Figure 6: The spectrum of the syllable [əm] form the word "American".



Studying the acoustic properties of reduced vowels in spontaneous speech, we found clear dependence of such vowels on consonant context. According to the mean formant values reduced vowels were divided into groups that correlate with their stressed counterparts: e.g. ɪ-like schwa type and stressed /ɪ/. However, their FI and FII values show that reduced vowels are closer. Such results correlate with those of Steven Greenberg and colleagues [Greenberg at al.].

3. PERCEPTUAL STUDY

In order to study perceptual characteristics of the reduced vowels and see whether the suggested division will have any reflection on native speakers' perception three experiments were carried out.

3.1. Experiment I

3.1.1. Subjects and procedure

The subjects were 12 native speakers of American English, who were not familiar with the talkers' voices. The subjects listened to the stimuli and performed the discrimination task. They were asked to put down what they hear using the letters of the English alphabet. They were also requested to set an example of a word or word combination, where they could hear such sounds. Each stimulus was recorded three times with the interval of 2 sec. The stimuli were not modified except for intensity increase in order to increase the volume of the sound and contained the reduced vowel in minimal consonant context CVC to reduce the chance that the listeners could guess the word the stimulus was taken from. The listeners were informed that those were not words but parts of words (taken from a word or two neighboring words).

It is a well-known fact that our perception is of lexical character: performing any discrimination task we always try to recognize words in any sound sequence. The perception of non-sense sound sequences does not differ much from sequences of words [Kingston et al, 2009: 310] except for quicker listeners' reaction to the latter [Fowler et al., 1981]. According T. Chugaeva [Чугаева, 1998, 2009] the important cues for native English speakers are 1) the beginning of the word, 2) stressed vowel, and 3) the number of syllables. Listeners look first of all at these three so-called "hyper features" of words and are able to recognize them leaving other phonetic features behind. So the variation degree of the latter being very high remains unnoticeable for the listener perceiving the flow of speech and thus having much wider context at their disposal. The stimuli were designed the way described above to deprive listeners of that wide context and make them concentrate much more on the acoustic features of the reduced vowels than they usually do.

3.1.2. Results

The results of the discrimination task demonstrated that the stimuli of Group 2 -- with the ɛ-like vowel were identified as the words containing phoneme /ɛ/ with the discriminating rate varying from 58% to 92% for different stimuli in this group. In some cases the listeners decided that such stimuli contained phoneme /ɪ/, which is supposed to be perceptually close to /ɛ/ (see more

about perceptual similarity of English vowels in [Nishi et al., 2008: 581]).

The discrimination rate of the stimuli of Group 1 – with the ɪ-like vowel – is quite high in general exceeding 60% and even reaching 100% for some stimuli (e.g. the one taken from *the student*). Very often this vowel was interpreted as perceptually close /ɛʔ**.** It coincides with the results of Group 2 discrimination pointing once again to perceptual similarity of the two vowels.

But in some stimuli containing this vowel the discrimination rate was unexpectedly low, like 35% for the one from the word "supposed". It might be the result of very small duration of the vowel.

The stimuli of Group 3 – with ʊ/ɔ-like vowel were interpreted as containing the /ɔ/ vowel with the discrimination rate of 75%-84%.

There were only two stimuli of Group 4 – with ɜ-like vowel. Only two speakers interpreted the vowel as in the word it was taken from. The recognition rate of the reduced vowel as /ʊ/ is 42%.

The perception of the second stimulus of the same group – the one segmented from the word "preservation" – was a little bit different. 33% of the listeners decided that the stimulus contains the /ʊ/ vowel, while 33% decided that it contains the /ɜ/ vowel. It should be taken into account that the /ɜ/ vowel is rotacized in American English – /ɜr/. As soon as rotacizing appears, the percent of its interpretation as /ɜ/ increases.

Non of the stimuli of Group 5 produced the major perceptual effect of schwa sound in spite of the fact that the formant values were very close to FI – 500, FII – 1500 that are given for this vowel. Thus the vowel in the stimulus taken from the word "Canada" was recognizes as /ɛ/ by 92% of the listeners. The discrimination rate of the vowel as /ɛ/ in the stimulus "domestic" was 57%.

The results of Experiment I demonstrate that the stimuli with the vowels /ɛ/, /ɪ/ have the highest rate of recognition, while the stimuli with the vowel /ɜ/ is expected the least successful. The stimuli with the schwa sound usually recognized as stimuli containing the sound /ɛ/.

3.2. Experiment II

3.2.1. Method

For Experiment II the same stimuli were used, but the listeners were given the variants from which to choose. The variants were designed taking into account the results of Experiment I. Another group of 21 volunteer speakers of American English participated in the experiment.

The listeners were to rate the variants written in the questionnaire on the 10-point scale, with the 10 points given to the most probable variant and 1 — to the least probable one.

3.2.2. Results

The results of Experiment II correlate with the ones of Experiment I. For Group 2 stimuli the words with the vowel /ɛ/ got 7 points. In general the rate of recognition was 70%. In some cases the listeners suggested perceptually close variant /ɪ/ as the other group of listeners did in Experiment I. For Group 1 stimuli the listeners gave from 7,8 to 9,9 points to the words containing the phoneme /ɪ/. For Group 3 stimuli the words with the phoneme /ɔ/ were rated from 5,6 to 7 points, which is above average.

As far as Group 4 stimuli with the /ɜ/ vowel, the listeners tended to give the maximum points to the words with the vowel /ɚ/. Thus the successful recognition of this vowel increased compared with Experiment I.

Group 5 stimuli were marked differently. The recognition rate of the vowels as /ɪ/ was 70% (7 out of 10 points).

Traditionally if not considered a phoneme the schwa sound is stated to be the allophone of /ʌ/, however the words with this phoneme got only 4 points when the listeners heard the stimuli with the schwa sound.

3.3. Experiment III

3.3.1. Method

The stimuli for Experiment III were divided into two blocks. For both blocks they were synthesized from the ones used in Experiments I and II. In the first block a stressed CVC or CV sequence containing acoustically the same vowel was added to the unstressed CVC used in Experiments I-II (e.g. Team + society = [ˊtɪsɪs]). In the second block stressed sequence contained a different vowel (involve+ society = [ˋvɔlsɪs]). Both unstressed CVCs and stressed CVCs/CVs were segmented from the same speaker.

21 American speakers participated in the second experiment. The subjects listened to the stimuli that were repeated three times with the interval of 2 seconds. The listeners were asked to listened to

the stimuli and put down what they hear with the letters of English alphabet.

3.3.2. Results

The study shows that Group stimuli containing the reduced ɹ̩like vowel again were most successfully recognized – 70-76%. These results correlate with the ones of the previous experiments. As in the previous experiments perceptually close /ɛ/ was suggested too.

The same tendency was observed for Group 2 with ɛ̩like vowel, which received 60% of /ɛ/ responses and 30% of /ɪ/ responses.

Group 3 stimuli with the ɔ/ʊ̩like vowel on the contrary with Experiments 1-2 were interpreted by the majority of participants (57%) as /ʊ/, /ʌ/ responses received 29%, /ə/ responses – 10%, /ɔ/ responses – the remaining 4%.

According to the Experiment III the stimuli containing acoustically the schwa sound were often perceived as having /ɪ/ or /ɛ/ replicating the results of Experiment II.

In Experiment III the percent of interpretations of Group 4 stimuli with ɜ̩like vowel as /ɪ/ and /ɛ/ increased to 40% and 27% correspondingly.

The percent of /ɚ/ responses did not change very much compared to Experiment I (23%). However, Experiment II demonstrates much higher percent of /ɚ/ interpretation.

The study of acoustic and perceptual properties of reduced vowels demonstrates much lower use of the schwa sound than it is traditionally considered. The expansion of the schwa sound in American spontaneous speech is not proved. Instead the use of close and mid-close front vowels /ɪ/ and /ɛ/ is increased due to very frequent forelingual consonant context that influences the reduced vowel quality.

4. CONCLUSION

The results of acoustic study enabled to split reduced vowels into 5 groups according to their formant structure. The analysis of the grouping and the consonants neighboring the vowels demonstrated a clear correlation between the choice of vowel type (Group 1 vs Group 2 vs Group 3 vs Group 4 vs Group 5) and the consonant context. Group 5 including the vowels with the formant structure of the schwa sound turned out the least frequent disproving the postulate about schwa expansion in the flow of speech.

The results of perceptual study show that in limited phonetic context native speakers perceive vowels according to their acoustic features. Under these conditions the expansion of schwa is not proven either.

The results raise a number of questions.

1. Is schwa just a convenient technical solution for dictionary makers?
2. Is it the same convenient for English as a foreign language learners who expect the same pronunciation seeing the same transcription symbol of schwa and do not find one?
3. What is the implication for learners in this connection?
4. What is the pragmatic value of schwa for certain types of modeling American English speech production and recognition.

These are uneasy questions that have to be addressed further.

5. REFERENCES

[1] Деркач С.В. 2003. *Фонетические свойства гласных в спонтанной речи: экспериментально – фонетическое исследование на материале американского варианте английского языка*. Дис. ... канд. филол. наук. Благовещенск.

[2] Чугаева, Т. Н. 1989. Механизмы аудирования родной и иноязычной речи. Дис. ... канд. филол. наук. Ленинград.

[3] Чугаева Т. Н. 2009. Звуковой строй языка в перцептивном аспекте (экспериментальное исследование на материале английского языка). Автореф. дис. ... докт. филол. наук. СПб: СПбГУ.

[4] Fowler C. A., Wolford G., Slade R., Tassinary L. 1981. Lexical Access with or without awarness. J*ournal of Experimental Psychlogy: General,* 110 (3), 341-362.

[5] Greenberg S., Carvey H., Hitchcock L. 2002. The relation between Stress Accent and Pronunciation Variation in Spontaneous American English Discourse. *ISCA Workshop on Speech Prosody*. Aix-en-Provence.

[6] Kingston J., Kawahara S., Chambless D. et al. 2009. Contextual effects on perception of duration. *J. Phonetics*, 37, 297-320.

[7] Nishi K., Strange W., Akahane-Yamada R. et al. 2008. Acoustic and perceptual similiarity of Japanese and American vowels. *JASA*, 124 (1), 576-588.

[8] Pickett J.M. 1999. *The acoustics of speech communication*. Boston, London.: Allyn and Bacon.

TONICITY ASPECT IN THE PROCESS OF TEACHING ENGLISH PRONUNCIATION

Elena Makarova

Ural State University of Economics, Russia
makarovayn@mail.ru

ABSTRACT

Tonicity, or location of the nuclear accent can differ significantly across languages. Such contrasts definitely influence the process of a foreign language learning. The exploration and comparative investigation of tonicity rules in English, Russian and Spanish can help reveal mechanisms of nuclear placement choice in the English speech of the Spanish and Russian speakers of English and find ways of improving teaching methodology to help learners progress towards the native-like production in the field of tonicity.

Keywords: tonicity, nucleus, focus, tonicity errors

An appropriate use of English suprasegmentals is often referred to as a" hurdle which an absolute majority of speakers of English never manage to cross" [1], "a problem child of pronunciation" [4] and even a 'fortress" [12]. According to M.A.K Halliday [7: 62] learners of English have to make three types of decisions as they speak, namely:

1. Distribute an utterance into tone-groups, called Tonality
2. Locate the nuclear accent, called Tonicity
3. Choose a tune, called Tone.

One feature of English intonation among these three which does not differ among the different varieties of the English language but can differ significantly across different languages is nuclear placement. Such contrasts can definitely influence the process of foreign language learning. There are opinions that tonicity is a separate choice from the choice of a nuclear tone and therefore should be taught separately [14: 286]. Errors in this field according to many phoneticians belong to one category of phonological errors which can cause difficulties in communication [9].

However, in teaching pronunciation programs the focus has traditionally been placed on the choice of a tune. Tonicity as a teaching target has never been a number one priority and has mostly been completely neglected. The exercises used in the classroom today are usually aimed at expressing a certain attitude in a contextualized situation by using this or that nuclear tone. However, the teaching experience proves that Russian students have difficulties in nucleus placement, especially when nucleus does not go on the last content word in an intonation group. Russian learners of English often shift the nucleus from the "correct" places in utterances serving as models. The question the learners often ask is: "Can a different word be a nucleus in the same context? If not – why?" The question teacher of a pronunciation should ask is: "Will my students be able to use those patterns in spontaneous speech situation in a different context and where will the nucleus be?"

It is common knowledge that tonicity is said to mark the focal point of a tone group. Focal material is the one presented as new and/or important. If the part of the intonation group is considered to be out of focus, it is usually not accented, as the speaker assumes the listener is familiar with this information. In a neutral pronunciation the nucleus is placed on the last accented syllable/word which contains new information and doesn't express contrast or emphasis – I gave him a PEN (the so-called broad focus). If a speaker wants to focus on a particular piece of information we speak about "contrastive" focus – I GAVE him a pen (I didn't take it from him).This is the so called narrow focus. It goes without saying that these rules, which do not differ much from the Russian tonicity rules, can be easily explained to the Russian learners of English as a foreign language (though the relation between nucleus placement and being in focus, as many linguists admit, is a complex affair). The tendency for main accent to be located on the so called lexical but not functional words is similar to the Russian tonicity rule and is not expected to present any difficulties to the Russian learners of English as well.

However, being aware of the simple rules mentioned it still can be difficult for a foreign language learner of English to predict accentual structure of an utterance quickly and easily . A number of authors [2, 11] claim that there are cases where an early nuclear accent may signal a broad focus and they are exceptions from the Last

Lexical Item Rule. Some of these patterns include:

The so called event sentences, which typically involve an intransitive verb which denotes appearance/disappearance of something or misfortune. The subject usually receives the nucleus though the whole sentence is in broad focus:

a. What happened?

- The MAIL arrived. The TRAIN'S late.

1. attributive constructions of different types.

a. Noun+ infinitive:

I've got a JOB to complete

b. Noun+ adjective or participle:

He left the GATE open.

2. Wh-questions with final verb. In this case nucleus is mostly located on a noun as well:

a. How many BOOKS have you brought?

b. What DRESS was she wearing?

3. Utterances with final adverbials of time :

a. I went to the ZOO yesterday.

b. Did you see the FILM last night?

4. Sentences with final items of low semantic weight, including final objects, when their reference is of general nature:

a. Let's DO something.

b. I'll ASK somebody.

In this case nuclear accent is normally on the verb.

Examples belonging to these groups can definitely present difficulties for foreign learners of English, as they demonstrate that the lack of nuclear accent does not necessarily refer the word to old information.

Russian speakers perform better in those areas where the nuclear accent patterns are similar in English and Russian ,for instance, when de-accenting old information is observed in utterances where old information is the matter of repeated words, as in: When do you think he will come? – I don't think he CAN come. They also don't experience much difficulty in expressing contrastive focus. In these cases the Russian speakers most probably enjoy positive transfer from their native language and place the nucleus on the correct place even though sometimes the word order in the corresponding Russian sentences may be different. At the same time one can predict various degrees of difficulty in those areas where the two languages differ. When analyzing students' performance 2 types of errors emerge - errors which may be explained by negative transfer from their native language and errors which do not follow either the English or the Russian tonicity rules. The former .manifest themselves in the English speech of the Russians in general questions as their accent patterns in the two languages differ. If in the English general questions: Do you like COFfee? Does she speak ENGlish? it is the last word which is accented, in the corresponding Russian translations the nucleus will be placed on the verb. So, Russian speakers most often chose verbs as nuclei in the English general questions. Another problematic area for the Russian learners is expected to include utterances with nuclear accents on auxiliary verbs and prepositions. Being a synthetic language Russian does not use grammatical words or particles to express syntactic relations within a sentence. There is no such a category as an auxiliary verb in the Russian language at all. Anyway, the teaching materials used in class provide a lot of examples of utterances where auxiliary verbs are accented, like:

a. He didn't make it that way.

– Which way DID he make it then?

b. Do you like it?

– I DO love it.

In the speech of many Russian learners of English the old information in these cases is re-accented:

Do you like it?

- I do LOVE it.

Speaking about English prepositions there can be two types of sentences where the nucleus is located on a preposition. Both involve special questions in which there are no content words:

Who is she WITH?

What's it ABOUT?

For the Russian learner of English the final position of the preposition looks very unusual as the corresponding Russian phrases will start with the prepositions: With who is she? About what is it? Choosing between a preposition and a pronoun, Russian speakers prefer placing nuclear accent on a pronoun which has to do with a wrong realization of neutral tonicity.

The second case involves prepositions immediately following a question word:

a. I've scored 60.

- What OF it?

b. You know my essay?

- What ABOUT it?

The Russian subjects often accent a final pronoun or a question word, avoiding prepositions acting as nuclei.

The methodology which can be used when carrying out this kind of comparative research is quite well known and described by A. Cruttenden [3]. The author conducted a cross-linguistic research with the Dutch speakers of English. Dutch has tonicity system extremely similar to that of English, so as the experiment proved neither speakers of English nor speakers of Dutch

experienced any problems in assigning sentence accents when speaking each other languages.

However, as it has already been mentioned, some languages differ in tonicity rules significantly and can be an interesting area for a comparative study of this type.The analyses of linguistic literature demonstrates that speakers of Romance languages have always had trouble in acquisition of tonicity aspect of English phonology. There hasn't been much research on tonicity errors by Spanish speakers of English though there are studies on focus and prosodic prominence in the speech of Spanish learners of English [13, 5, 10, 6]. A review of the literature shows that tonicity contrasts between the two languages – Spanish and English are much greater than similarities. First, in Spanish nuclear stress is located on the last word of the intonation phrase in broad focus context and as it has been shown above there are quite a few exceptions to the Last Lexical Item Rule in English. Moreover, English is very sensitive to de-accenting given information. The rule in Spanish is that the nucleus falls on the last word/syllable of the tone unit, no matter given or new. Several authors claim that also the shifting of nucleus is also possible in Spanish a more common procedure for highlighting words is carried out by lexical and syntactic means, mostly word order variation. As for expressing contrastive focus some linguists admit that in this case Spanish allows to use stress shift to signal contrastive focus, but not for other types of narrow focus [8].

There is no doubt that further exploration of the accent patterns in English utterances in both broad and narrow focus in the speech of Spanish speakers of English deserves linguists attention.

The comparative investigation of this type can help answer several questions:

1. Do the contrasts/similarities between Spanish and English tonicity have an influence on the second learners' accentuation strategies?
2. Which areas constitute a problem area/s for the Spanish learners of English?
3. Do they correspond to the problem tonicity zones revealed in the speech of Russian learners of English?
4. How serious are these types of errors and what is the reaction of the native speakers?
5. Can they cause as J.Jenkins claims a breakdown in communication?
6. Can this type of errors be corrected easily?
7. If yes, how?

Having subjects of different languages and different English language proficiency levels may help explore how Russian and Spanish learners of English progress towards the native like production in the field of tonicity.

REFERENCES

[1] Banjo, A. 1972. Beyond Intelligibility in Nigerian English. In: *Varieties and Functions of English in Nigeria*. Irban: African Universities Press, 7-13.

[2] Cruttenden, A. 1986 Intonation. Cambridge University Press.

[3] Cruttenden, A. 1993. The de-accenting and reaccenting of repeated lexical items. In *Proceedings of the ESCA workshop on Prosody, Lund*, 16—19.

[4] Dalton C., & Seidlhofer B. 1994. *Pronunciation*. Oxford: Oxford University Press.

[5] Diez, F. G. 2005. Intonation Focus in the Interlanguage of a Group of Spanish Learners of English. *Revista Alicantina de Estudios Ingleses,* 18, 129-147.

[6] Estebas-Vilaplana E. 2001. Teaching Accentual Focus to Catalan Students of English. Paper presented at Phonetics *Teaching and Learning Conference*. Engham, England. [cited 20 November 2004] Available from http://www.phon.ucl.ac.uk/home/johnm/ptlc2001/pdf/estebas.pdf.

[7] Halliday, M. A. K. 1967. *Intonation and Grammar of British English*. The Hague: Mouton.

[8] Hoot B. *Dissertation Prospectus*. http://www.uic.edu/depts/ing/BRL/BRL_research_Dissertation%20Prospectus%-%20Brad%20Hoot.pdf

[9] Jenkins, J. 2000. *The phonology of English as an international language: New models, new norms, new goals*. Oxford: Oxford University Press.

[10] Jurado, M. E. 2005. Revision between differences between Nucleus Placement in English and Spanish. *Proc. Of the 4th International Conference on contrastive Linguistics*, Compostela, 603 – 611.

[11] Ladd, D. R. 1980. *The Structure of Intonational Meaning*. Indiana University Press, Bloomington, Ind.

[12] Oldin, T. 1989. *Language Transfer: Cross-linguistic influence in language learning*. Cambridge University Press.

[13] Ortiz-Lira, H. 1994. *A contrastive analysis of English and Spanish sentence accentuation*. Thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the faculty of Arts. Unpublished Ph.D. Thesis.

[14] Wells, J. 2006. *English Intonation: an Introduction*. Cambridge University Press.

LENGTH AND DURATION OF EVENKI VOWELS

Olga Morozova

The Amur State University, Russia
morozova_olga06@mail.ru

ABSTRACT

The paper presents the quantitive characteristics (i.e. length) of Evenki vowels. The data examined were high quality sound recordings of speech of four native speakers of the Selemdza accent of the Evenki language.

Keywords: phonetics; vowels; the Evenki language.

1. INTRODUCTION

Language is the heart of culture. Language holds people together and establishes their identity as a group. But nothing is more ephemeral than the sounds of a language. In case a language is disappearing, in other words, endangered, the sounds of this language will live only as long as the language is spoken [1]

Nowadays the Evenki language is considered a severely endangered one. It is registered in the UNESKO Red Book of Endangered Languages. The Amur Evenki varieties are spoken by grandparents and older generation. The parent generation may understand it, but they do not speak it to children or among themselves.

Evenks are the only Tungus speaking people living in the Amur Region (Russia). According to the Census (2009) The Amur Evenks (about 1500 people) live collectively in 5 national villages in Tyndinsky, Zeysky and Selemdzhynsky areas. The Selemdzha Evenki accent is in the worst condition. The territory, where the Selemdzha type is spoken, is extended for no more than 46 thousand square kilometers. About 30 Evenks lead a nomad's life on the territory of Mazanovsky Area. Young Amur Evenks have no possibility to listen to the native oral Evenki speech. This problem makes it particular useful to record the unique Amur types of the Evenki language.

There are very few studies language of Amur Evenks. There is no consistant description of phonetic patterns of Amur Evenki varieties. One may hear old fluent speakers of Amur Evenki regional types only in taiga that is far away from populated areas. Teachers of the Evenki Language say that schoolchildren learn it as any other foreign language. Children use their native language only in classrooms and do not speak it anywhere. They need not do it, because everybody speaks and understands Russian [2, 99-100].

2. MATERIAL AND METHOD

The speech material, used for researching length for Selemdzha Evenki vowels is isolated words pronounced by four native speakers (one male and three females) of the Selemdzha variety of the Evenki language (see Table 1).

Table 1: The speech material for researching length for Selemdzha Evenki vowels (transcription on the basis of International Phonetic Alphabet).

№	Selemdzha Evenki*	Standard Evenki**	English
1	**mukti: silukta**	i:	blind gut
2	u:	u:	scraper
3	dʲu:	dʲu:	tent (of skins or bark)
4	**ajmakin**	da:	relative
5	do:n	do:	internal organ
6	lu:	lu:	resin
7	mo:	mo:	tree
8	mu:	mu:	water
9	dʲa:n	dʲa:n	ten
10	**sokso**	la:n	trap
11	bo:na	bo:na	hail
12	**e:ɣa**	e:sa	eye
13	tʃa:ski	tʃa:ski	farther
14	**ko:to:**	pu:rta	knife
15	**sʲa:kso:**	sɛº:ksɛ	blood
16	**a:ʧaβ**	a:răn	had a nap
17	**sa:dʲerɛn**	sa:răn	knows
18	**bu:ʧa**	bu:rɐn	gave
19	e:kun	e:kun	what
20	**ariʧa**	ɛº:rim	called
21	**dʲo:ʧa**	dʲo:nan	remembered
22	**nʲo:ʧaʧa**	mo:rʧan	fight
23	a:ʧin	a:ʧin	no
24	**kure:**	βo:tar	fence
25	**i:ʧa**	i:-	to come
26	**bu:ʧa**	bu:-	to give

27		da:-	to notice
№	Selemdzha Evenki*	Standard Evenki**	English
28	**do:ʧa**	do:-	alight
29	**ju:ʧa**	ju:-	go out
30	**na:ʧa**	nɛ:-	to put
31	**ha:βʧa**	ha:β-	to add
32	-	ta:n-	to draw out
33	e:lla	e:lla	coal
34	ge:	ge:	the second
35	se:n	se:n	ear
36	**hagdɨ ir^{j}akta**	he:ki	a big larch
37	**hēмуργи мӯ**	he:mur	cold
38	**ʧu:linka:n ʧinɨrɨka:n**	he:n^{j}aki:	mole
39	urɛ^{o}:	urɛ^{o}:	mountain
40	adu:	adu:	how many
41	tadu:	tadu:	there
42	d^{j}uγa	d^{j}uγa	in summer
43	-	korme:	flap
44	ʧutʃu:n	ʧutʃu:n	scraper
45	**ɛri:ʧa**	ɛri:m	breathed
46	d^{j}agda:	d^{j}agda:γ	pine forest
47	**kuli:n**	u:rɛ:	worn
48	**a:mi:d^{j}aran**	a:mi:	sleeping
49	-	d^{j}u:du:n	in the tent
50	**mo:**	mo:ka:n	stick
51	i:rinak	i:rinɛk	anthill
52	-	tu:kala	locust
53	**du:nna**	tu:kala	soil
54	**ʧa:lbon**	o:sikta	star
55	**aγi:kpi**	asi:nan	with wife
56	**oγo:**	ama:ka:	bear
57	**hi:ma:t**	ama:ka:n	soon
58	**nunan oronɛn**	oronmo:n	his deer
59	**tɛma:tna**	tɛγɛmi:	tomorrow
60	kɛdɛra	kɛdɛrɛ	a tool for leather dressing
61	**toktʃoko:**	ke:ta:ra:	one-eyed
62	ɛmɛd^{j}ɛŋa:n	ɛmɛd^{j}ɛŋɛ:n	will come
63	**ʃurud^{j}iŋa:n**	dukud^{j}aŋa:n	will write
64	**ʃuruʧo: bojuʧa:**	mo:tima:răn	went hunting
65	**ho:gdɛ aγi:**	asika:kun	a big woman
66	**kirakan ir^{j}a:ktɛtka:n**	irɛ:ktɛkɛ:n	a small larch
67	**ilan^{j}**	ilani:	the third
№	Selemdzha Evenki*	Standard Evenki**	English
68	amargu:	amargu:	hind
69	horoki:	horoki:	cock of the wood
70	**lu:ma:m**	lu:la:	to go for resin
71	**bojuʧa**	motɨma:	to go moose hunting
72	-	ilanma:n	triple
73	-	va:ma:n	passionate hunter
74	**bakaldɨmi**	bakaldɨna:	to go meeting
75	**hol**	ilaŋa:t	fuel
76	-	taγinka:	riddle
77	**noŋan ŋa:loβa:n**	na:lɛβa:n	his hand
78	**noŋan ull^{j}o:n**	ullɛβɛ:n	his meat
79	**noŋan kokoll^{j}o:n**	kokolloβo:n	his mitten
80	aja	aja	good
81	**ajami:**	ajavd^{j}ă	to love
82	haβa:	haβa:	work
83	**haβald^{j}aran**	haβald^{j}a	to work
84	**aγi:**	asi:	woman
85	**hogdɨŋo: aγi:**	asika:kun	a big woman
86	ʤagda	d^{j}agda	pine
87	**ira:ktɛ**	irɛ:ktɛ	larch
88	**ira:k**	irɛ:ktɛ:γ	larch forest
89	**ʧuβka**	ʧuka	grass
90	-	ʧukaγ	meadow

*Selemdza words are given in acoustic transcription
** Poligus words are given in phonemic transcription

As shown in Table 1, one of the Evenki vowels phonological features is length.

Isolated words for experiment are ones of Standard Evenki and taken from [9]. But during the experiment it turned out that the difference between Standard and Poligus Evenki accents is so great that 70 % Selemdzha words (57 of 90 items) do not coincide with the main experimental bulk. This fact made us suggest the other list of words to our speakers. The list included minimal pairs that illustrate length of Evenki vowels: **имӣ** */imi:/ to dress a fell* - **ӣмӣ** */i:mi:/ to come;* **икэн** /ikɛn/ *chin* - **икэ̄н** /ikɛ:n/ *a song;* **амӣ** */ami:/ father* - **а̄мӣ** */a:mi:/ to sleep;* **омӣ** */omi:/ sole* - **о̄мӣ** */o:mi:/ to do;* **бӯмӣ** */bu:mi:/ to sleep* - **бумӣ** */bumi:/ to die/.*

The data presented in this paper result from perceptual and acoustic study of Selemdzha Evenki vowels, segmented from monosyllabic words.

The speech material was recorded in the Laboratory of Experimental Phonetics in the Amur State University (Blagoveshchensk, the Amur Region, Russia). The material was transcribed using the

symbols of the International Phonetic Alphabet. In accordance with international practice, phonemes and phonemically written words in this paper are given between slanting lines, allophones of phonemes in square brackets. PRAAT was used for acoustic measurements.

3. RESULTS AND DISCUSSION

Length is a distinctive feature in the vowel system of the Evenki language. There is no symmetry between long and short vowels: the Evenki language has 6 long vowels /i:, e:, a:, ɛ:, u:, o:/ and 5 short vowels /i, a, ɛ, u, o/ [3, 4, 5, 6, 7, 8, 9].

The following minimal pairs illustrate phonological length of Evenki vowels:

/i/ - /i:/ **илэ** /ilɛ/ *a man* - **ӣлэ̄** /i:lɛ:/ *where;*

/ɛ/ - /ɛ:/ **турэн** /turɛn/ *language* - **тӯрэн** /tu:rɛn/ *he trod;*

/a/ - /a:/ **няма** /nʲama/ *it's warm* - **нямā** /nʲama:/ *a hundred;*

/o/ - /o:/ **оран** /oran/ *rapid s*- **ōран** /o:ran/ *he did;*

/u/ - /u:/ **бумӣ** /bumi:/ *to die* - **бӯмӣ** /bu:mi:/ *to sleep* [7].

Length of an Evenki vowel is not connected with its position in the word. Long vowels can be vowels of any syllable, i.e. at the beginning, in the middle or at the end of a word, e.g., /bo:ktan/ *tied*, /duku:n/ *a letter.*

The length of an Poligus Evenki vowel varies depending on a syllabic structure of a word, number of syllables, an intrinsic duration of a definite vowel, context, stress. The longest vowels are those in monosyllabic words [9].

In the Tommot Evenki variety both short vowels and long vowels in checked position are shorter than in free position. In other words any Evenki vowel is the longest in V-, CV- syllables and the shortest in CVC- syllable [3].

Vowels duration measurements carried out by other researchers enable to say that it depends on (i) the vowel quality, (ii) syllabic structure of a word and (iii) whether there are other long vowels in a polysyllabic word. In disyllabic words 69 % of all vowels are long in the first syllable and only 31 % of all vowels are long in the second one. The distribution of long vowels in polysyllabic words is different. 91 % of all long vowels of polysyllabic words are located in the last third syllable [4, 296-297].

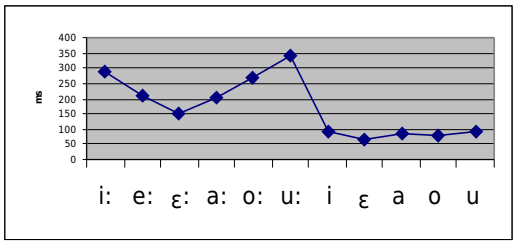
As for Selemdzha Evenki variety there is no information concerning vowel quality and duration.

Judging by the results of the acoustic experiment carried out the current study, in the Selemdzha Evenki variety there is a distinction between long and short vowels (see table 2 and fig.1). In general mean long vowel duration is 244 ms, mean short vowel duration is 83 ms. In other words, long Selemdzha Evenki vowels are three times longer than short ones.

Table 2: Mean duration of vowels (ms).

Vowels	D1	D2	D3	D4
i:	105	455	190	161
e:	134	318 [ʲe]	262	124
ɛ:	109	177	195	115
a:	135	223	288	165
o:	173	334	302	263
u:	240	499	355	283
i	60	104	91	104
ɛ	69	65	72	66
a	67	88	96	88
o	65	86	67	87
u	58	97	129	97

Figure 1. Mean duration of vowels (on the whole experiment).



It is a common knowledge that vowels (in similar contexts) usually have some inherent properties as far as duration and intensity are concerned. General rules are the following: the more open and the lower a vowel is the higher intensity and the longer duration it possesses. Usually a-like sounds are the longest and most intensive. I-like and u-like sounds have the smallest duration and intensity. Differences mentioned above are relatively slight and are greatly influenced by prosodic factors such as speech tempo, pitch pattern, rhythm, etc. [10]

General rules about inherent vowel duration do not work for Evenki vowels. Our experimental material indicates that the longest vowels in the Selemdzha Evenki are the closest ones. (see fig. 2).

It should be noted that the tendency to be the longest is true for long close vowels /i:/ and /u:/.

In the subsystem of short Evenki vowels differences in inherent length are not so large. Length of

all short Evenki vowels is approximately the same and ranges from 68 ms up to 95 ms. But again the short Evenki vowels /i/ and /u/ are longer than the other ones. However it cannot be stated that duration differences for /i-a-u/ are essential. The variation is within 5-10 ms. The pair of short mid Evenki vowels /ɛ-o/ can be characterized as the shortest ones, 68 and 76 ms correspondingly (see fig. 3).

Figure 2. Mean duration of long vowels (ms).

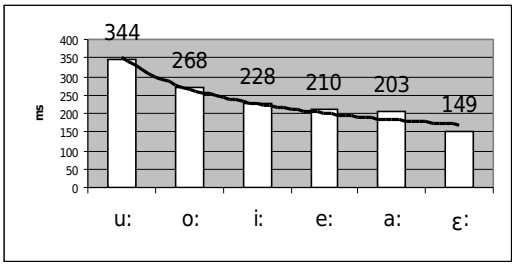
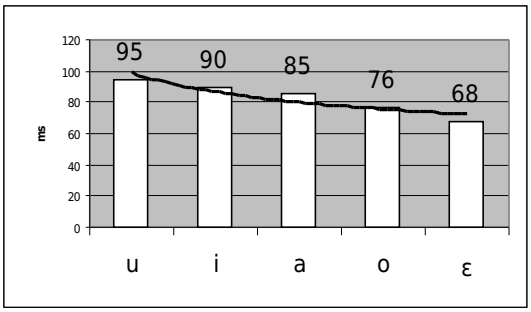
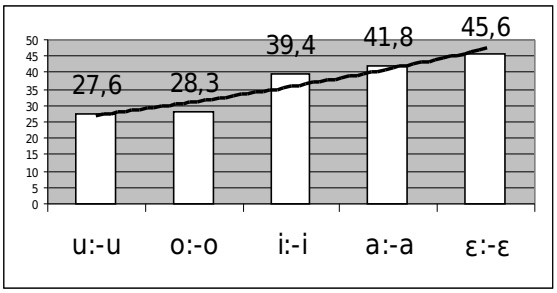


Figure 3. Mean duration of short vowels (ms).



Besides it should be mentioned about one thing that is true for both long and short Evenki vowels. Analysing the opposition front-back vowels one can see that the former are shorter but the latter are longer. In the subgroup of long Evenki vowels mid front vowel /ɛ:/ is the shortest one (149 ms) and close back /u:/ is the longest one (344 ms). In the subgroup of short Evenki vowels one can see the same rule. Mid close /ɛ/ is the shortest (68 ms) and close back /u/ is the longest (95 ms).

Figure 4. Correlation between long and short vowels (%).



If we consider the correlation between long and short vowels the results will be following. The duration difference between long and short back vowels /u:-u/, /o:-o/ is less than between long and short front /i:-i/ and central vowels /a:-a/, /ɛ:-ɛ/ (see fig. 4).

4. CONCLUSION

The current study demonstrated rather peculiar relations between Evenki vowels length and their quality that need to be studied further.

5. ACKNOWLEDGMENTS

For useful feedback the author would like to thank professors Kim Yuwon, Ko Dongho, T. Andreeva. I am also immensely grateful for the time and expertise contributed by our Selemdzha Evenki Language consultants.

6. REFERENCES

[1] Ladefoged, P. Preserving the sounds of dissapearing languages */www. linguistics.ucla.edu/people/ladefoge /Preserving%20sounds.pdf*

[2] Morozova, O. 2011. Duration of Evenki Vowels (on experimental data). *Proc. 10th Seoul International Altaistic Conference "Reexaminations of Objects and Methods of Research into Altaic Languges and Cultures"*. Sunchon National University. The Altaic Society of Korea, 99-104.

[3] Andreeva, T. 1988. *Zvukovoy stroy tommotskogo govora evenkiyskogo yazyka (The sounds of the Tommot variety of the Evenki Language)*. Novosibirsk, "Nauka".

[4] Andreeva, T. 2006. Phonetic Parallels in Altaic Languages. *Proc. 7th Seoul International Altaistic Conference "Perspectives on the Studies of Endangered Manchu-Tungus"*. Seoul National University. The Altaic Society of Korea. 291-304.

[5] Afanasyeva, E. F. 2007. Evenkiyskiy Yazyk v Buryatiyi / *Almanach 2007 "Yazyki narodov Rossiyi"*. St. Petersburg, St. Petersburg University, 17-20.

[6] Boldyrev, B. 2000. *Evenki-Russian Dictionary.* Part 2. Novosibirsk. Siberian Branch of the Russian Academy of Science. "Geo".

[7] Bulatova, N. Y. a. 2003. *The Evenki Language in Tables (Evenkiyskiy yazyk v tablitsah)*/ St. Petersburg. "Drofa".

[8] Bulatova, N. Y. b. 1999. *The Language of the Sakhalin Evenki.* St. Petersburg, St.Petersburg Fund "North without Conflicts".

[9] Konstantinova, O. 1964. *Evenkiysky yazyk (The Evenki Language)*. Moscow-Leningrad, "Nauka".

[10] Kodzasov S., Krivnova O. 2001. *Obshchaya phonetika (General Phonetics).* Moscow. Russian State Humanitarian University.

ACOUSTIC CORRELATES OF BRITISH CONSONANTS IN POLITICAL DISCOURSE

Elena Musaeva

Amur State University, Russia
blagmeg@yandex.ru

ABSTRACT

The article gives a review of consonants' acoustic features in British political discourse. The groups of plosives, fricatives, nasals and approximants (liquids and glides) were involved into the analysis. The special attention was paid to such characteristics as duration, intensity and noise components of the consonants under study. The analysis was conducted on the basis of political speeches provided by native speakers of British English. The results illustrate some difference between the generally accepted acoustic features which can be found in spontaneous British speech and those realized in British political discourse.

Keywords: British English, consonants, acoustic properties, duration, noise components, spectrogram.

1. INTRODUCTION

The acoustic analysis undertaken is to find out peculiar acoustic properties of British consonants in the political discourse.

The type of discourse under study is attributed to a prepared form of speech, usually text-supported and is characterized by distinct articulation and clear rhythmic structure. The speech of politicians, as the most educated representatives of a nation, has always viewed as phonetically correct.

All these promote the idea of a sui generis "standard" of sounds' realization in the political discourse.

2. MATERIAL AND METHODS

2.1. Corpus

The material employed in the course of analysis has been obtained from the official website of British government (www.pm.gov.uk) in *mp3* format and then converted into *wav* format to facilitate further analysis of speech sounds.

The speakers are British Prime Ministers: Tony Blair, 64 years old; Gordon Brown, 66 years old; David Cameron, 44 years old. The speeches of the three Prime Ministers follow the pronouncing standard of British English, cover political topics and are considered to be quasi-spontaneous. The total duration of audio speeches is 1.45 hour.

2.2. Method

The acoustic properties of English consonants can be visualized and measured with the help of a spectrogram. Yet, there is no single acoustic measurement that can be applied to distinguish all consonants in English. Due to varying manners of articulation that underlie the different consonants of English, it is always useful to group them accordingly for an acoustic analysis [1].

So, in the present study the consonants were grouped as follows:plosives /p/, /b/, /t/, /d/, /k/, /g/, fricatives /f/, /θ/, /s/, /ʃ/, /h/, approximants /w/, /l/, /r/, /j/ and nasals /m/, /n/, /ŋ/.

The acoustic properties of the above mentioned consonants were analyzed in a stressed syllable (CV type) as it is commonly known that preposition to a vowel in a stressed syllable facilitates consonants to reveal their properties to the best.

2.2.1. Acoustic properties of plosives in British political discourse

The spectrum of English plosives is characterized by an apparent weakening of the signal or its total absence. This acoustic interval corresponds to the articulatory closure and has various duration depending on the prosodic conditions. It is followed by a short burst of noise which indicates the release of the closure [2, 3, 4].

The following acoustic features were taken into consideration in the study of plosives: duration of the closure, presence or absence of F0, intensity of burst stage, typical features of fricative and aspirate noise (their intensity and duration).

In the course of study employing British political speech we managed to reveal the following characteristics of plosives:

1. The duration of the closure part is longer for the voiceless consonants /p, t, k/ in comparison with the voiced ones /b, d, g/. For example, closure duration for voiceless bilabial /p/ varies from 90 up to 140 ms, equaling 110 ms on average. The same measurements for its voiced correlate vary

from 65 up to 90 ms, the average of which is 75 ms (See Fig.1 and Fig.2 for comparison).

Figure 1: The spectrogram of the word “pension”.

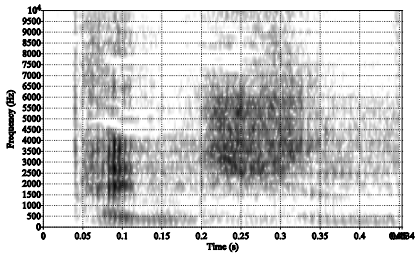
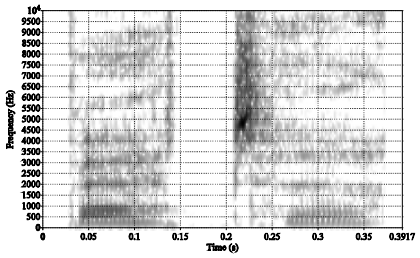


Figure 2: The spectrogram of the word “better”.



2. The burst stage is indicated by the presence of pulse which is seen as a vertical stretch (spike) of energy. It is usually more intensive in case of voiceless consonants /p, t, k/, of which /t/ tends to be the most intense (See Fig. 3 and Fig. 4 for comparison).

Figure 3: The spectrogram of the word “tax”.

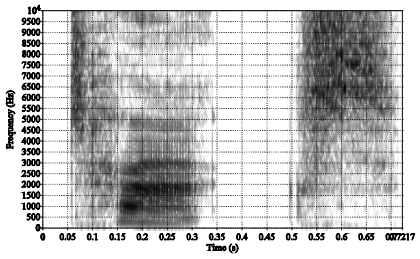
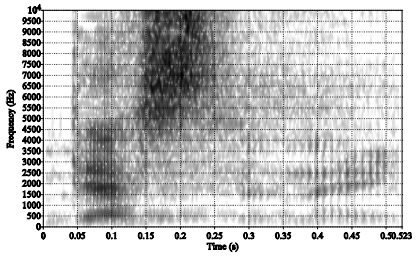


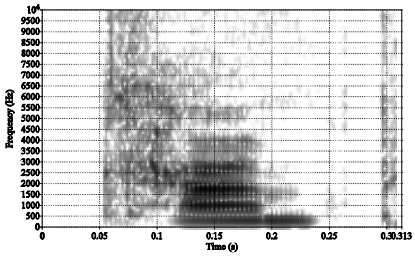
Figure 4: The spectrogram of the word “destiny”.



The analysis of the spectrograms for this sound shows that in the majority of cases the burst stage has a wide range and takes lower and middle areas of the spectra (from 1500 to 3000 Hz).

3. Aspiration is noticeable in the spectra of voiceless consonants /p, t, k/ in the preposition to a vowel in a stressed syllable. The duration of aspiration equals ascending from 30 ms for the bilabial /p/ and up to 50-60 ms for the velar /k/ (See Fig. 5).

Figure 5: The spectrogram of the word “camp”.



2.2.2. Acoustic properties of fricatives in British political discourse.

The general acoustic property characteristic of fricatives is noise that occurs through a narrow constriction in the vocal tract. The constriction causes turbulence in the airstream, which is associated with noise in the acoustic signal. This noise is much more intense in the sibilant fricatives, such as /s, z, ʃ, ʒ/. The voiced fricatives /f, v, θ, ð/ and the voiceless glottal fricative /h/ only have a very faint formant structure with little energy [5, 6].

High-intensity noise and energy concentration in relatively high frequencies are typical for the spectrograms of hole fricatives /s/ and /z/. While slit fricatives /θ/ and /ð/ are of lower intensity and are characterized by a more homogeneous distribution of energy within the spectra, noise components being of a more diffuse nature (See Fig.6 and Fig.7).

Among the acoustic properties of interdental fricative /θ/ is a vast, low-intensity noise randomly distributing from about 1000 Hz onwards (Fig. 7). Its mean duration in a stressed word initial syllable equals 119 ms.

The spectra of the labiodental fricative /f/ is marked by a low-intensity homogeneous noise at the spectral range of 800 – 10000 Hz. Its mean duration in a stressed word initial syllable equals 122 ms (Fig. 8).

The spectrogram of sibilant /ʃ/ is also marked by a homogeneous noise being the most intense at the

range of 2000 – 6000 Hz. Its mean duration in a stressed word initial syllable equals 118 ms (Fig. 9).

When comparing the three fricatives in the words "fit", "sip" and "shiver" we can see a significant diversity of energy concentration – in case of hole fricatives /s/ and /ʃ/ noise intensification is observed. There is also a marked difference in frequency data – the alveolar fricative /s/ displays the most intense noise component in the range of 4000 – 9000 Hz, while the palate-alveolar fricative /ʃ/ displays the same features at the range of 2000-7000 Hz of the spectrogram.

Figure 6: The spectrogram of the word "sip".

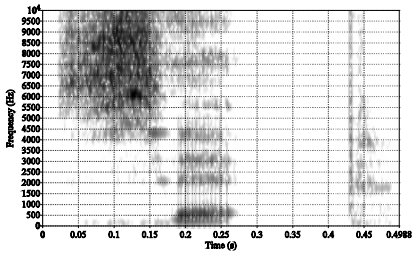


Figure 7: The spectrogram of the word "think".

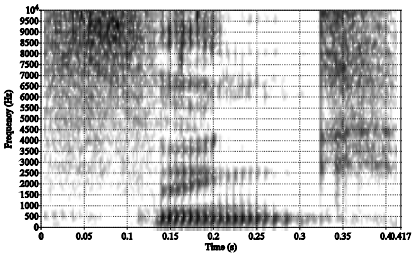


Figure 8: The spectrogram of the word "fit".

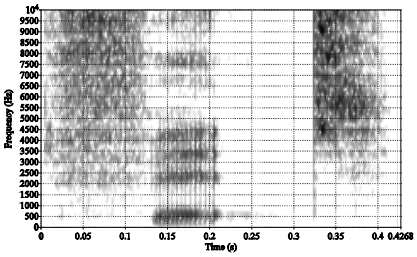
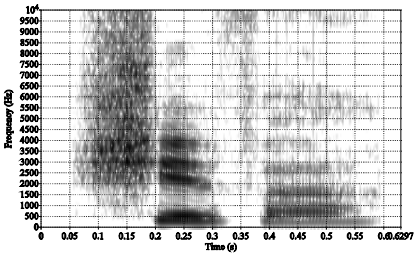
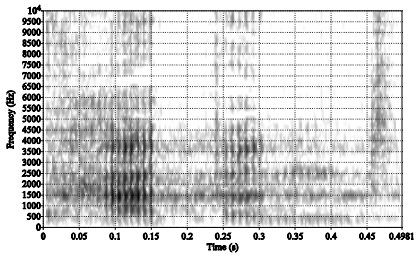


Figure 9: The spectrogram of the word "shiver".



Getting down to the glottal fricative /h/ it's worth mentioning that there is no distinctive voiceless/voiced opposition such as characterizes the other fricatives. The glottal fricative /h/ functions essentially as a voiceless syllable-initial phoneme and easily adopts to the formant movements of the vowel posterior. Physically this sound presents aspiration which in the spectrogram looks rather diffuse (Fig. 10). The signal is relatively weak and has a tendency to vocalization medially between voiced sounds, e. g. in words such as *ahead, behind.* The duration of the glottal fricative in a stressed word initial syllable varies in the range of 40 – 80 ms.

Figure 10: The spectrogram of the word "happened".



2.2.3. Acoustic properties of nasals in British political discourse.

As widely known there are three nasal phonemes in English which differ in the place of articulation: bilabial /m/, alveolar /n/ and velar /ŋ/. Nasals are frequently compared to vowels or approximants as they can be characterized largely in terms of their formant frequencies, but they differ in that the formants are not as marked as they are in vowels [7].

Figures 11, 12, 13 show the spectrograms of the three nasals. Sounds /m/ and /n/ are represented in intervocalic position and /ŋ/ in syllable-final as it is the only position the latter can occur. All the three show slowly varying waveforms – a characteristic feature of signals with little energy in the high frequency regions. Just after the vowels preceding the nasal there is a sharp discontinuity (marked by an arrow) which identifies the lips or tongue movements. After this point less amplitude in the nasals is left. The spectrograms display that the nasals have a prominent low frequency F1, referred to as a nasal formant (the dark area at the low frequency range of the spectrogram – usually 200-500 Hz). For each of them there is also another formant visible in the neighborhood of 2500 Hz, but there is comparatively little energy in the region normally occupied by the second formant (at the range of 1600-1650 Hz).

The duration measurements for the three nasals are the following (min-max):

1) bilabial nasal /m/: 70-115 ms;
2) alveolar nasal /n/: 50-94 ms;
3) velar nasal /ŋ/: 72-121 ms.

Figure 11: The spectrogram of the word “diplomatic”.

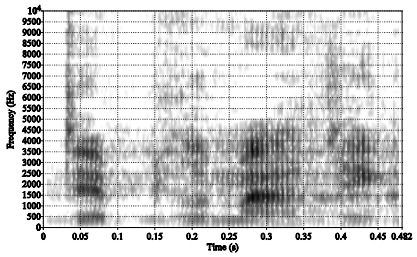


Figure 12: The spectrogram of the word “financial”.

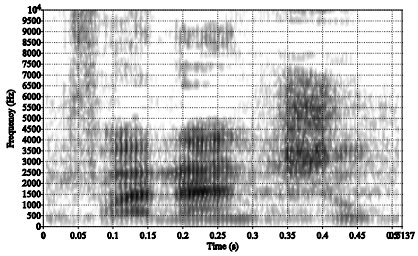
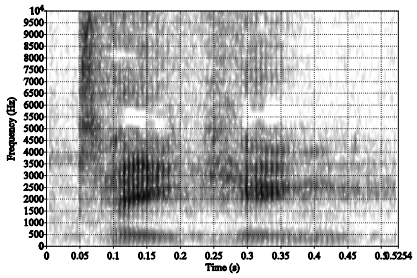


Figure 13: The spectrogram of the word “taking”.

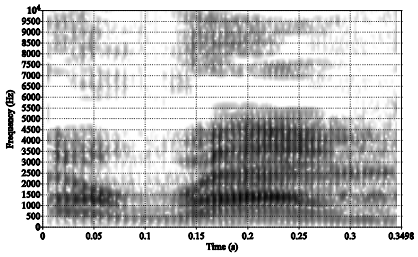


2.2.4. Acoustic properties of approximants in British political discourse.

The liquids /r/ and /l/ and the glides /w/ and /j/ have vowel-like waveforms and are identifiable by their formant configuration. Since high constriction is not characteristic to the liquids and glides, the formants are more continuous at the vowel junctures.

The spectrogram in Figure 14 shows the liquid /r/ in an intervocalic context. The formant configuration of /r/ has low values for all the three formants. This sound is easy to identify as it has a low F3 (below 2000 Hz). At the spectrogram the liquid is notable for the formants’ onward movement when a posterior vowel occurs.

Figure 14: The spectrogram of the word “around”.



The spectrogram of the lateral liquid /l/ shown in Figure 15 is characterized by faint formant bars – F1 at a very low frequency (around 500 Hz) and F2 at the range of 900-1600 Hz, a value at the low end indicating a dark /ɬ/. The difference between the “clear” and the “dark” /l/ can be seen through the specific features of F1 and F2. In clear /l/, F1 and F2 are farther apart, and closer and lower in dark /l/ [8, 9].

The group of glides is represented by 2 phonemes: labial-velar /w/ and unrounded palatal /j/ which are viewed as vocalic glides possessing acoustic features similar to those of vowels, i.e. a characteristic two- or three-formant structure similar to that of /iː/ /uː/. In /w/, F1 and F2 are very close, while in /j/ are wide apart [10].

It’s worth mentioning that the labial-velar glide /w/ has a low position of all formants with a sharp rise in F2 (Figure 16). The palatal glide /j/ has a low F1 and a high F2 which move towards each other (Figure 17). The starting point of F2 is about 2300-3500 Hz for the palatal glide /j/ and within the range of 350-850 Hz for the labial-velar glide /w/.

Figure 15: The spectrogram of the word “large”.

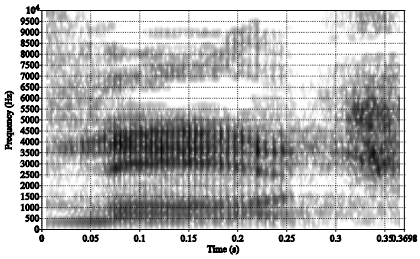


Figure 16: The spectrogram of the word wick”.

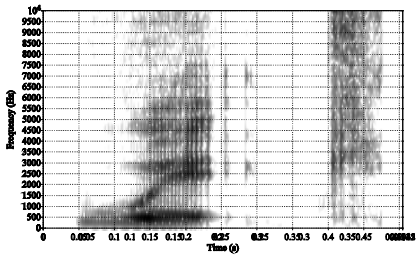
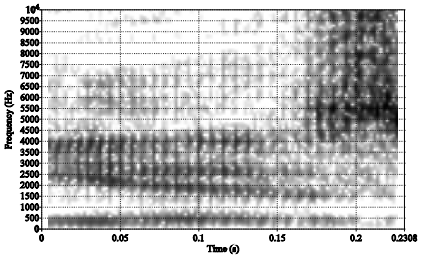


Figure 17: The spectrogram of the word “years”.



3. RESULTS

The acoustic data obtained in the course of study of consonants in British political discourse generally correspond to those presented in original sources on acoustic phonetics. The results support the idea of a standard-form pronunciation of British politicians and give the sui generis standard for the acoustic properties of groups of British consonants which is undoubtedly significant for further research of consonants’ realization in the prepared form of discourse.

4. REFERENCES

[1] Gut, U. 2009. *Introduction to English Phonetics and Phonology.* Frankfurt am Main: Peter Lang, GmbH.

[2] Halle M., Hughes G. W., Radley J. P. 1967. Acoustic Properties of Stop Consonants. In Lehiste I. *Readings in Acoustic Phonetics*. Cambridge, MA: MIT Press, 170-179.

[3] Harrington, J. 2010. Acoustic phonetics. In J. Laver & W. Hardcastle (Eds.). *The Handbook of Phonetic Sciences*. Blackwell, 81-129.

[4] Peterson, H. On British English p-, b-, sp- and -s+b- [Text] // Annual Report of the Institute of Phonetics University of Copenhagen. Copenhagen, 1972. - PP. 269-304.

[5] Stevens, P. 1967. Spectra of fricative noise in human speech. In I. Lehiste. *Readings in Acoustic Phonetics*. Cambridge, MA: MIT Press, 202-219.

[6] Heinz J. M., Stevens K. N. 1967. On the Properties of Voiceless Fricative. In I. Lehiste. *Readings in Acoustic Phonetics*. Cambridge, MA: MIT Press, 220-227.

[7] Fujimura, O. 1967. Analysis of Nasal Consonants. In I. Lehiste. *Readings in Acoustic Phonetics*. Cambridge, MA: MIT Press, 238-248.

[8] Cruttenden, A. 2001. *Gimson’s Pronunciation of English.* Beijing: Edward Arnold (Publishers) Limited.

[9] Recasens, D. Darkness in /l/ as a gradual phonetic property http://www.aune.lpl.univ-aix.fr/-labphon10/abstracts/170pdf.

[10] Yavas, M. 2011. *Applied English Phonology*. Chichester, UK: Blackwell Publishing.

THE SOURCES OF GERMAN ORTHOGRAPHY

Vladimir Naumov

St.-Petersburg University of the Humanities, Russia
vladimir_naumov_@mail.ru

ABSTRACT

The history of German Orthography is rather controversial. It took German Orthography around 400 years to be formed. However, today it is considered the simplest and the most effective graphic code among all European languages. Remarkable achievements in German spelling and normalizing and standardizing of oral speech belong to the Church. First translations of the Holly Bible were the first successful attempts to unify the graphic coding of the phonological system of the German National Language. The analysis of those translations demonstrate changes concerning a number of graphemes, which reflect tendencies of speech of the 16th century compared to that of the 14-15th centuries. It was a great attempt to bridge the gap between the language and its graphic reflection. Unfortunately, in the 17th century certain unfavorable exterior conditions widened the gap. The situation continued deteriorating in the following century. The second half of the 18th century was the time of synthesizing traditional historical and phonological principles in German orthography.

Keywords: Graphic code, unification, national language, phonology, norm

The uniqueness of forming the National Literary language in Germany is determined by deep controversy extrinsic and intrinsic factors of the language development of the country. Political, economic, and cultural decentralizing of Germany in 13-16 centuries on the one hand and as a consequence the lack of any unification of the National language phonological system realization and its graphic code on the other lead to total anarchy in means and forms of fixing and transferring information. The main function of the language — its communicative function — was made difficult to perform.

The characteristic feature of German literary language developing is that it grew up and was formed out of graphics where Status quo had been achieved before the premises for the unified sound system were created.

The history of German Orthography is rather complicated and controversial. Forming unified writing code in Germany was not consecutive or progressive. It took German Orthography around 400 years to be formed. The first orthographic dictionaries appeared in Germany only at the end of the 19th century. It is necessary to note however that such a long way of the National System of writing development was not in vain. today it is considered the simplest and the most effective graphic code among all European languages.

Until the middle of the 13th century the official written language on the territory of Germany was Latin. However from the end of the 12th century the first written works of secular literature in the German language characterized by more or less unified orthographic rules appeared.

The other way of forming the basis for the unified national language stems from chancelleries. The State was making its first attempts to take spelling rules under control in order to use it as one of the efficient means for manipulating public consciousness.

However, neither secular literature nor chancellery acts could not aspire for the role of the source of the National system of writing first of all because of their limited influence on the language community. The nation needed the model of the written text that would be maximally close to various layers of society.

Such text had to possess tremendous social value, which in its turn could become the basis for introducing the most acceptable system of graphic coding of the sound form of the language deprived of bright dialectal features. That was a very difficult task for the linguistic situation in Germany at the end of the 16th century.

An important factor in forming German orthography and then normalizing of the phonetic pattern of the language was the Church. It became the most conservative bearer of the official German written speech and the conduit of the newly forming traditions to the common usage. The most valuable in that aspect were the first translations of the Bible which were considered the first attempts of the uni-

form graphic coding of the phonological system of the National language.

High social and cultural value of the text of the Holy Bible prevented the wide range of graphic variation in its translation. Being aware of that value the translator M. Luther, tried to free his variants of translation from dialectal features. However the influence of the Eastern-mid-German dialect which basis was the written speech of the Saxon Chancellery was strong enough in his first translations. Thus in the translation performed in 1580 there were the following dialectal graphic peculiarities: the change of voiceless consonants into the voiced ones in Anlaut and Inlaut as in *Dach – Tag, örde – Örter, Gades – Gottes*, Umlaut of /u/,/o/ vowles, voiceless consonants in word-final position as in *Godt, gutt*. All these features were not inherent in other dialects of the German language like Bavarian-Austrian, Swabian, Eastern-Frank.

In the 16th century Hansa dialect, that was widely spread on the territory of Germany in 14-15 centuries gradually loses its influence. Newly formed phonological oppositions start to be reflected in M. Luther's Bible translation graphics.

The analysis of the Hole Bible translation dated by 1580 demonstrates that graphemes [o], [u] aquire their [ö], [ü] pairs in words like wöste, düster, Düpe as a result of their Umlaut process.

The phenomenon of diphthonging is represented in the translation in a rather contradictory way. The diphthong /ae/ has three ways of spelling: ei, ey, y. Grapheme [y] is used to denote short [i] as in *Lycht*. The diphthongs / oy / and / ao / do not have any adequate graphic expression and are written as by M. Luther as monographs [ö] as in *Böme*, and [u] as in *upghan*.

Vowel length is coded by three different ways. First by doubled graphemes after the short vowels as in *settede, krüppet*; second, by mute [h] in post vocal position denoting the the previous long vowel as in *upghan*. Besides, long vowels are expressed through gemination as in *seer*.

In most cases M. Luther uses Grapheme [sch], which is new for that period of time. Ir is done in accordance with the existing distribution rules that is before vowels and sonorants. In other positions [s] as in *swenede* is written.

The analysis showed that M. Luther tried to reflect the real condition of German written speech understanding that if the standard were the grouping of ideal rules used by no one it would never have perspective and would be doomed to disappearing (see more about it in [Наумов, 2010]).

M. Luther was a great pragmatist. He was one of the first in the history of German National language development to make an attempt of creating phonetic-based written code of the dialect that was most prestigious in those times. This was meant creating spelling rules available for the majority. At the same time taking only one dialect into account inevitably gave way to the means that later became an obstacle for forming unified German orthography and slowed down the process of pronunciation regulation.

Written form of the National language introduced by M. Luther in the second half of the 16th century took time to be widely accepted in the country. It spreads irregularly. In central and western parts of the country the unification process is much faster, while south-eastern parts (Bavaria and Austria) tried to preserve their social, political and cultural identity which led to separatism in language as a whole and its written manifestation in particular.

G. Gabelenz pointed at dangerously widening gap between the language and its written form in the second half of the 16th century: «The wider is the gap between the orthography and the sounds of the language, the more drastic should be the steps to move spelling towards pronunciation» [Gabelenz, 189: 142].

The 17th century should be considered the most inefficient in the history of German National language. When the Thirty-years War in Germany was over there were no traces of the initial tendency of unification of the language. According to K. Hoffmann, «there is a complete anarchy not only in jurisprudence and governing but in writing as well [Hoffmann, 1875: 16].

M. Raschke calls the 17th century «the century of oblivion of eveything created in the previous century» [Raschke, 1987]. The ties between generation in language development are lost due to unfavorable extrinsic factors. According to K. Hoffmann «There was almost no single sound and no single letter to give unambiguous unified way to read what was written. Most sounds had various graphic expression and most letters had different ways of reading» [Hoffmann, 1875: 20].

The urgency of that problem was felt harder in the second half of the 18th century when political, economic and cultural situation in the country required unification of writing. The nation needed the system of writing that would reflect the centralizing tendency in society directed to creating the new unified social structure. Newly-born capitalism needed not only regularity of production process and welfare distribution but improving all means to achieve that regularity.

Language and its written form were not the least important among those means.

Specialists suggested two ways out of the situation: (i) to return orthographic rules to initial historic state not taking any changes into account (ii) to create new rules based on phonetic principles because historic orthograph in its pure form is unacceptable because of its inconvenience as a result of unnecessary graphic profusion» [Gabelenz, 189: 141]. The same problems arose in other European countries like England and France where sticking to traditional principles forced the society to follow historical rules for centuries.

Somehow or other only one of the leading principles of orthography had to be used — either the one the ear was accustomed to (the phonetic basis of the language) or the one, the eye was accustomed to, which used to be the basis of the written form of the language.

In 1757 K. Gottsched, the supporter of phonetic rules, wrote in his book «Deutsche Sprachkunst»: «Every syllable should be spelled it is clearly heard in good pronunciation». In 1788 J. Adelung in his research «Vollständige Anweisung zur deutschen Orthographie» declared the necessity for spelling rules to reflect rules of pronunciation: «Schreib der besten Aussprache gemäß».

However the question about which pronunciation should be taken as a model because the phonetic form of the language had not been unified yet and there was only dialectal notion of pronunciation standard. That is why Gottsched considered Meissen pronunciation as the model one while Adelung being a Saxon resident preferred Upper-Saxon one.

The activity of J. Grimm played a special role in building the theoretical concept of unified spelling rules in German. The influence of writing on the German historic school is significant. According to Grimm, the purpose of spelling is to give simple and adequate expression of the phonetic form of the language. However it is impossible to count on its total etymologization [Raschke, 1862: 21], advocated by W. Wakkerhel and K. Weinholt.

Grimm understood that orthography was not able to represent the history of certain words. At the same time he did not deny graphic traditions of the past completely.

Since the very beginning Grimm stressed the necessity of Latinization of the graphics because it facilitated the process of teaching writing. In general Grimm tended to simplify spelling rules getting rid of unnecessary forms. Thus Grimm advocated abandoning vowel length symbols. As for the consonants, Grimm promoted abolishing of double graphemes as in *Ros, Kus, Schif*. Being a man of principle, Grimm considered the usage the greatest yardstick of linguistic status of any form or spelling rule. It explains the fact that Grimm who was the first to introduce the grapheme ß], abandons his creation and never mentions it again having decided that it would not be accepted in German system of writing: «bloß aus Nachgibigkeit gegen den Gebrauch» (Наумов: 106). As far as foreign words and proper names Grimm followed Latin graphics. The only way he spelled his name was *Jacob*.

The most influential and logical advocate of language unification in Germany in the first half of the 18th century was K. Gottsched who promoted regulation of spelling based on Upper-Saxon or Meissen pronunciation rules.

Many writers including G. Lessing, V. Schiller, W. Goethe were aware of the necessity of unified orthography. Thus the language of the writers of the 18th century became the second leading factor regulating the spelling rules and the National German language as a whole.

The analysis of their literary works enabled linguists to understand the proportion of social, functional-stylistic components of the literary language, compare phonetic forms with their graphic expression, which in their turn were crucial in further development of spelling rules.

REFERENCES

[1] Наумов, В.В. 2010. *Государство и язык. Формулы власти и безвластия*. Москва, УРСС.

[2] Adelung, J. 1788. *Vollstängige Anweisung zur deutschen Orthographie*. Leipzig.

[3] Gabelenz, G. 1891. Die Sprachwissenschaft, ihre Aufgabe, Methoden und bisherige Ergebnisse. Leipzig.

[4] Gottsched, K. 1757. *Deutsche Sprachkust*. Berlin.

[5] Hoffmann, K. 1875. *Die neuhochdeutsche Rechtschreibung vom Standpunkte der Sprachphilologie und Sprachgeschichte*. Arnstadt.

[6] Raschke, M. 1987. Proben und Grundsätze der deutschen Reform der Rechtschreibung. *Frankfurter Allgemeine Zeitung,* 114.

ST. PETERSBURG PRONUNCIATION STANDARD IN RUSSIAN HARBIN

Elena Oglezneva

Amur State University, Russia
slovoamgu@yandex.ru

ABSTRACT

For two hundred years there have been two variants of the Pronunciation Standard of the Russian language equal in their status: Moscow and St. Petersburg Pronunciation Standard. However, they are moving towards a single integrated pronunciation standard. The paper attempts at analyzing phonetic features of oral speech of the last representatives of the Russian diaspora in Harbin. The results of the analysis are compared to the earlier data.

Keywords: pronunciation standard, vowels, consonants.

1. BACKGROUND

There was an assumption about the domination of St. Petersburg Pronunciation Standard in Russian Harbin [Дземешкевич, 1998; Райан, 2005; Таут, 2001 etc.].

It is a well-known fact that for two hundred years there have been two variants of the Pronunciation Standard of the Russian language equal in their status: Moscow and St. Petersburg (in Soviet period referred to as Leningrad) Pronunciation Standard further referred to as SPPS and MPS correspondingly. The two variants diverged in terms of some vowel and consonant peculiarities as well as certain grammar forms. The main reason determining the differences is phonetic texture of the dialects in ares surrounding Moscow and St. Petersburg [Щерба, 2004; Аванесов, 2005; Вербицкая, 2001].

There are the following phonetic features of MPS: 1) preference of i-sound to e-sound in unstressed reduced syllables; 2) predominance of a-sound in the first pre-stressed syllable after non-palatalized consonants, which is characterized as an open long sound (often as long as the stressed one or sometimes even longer than the stressed one) [Высотский, 1977, Китайгородская, Розанова, 2005]; 3) pronouncing the letter «*щ*» as long palatalized [ш`:]; 4) instability of word-final labial consonants that can be non-palatalized in some words and palatalized in others; 5) the presence of consonants palatalizing due to high magnitude assimilation process [Панов, 2002].

SPPS in its turn is characterized by the tendency of tying the sound to the letter, which is the direct influence of spelling on pronunciation [Аванесов, 2005]. It included the following features: 1) the absence of unstressed e-sound reduction to i-sound; 2) predominance of mid-open or mid [ʌ] in the first pre-stressed syllable after non-palatalized consonants; 3) pronouncing the letter «*щ*» as [ш`ч`]; 4) only non-palatalized word-final labial consonants; 5) the absence of consonants palatalizing due to assimilation with the neighboring palatalized consonants compared to stable palatalizing in MPS. In addition to that, M. V. Panov wrote about «spelling determining the pronunciation of certain grammatical forms» as one of the characteristic features of SPPS. Other peculiarities of SPPS are introduced in L. A. Verbitskaya's works quoted above. The data have been received from different sources of different time periods beginning with the middle of the XIX century up to 1970s and as a rule are based on the researcher's oral perception [Вербицкая, 2001]. In the speech of contemporary St. Petersburg residents following the Standard there were only 19 out of 39 peculiarities traditionally spoken about, which obviously proves that SPPS has undergone changes and that both MPS and SPPS are moving towards a single integrated pronunciation standard [Вербицкая, 2001].

2. CURRENT STUDY

The paper presents an attempt to analyze oral speech of the last representatives of the Russian diaspora in Harbin as far as their phonetic features are concerned. I also aim at comparing the results of the analysis to the earlier data on Russian diaspora in Harbin speech and the currents SPPS and MPS described in scientific works. Further more, the purpose of the analysis is to understand which of the two standards dominated for Russian emigrants in Eastern countries, particularly in Harbin. Finally, I will try to confirm or disprove the opin-

ion about the domination of SPPS in Harbin that exists in memoirs and scientific non-linguistic works. The perspective for future studies will be determining the causes of such domination.
Besides the 5 major differences between SPPS and MPS mentioned above there are some peripheral ones included to help diagnose orthoepic status of the speakers used in the current experiment. The number of phonetic peculiarities of SPPS chosen from the complete list performed by Ludmila Verbitskaya [Вербицкая, 2001] was 15.

2.1. Subjects

The subjects were 8 speakers of Russian diaspora including re-emigrants from China to Australia who speak English — the official language of the country where they currently live. The recordings were done in a quiet room and later digitized for further analysis. Acoustic analysis enabled to get objective data about the variety of of the phonetic features under consideration.
Speech samples from different speakers turned out to be of different size which made it difficult to analyze certain positions, therefore some pronunciation features of some speakers, including the re-emigrants from China to Australia, remain unclear.

3. RESULTS AND DISCUSSION

The results of the analysis are presented in tables 1-3 (see the Appendix). Thus, tables 1-2 demonstrate the domination of SPPS both in the speech of the last representatives of Russian Diaspora in Harbin and in the speech of the Russian re-emigrants from China to Australia.
Table 3 contains the information about the differences concerning MPS and SPPS contemporary to Harbin emigration period, Harbin pronunciation variant itself and modern pronunciation standard. The data enable to compare the variants formulating their common features and differences in certain positions. They also demonstrate unidirectional tendencies and divergences of the modern pronunciation standard of the Russian language of the Metropolis and the one for Russian Diaspora in Eastern countries at the beginning of the XXI century.
Data analysis performed during the comparison of the pronunciation variants mentioned enabled to state that the phonetic features of Russian speech in Harbin were very close to the ones of SPPS, which still existed in the Metropolis at the beginning of the XX century. However it was gradually being destroyed moving towards a single integrated pronunciation standard. Our data show that in oral speech of Russian emigrants to eastern countries pronunciation standard close to SPPS endured for longer period of time.
The correspondence of the Russian-Harbin pronunciation is proven in 12 out of 15 positions. The brightest peculiarities are: unstressed [e] realization; mid-open or mid [ʌ] predominance for both «*a*» and «*o*» letters in the first pre-stressed syllable after non-palatalized consonants; mostly non-palatalized word-final labial consonants, and the absence of assimilation in reflexive forms of the Verb.
Features of MPS (old MPS) were few and unstable. They could be seen only in the speech of some subjects (e.g. Subject EN who in some cases pronounced [ut/jut] in the third person plural verb forms).
Most phonetic peculiarities fixed in the speech of Russian emigrants living in Harbin coincide more often with SPPS and less often with MPS and thus demonstrate certain common features with the contemporary single integrated pronunciation standard. So we can see the similar tendencies of pronunciation standard moving towards unification both with the residents of the Metropolis and Russian emigrants to eastern countries.
It turns out that intuitive assumptions of Harbin Russian speech reflecting SPPS are proven true according to a number of features.
This raises two questions: (i) How could only one of the two pronunciation standards appear and develop for a long period of time in the speech of Russian emigrants to Harbin and their descendants? (ii) How could it resist the influence of other languages and dialects of the Russian language that contacted the Russian Literary Language in Harbin?
M. V. Panov in his book «The history of the Russian Standard Pronunciation of the XVIII-XX centuries» wrote about the decline of the orthoepic culture in society and the search for the causes of that phenomenon. He addressed the post-revolutionary period in the history of our state in general and the history of the Russian language in particular: «After the revolution the local phonetic varieties of the Russian Literary Language such as Moscow variety and Leningrad variety. Mixing of different masses of people during political turmoil was a constant large-scale process in Russia, therefore there was no chance for the local features to be kept intact. People's desire to get acquainted with the national culture, destroying social barriers, constant interaction during speech process led to the unification of the orthoepy and high orthoepic culture, did they not? Unfortunately they did not.

Literacy of speech went broadway but not deepway. Mediocre level of speech spread out everywhere. That was the level of pronunciation at the end of 1930s and it remained in stagnation» [Панов, 2002]. The reason, of according to M.V. Panov, is the «absence of the «mountain»: the bog has no landmark to dry out and go up. There is no socially acceptable prestigious standard in the realm of pronunciation culture».

However, this standard did not disappear for Russian emigrants in Eastern countries and was widespread particularly in Harbin up to the period of leaving China and then remained intact in the speech of the last representatives of Russian Diaspora in Harbin and re-emigrants from Harbin to other countries. The ideal model for imitation in Harbin was the speech of educated people — the ones who arrived to build the Chinese-Eastern Railroad. There were many residents of St.Petersburg among them in the first wave: rail-road engineers, medical doctors, teachers. They got excellent education in Russia and brought it to Harbin to create the cultural environment that was an application of an ideal model that included good literate speech that started being used by the rather varied Russian population of Harbin. The same tendency was observed during the second wave when educated elite from provincial towns of Siberia and the Far East came to Harbin. The pronunciation standard formed during those two waves started to be taught in Russian lyceums and gymnasiums and later in higher educational institutions in Harbin.

Thus, it is possible to consider that SPPS in Harbin appeared with the appearing of its native speakers from St. Petersburg and then was spread further as the model pronunciation. Being very effective in case the speaker understand the necessity of ethnic self-preservation when the Metropolis is far away, deliberate imposing of the norm however needs to be supported by a number of objective factors. One of these factors plaid a crucial role in Russian-Harbin Standard development.

SPPS is traditionally defined as the one tied to spelling or as pedantically bookish. M. V. Panov wrote: «it appeared in the city that was inhabited in a hurry and turned out to be a mess of dialects. The only possibility to achieve some unification of pronunciation in this amalgamation of dialects was to tie it to spelling. There was no other uniting factor».

M. V. Panov on the one hand blamed SPPS for being tied to spelling, but on the other hand he said that SPPS bridges the gap between orthography and pronunciation.

The Russian population of Harbin varied a lot representing different social groups of different regions of the Metropolis. Pronunciation standard tied to spelling was supported in Harbin society because it was the only possible variant to unite all Russian-speaking emigrants who had their own regional and social peculiarities in their speech.

The similar tendency was developing in the Metropolis. It resulted in gradual displacing MPS by SPPS.

4. CONCLUSION

Keeping in mind everything mentioned above I conclude that pronunciation standard of the Russian language in Harbin had certain peculiar features connected with the distribution of possible variants of pronunciation. The dominating variants were the ones that coincided with SPPS and later the single integrated pronunciation standard of the Russian language, that became the modern standard. The common feature of those variants is their closer connection with the direct sound content of the letter.

5. REFERENCES

[1] Аванесов, Р. А. 2005. Русское литературное произношение. 5-е изд. Москва: КомКнига.

[2] Вербицкая, Л. А. 2001. Давайте говорить правильно. 2-е изд. Москва: Высшая школа.

[3] Высотский, С. С. 1977. Звук речи в контексте. Диалектологические исследования по русскому языку. Москва, 24-38.

[4] Дземешкевич, Л. 1998. Харбинцы. Омск.

Китайгородская, М.В. Розанова Н.Н. 2005. Речь москвичей: коммуникативно-культурологический аспект. Москва: Научный мир.

[5] Панов, М. В. 2002. История русского литературного произношения XVIII-XX веков. 2-изд. Москва: Едиториал УРСС.

[6] Райан, Н. В. 2005. Россия-Харбин-Австралия: сохранение и утрата языка на примере русской диаспоры, прожившей XX век вне России. Москва: Русский путь.

[7] Таут, М.П. 2001. Сбереженный русский язык. Русская Атлантида, 6, 35-38. Челябинск

[8] Щерба, Л.В. 2004. Краткий очерк русского произношения. *Языковая система и речевая деятельность*. 2-е изд. Москва: Едиториал УРСС. 171-175.

APPENDIX

Table 1: Pronunciation peculiarities of last representatives of Russian diaspora in Harbin. «+» is used to denote 100% matching SPPS; «–» – not matching SPPS; «+/–» – matching SPPS for the majority of cases; «–/+» – not matching SPPS for the majority of cases.

St.Petersburg Standard peculiarities	Speaker 1	Speaker 2	Speaker 3	Speaker 4
Vowels				
1.[e] or [eᵘ] in the first prestressed syllable after palatalized consonants	+ [л` eᵘ ]нт**я**[eᵘ ]в [н` eᵘ]в**е**стка	+ [н` eᵘ ]л**е**по ма[т` eᵘ]рьялов	+ [р` eᵘ ]ш**и**ли [п` eᵘ]т**а**к	+ [б` eᵘ ]р**и**те [т` eᵘ]п**е**рь
2.mid-open [aᵡ] in the first pre-stressed syllable after non-palatalized consonants	+ гов[aᵡ]р**и**ла з[aᵡ]б**ы**л	+ х[aᵡ]т**е**л т[aᵡ]нц**у**ют	+ л[aᵡ]т**о**к ск[aᵡ]зал	+ б[aᵡ]льн**и**ца вл[aᵡ]д**ы**ка
Consonants				
3.Pronouncing letter *щ* as [ш`ч`]	– уч**и**ли[ш`ш`]е	– [в**о**бш`ш`ем]	– е[ш`ш`]ё	+/– [јеш`ч`**о**] п**о**мо[ш`]
4. Pronouncing non-palatalized word-final or pre- [j] consonants	+ [в**о**с`ем] [с`емј**а**] [кров](кровь)	+/– любо[ф] церко[ф`]	+/– восе[м] церко[ф`]	+/– [в**о**с`ем] церко[ф`]
5.No palatalizing as assimilation process	+/– е[ст`]**е**ственный бол**е**[з`н`]и	– п**е**[н`с`]ия сл**а**бо[с`т`]	–/+ ко[нф`е]тки пе[н`с`]ия ве[з`д`]де	–/+ [дв`**е**с`т`и] се[н`т`]ябрь

Table 2: Russian re-emigrants from Harbin Australia and their descendants Russian speech peculiarities. «+» is used to denote 100% matching SPPS; «–» – not matching SPPS; «+/–» – matching SPPS for the majority of cases; «–/+» – not matching SPPS for the majority of cases.

St.Petersburg Standard peculiarities	Speaker 6, born in 1939	Speaker 7, born in 1943	Speaker 8, born in 1938	Speaker 9, born in 1975
Vowels				
1.[e] or [eᵘ] in the first prestressed syllable after palatalized consonants	+ [с`eᵘ]рь**ё**зный	+ [т` eᵘ]б**е**	+ у[м` eᵘ]р**е**ть	+ [н` eᵘ]-знаю до-[п` eᵘ]ти лет
2.mid-open [aᵡ] in the first pre-stressed syllable after non-palatalized consonants	+ п[aᵡ]кл**о**н в к[aᵡ]р**е**те	+ п[aᵡ]ртр**е**т зн[aᵡ]к**о**мый [дъ]г**а**дливый	+ с[aᵡ]шьёт п[aᵡ]правл**я**ет с[ъ]м**а**	+ т[aᵡ]**к**а**я** п[aᵡ]-р**у**сски с[ъ]л**и**сты
Consonants				
3.Pronouncing letter *щ* as [ш`ч`]	– е[ш`ш`]ё	– е[ш`ш`]ё	– **и**[ш`ш`]ешь	– е[ш`ш`]ё
4. Pronouncing non-palatalized word-final or pre- [j] consonants	+/– [в**о**с`ем] церко[в`]	+/– се[м] церко[в`] [с`емј**а**]	+ в**о**се[м] се[м] [с`емј**а**]	+ [в**о**с`ем]
5.No palatalizing as assimilation process	–/+ е[с`т`] р**а**[зн`]ица [с`н`]ять	–/+ [дв`]ерь вм**е**[с`т`]е [з`д`]есь	– **о**бла[с`т`] **е**[з`д`]или жи[з`н`]	– [з`д`]есь жи[з`н`]

Table 3: Generalizing and comparison

MPS	SPPS	Harbin pronunciation	Modern Russian norm
Vowels			
1.after palatalized consonants in unstressed position: [i]/[e]			
1.**[i]**. for the letters **е** and **я** in all unstressed positions [Pronunciation Dictionary ed. by Ushakov: 11]	1.**[e]** for the letters **е, а, я** after palatalized consonants [Verbitskaya, 2001: 59]	1. [e] or [eᵘ]. matching SPPS	1.**[i]** [Ozhegov's Pronunciation Dictionary: 7];
2. Vowels after non-palatalized consonants			
2. open [a] in the first pre-stressed syllable for the letters **а, о** [Pronunciation Dictionary ed. by Ushakov: 11]; [Panov: 54].	2. mid-open [aˇ] or **[ъ]** in the first pre-stressed syllable for the letters **а, о.** [Panov, c.5 4].	2. **mid-**open [aˇ] or **[ъ]** in the first pre-stressed syllable for the letters **а, о.** matching SPPS	2. open [a] in the first pre-stressed syllable for the letters **а, о** [Ozhegov's Pronunciation Dictionary: 7].
Consonants			
*3. Pronouncing letter **щ***			
3. as long palatalized [ш`:] [Pronunciation Dictionary ed. by Ushakov: 12] [Panov: 54]	3.as [ш`ч`] [Panov: 54].	3. as long palatalized [ш`:] matching MPS and modern standard	3.as long palatalized [ш`:] [Ozhegov's Pronunciation Dictionary: 8].
4. Pronouncing word-final labial consonants and labials before [j]			
4. in some words — non-palatalized, in others — palatalized. [Panov, c. 54].	4. only non-palatalized [Panov: 54]; [Verbitskaya, 2001: 60].	4. only non-palatalized except for the word *церковь (church)*. Matching SPPS	4. palatalized for –мь, -бь, -пь, -вь, -фь letters [Pronunciation Dictionary: 651].
5. +/-- palatalizing as a result of assimilation			
5. + for dental consonants (letters **д, т, с, з, н**), also for letters **р** and **ф** before palatalized consonants [Pronunciation Dictionary ed. by Ushakov: 12]; [Panov: 54].	5.+ for consonants of the same speech organ -- for consonants of different speech organ [Verbitskaya, 2001: 60]. [Panov: 54].	5. + for consonants of the same speech organ (or sometimes even --) -- for consonants of different speech organ matching SPPS and modern standard	5. Palatalizing might be present but it is fading in modern language [Pronunciation Dictionary: 653]; «No palatalizing is becoming more and more characteristic» [Ozhegov's Pronunciation Dictionary: 8].

PROSODIC FEATURES IN SPONTANEOUS MONOLOGUE PERCEPTION

Marina Pirogova, Tatiana Shuiskaya

Amur State University, Russia
pirogova_marina@mail.ru

ABSTRACT

This paper summarizes recent research concerning the relationship between intonation and spontaneous speech perception.
After introductory comments on the research arrangement, we discuss the relationship between some prosodic features and the process of speech perception. The focus of the current study is the correspondence between prosodic arrangement of spontaneous monologue and its highly informative and low informative segments perception.

Keywords: prosodic features, informative segments.

1. INTRODUCTION

Recent reviews of the relationship between prosody and speech perception [Ayers, 1994; Birch, Clifton, 1995; Fowler, Housum, 1987; Ito, 2002; Ladefoged, 2003; Бондарко et al., 2000] show how prosodic structure of spoken language influences the process of defining so called highly informative segments (HIS) and low informative segments (LIS).
In this paper we concentrate on the relationship of prosody and sense division. The main questions which we discuss in this paper are the following:

- What are prosodic features which mark so called highly informative segments (HIS) and low informative segments (LIS) in spontaneous speech?
- Do these features really help the listener differentiate these parts of the spontaneous monologue?
- Does the listener's choice of HIS depend only on prosody?

2. METHOD

The aim of the research was to examine the basic peculiarities of sense division of spontaneous monologue, to examine the basic peculiarities of spontaneous speech perception by native speakers and to find out the most important prosodic features which may influence the process of spontaneous speech perception.

The prosodic phenomena were examined in connection with their realization in different sense segments (we call these segments "highly informative segments (HIS)" and "low informative segments (LIS)".

2.1. Materials

We used speech material representing spontaneous monologues of three native speakers of General American (without any sharp dialect features) with the age varying from 27 to 50. The speakers did not report of any speech and hearing disorders.
The monologues of total duration of 20 minutes were recorded under quiet conditions. The recordings were digitized at a sampling rate of 44 kHz and 16 bit quantization.
Ten professional linguists marked important words in each phrase of the monologues. The criterion for including words marked by the experts in HIS was 60 % of experts agreement.

2.2. Participants and procedure

There were 10 participants – native speakers of General American English aged from 25 to 50 without linguistic education (naive listeners). None of the listeners reported of any speech or hearing disorders. They were instructed to listen to spontaneous monologues phrase by phrase and to write out the most important from their personal point of view words in each phrase. Each listener worked separately and they were not given written samples of the monologues they were listening to. The number of listening was limited to three times.
According to the results of the naive listeners perception all the words of the monologues were divided into two groups: a) highly informative segments (HIS) b) low informative segments (LIS). The words were included in HIS if they were marked not less than by 60 % of the listeners.
Comparing the results of the naive listeners perception to the experts' marking of the words we saw a discrepancy between them: not all HIS that were marked by the experts were written out by the naive listeners. It clearly led to the necessity to in-

troduce one more group of words for our phonetic analysis – the third group – c) HIS not marked by the naive listeners.

In this paper we present the results of two groups' analysis: highly informative segments (HIS-1) marked by both the experts and the naive listeners and highly informative segments (HIS-2) marked by the experts but not marked by the listeners. HIS-2 group included 22 % of words marked by the experts.

These two groups – HIS-1 and HIS-2 – are highly remarkable. Acoustic study demonstrates that all HIS definitely have certain kind of prosodic prominence. However, some of these segments were not marked by the experts.

We examined prosodic features of these words with the help of the Programs of Computer Speech Analysis "Sound Forge (Sonic Foundry)" and "EDS" (made by SIAL). We were examining the following prosodic features of spontaneous monologue: tempo, intensity, pitch level and pauses.

3. RESULTS AND DISCUSSIONS

While listening to the spontaneous monologue the listeners get some pieces of information which they consider most important, informative.

It is common knowledge that certain prosodic features of these segments play the main role in this process. However, the results of the research show that sometimes these criteria are not enough. It can be proved by the fact that sometimes listeners did not mark HIS as important parts of the utterance. The reason might lie within non-standard exploitation of prosodic features in spontaneous speech vs prepared speech.

Spontaneous speech is characterized by enormous variation of different prosodic features including tempo, character of pauses, intensity.

Speakers use various hesitation phenomena unintentionally. This enables them to have some time to "search for" a suitable way of expressing their ideas clearly and properly, thus not interrupt the flow of speech. The length and type of pauses differ throughout the unprepared monologue and the amount of pauses varies enormously in prepared and non-prepared speech. There are many examples of the absence of syntagmatic and phrasal pauses in spontaneous speech. We also noticed a lot of examples of filled and non-filled intersyntagmatic pauses in spontaneous monologues of all the speakers under analysis.

Using definite intonation pattern in spontaneous speech is situation-dependent. First, there were a great number of phrase-final segments and even text-final utterances in spontaneous monologues of S1, S2, and S3 that were pronounced with Rising or Level intonation unlike in prepared narrative that dictates Falling tone for such segments and utterances.

Second, the speech tempo was from time to time faster in HIS and slower in LIS providing smaller duration of HIS compared to that of LIS although in a prepared narrative it should be vice-versa.

The basic differences and peculiarities of prepared and unprepared speech production influence the realization of prosodic features in various oral speech types (e.g. reading vs spontaneous monologue).

We analyzed prosodic features – pitch level, tempo, pauses, and intensity – of 76 words (tokens) which experts defined as HIS. We tried to find an answer for the following question. What are prosodic features (or combinations of prosodic features) which help the listener correspond with these words with HIS?

Duration of the tokens and and pauses was measured using "Sound Forge". Other acoustic measurements – pitch direction, pitch level, and intensity – were made using "EDS".

During the acoustic analysis we found out that in some of them only one prosodic feature was strongly exploited to make the word prosodically prominent, while other words several prosodic features were used to make them prosodically prominent.

We had 16 possible variants of prosodic prominence received by:

1) pitch level (Pl)
2) intensity (I)
3) tempo (T)
4) pause (P)
5) pitch level + intensity (Pl + I)
6) pitch level + tempo (Pl + T)
7) pitch level + pause (Pl + P)
8) intensity + tempo (I + T)
9) intensity + pause (I + P)
10) tempo + pause (T + P)
11) pitch level + intensity + tempo (Pl + I + T)
12) pitch level + intensity + pause (Pl + I + P)
13) pitch level + tempo + pause (Pl + T + P)
14) intensity + pause + tempo (I + P + T)
15) pitch level + intensity + tempo + pause (Pl + I + T + P)
16) no distinct prosodic arrangement

All the words from HIS-1 were prosodically prominent. In most cases major prosodic feature was PL change (79,6% of the tokens). Intensity and tempo

helped the listeners mark HIS in 42,7% and 37,2% of the tokens correspondingly.
Less than 11,1% of all the analyzed tokens had all the mentioned prosodic features involved – Pl + I + T + P. 3% of the tokens had no prosodic prominence. However, the experts marked them as HIS.
Here are some examples of the analyzed words given in the context:

EXAMPLE 1
The service industry in America is continually growing. (S1)

EXAMPLE 2
I really enjoyed studying, being challenged intellectually, meeting new friends, meeting new people and a lot of the leisure activities that I was involved in at Stanford. (S2)

EXAMPLE 3
They would really want to know about me, they'd want to know about life in America and they are very sincere. (S2)

EXAMPLE 4
Wherever my opinion was before I came, it has changed as I say even in spite of the problems and the difficulties that people are facing here. (S3)

All words from HIS-2 group were also characterized by certain kinds of the mentioned 15 variants of prosodic prominence. In 4,6% of HIS-2 tokens the major prosodic feature was PL change, and in 4,6% major prosodic feature was higher intensity. Neither pause nor tempo seemed to play any important role in listeners' choice.
In 45,2% of all the words from this group the following combinations of prosodic features were defined: Pl + T + P; Pl + T; Pl + I + T – 18,2%, 13, 5% and 13,5% correspondingly.
Another group of words (where major prosodic feature was intensity) seems to be less impressive and accounts for 50% of all the words from HIS-1 group.
The combination of prosodic features Pl + L + T + P was found in 5 words from HIS-2 group. Although prosodic prominence of these words was obvious from the acoustic point of view and might influence the listeners' choice, the listeners didn't find them highly informative. It enables us to think that there are certain instances when prosodic features are not enough for listeners to interpret them as highly informative and that the role of different prosodic features in the process is not the same.

Here are some examples of the analyzed words given in the context:

EXAMPLE 1
Moving and living in another country is the question. (S1)

EXAMPLE 2
They don't even a language lab in the classroom and especially teaching language is beyond comprehension. (S2)

EXAMPLE 3
Usually undergraduate is four years and Master's degree is another two years, but I managed to do in five. (S3)

EXAMPLE 4
Making preparations it is very difficult to know what you need in another country. (S3)

EXAMPLE 5
She sends me job advertisements from the university, she sends me information, she tells me what's going on there and she'll be encouraging me when I go back to the U.S. to go in for some informational interviews to get some information about job opportunities in the field I'm interested in. (S1)

The results of the research show that the intensity of HIS tokens in spontaneous speech of different speakers vary. The intensity of HIS is higher then of LIS. But it is very remarkable that HIS, which were not marked by the experts also had very high intensity, but it did not influence the listeners' choice. We may suppose that this prosodic feature is not enough for marking the HIS.
The pitch variation in HIS in comparison with the average pitch level of the phrase might have helped the experts to mark these words as HIS.
The tempo of speech plays the most important and vivid role in HIS perception. The most important words are pronounced more slowly than LIS. But this feature turned out speaker-dependent to a great extent. The results of the Computer Analysis prove the fact that there are vivid contrasts between HIS and LIS. The more expressive and emotional the speaker's speech is the sharper these contrasts are. Here we should pay attention not only to some particular features of the speaker's speech, but also to the nature spontaneous monologue as to the means of communicating and sharing information.
The duration of pauses in spontaneous and prepared speech vary due to some psychological rea-

sons: unprepared speech is highly emotional, there is not much time for thinking about the plan of the narration, a person may feel excitement, hesitation, he speaks faster then usually etc. In spontaneous speech there are less syntagmatic and phrasal pauses than in reading. The phenomena "prepared – non-prepared speech" may explain the absence of intersyntagmatic pauses in reading and their presence in spontaneous monologue. The results of the research demonstrate that syntagmatic and phrasal pauses (either filled or non-filled) and hesitation pauses of all kinds were the longest in spontaneous speech.

Studying basic perception models of HIS of either spontaneous or prepared speech is very useful and practice-oriented. The results of such experiments can be useful for solving a number of Computer Speech Analysis problems and in teaching Phonetics and Speech Techniques.

4. CONCLUSION

Data described show that the problem of speech perception is more complicated than it seems to be. A number of important questions remain in this area, which we simply list here as indications of future research. How do prosodic features interact with other information sources, such as the semantic features of lexical items to determine sentence comprehension? Is there a direct correlation between prosodic arrangement of HIS and LIS and their perception in spontaneous speech?

5. REFERENCES

[1] Ayers, G. 1994. *Discourse functions of pitch range in spontaneous and read speech.* OSU Working Papers in Linguistics. 44, 1-49.

[2] Birch S., Clifton C. E. 1995. *Focus, accent, and argument structure: effects on language comprehension.* Language and Speech. 38, 365-391.

[3] Fowler C. A., Housum J. 1987, Talkers' signaling of "New" and "Old" words in speech and listeners perception and use of the distinction. *J. of Memory and Language*, 26, 489-504.

[4] Ito, K. 2002. *The interaction of focus and lexical pitch accent in speech production and dialogue comprehension.* Evidence from Japanese and Basque, PhD, Univ.III., Urbanna-Campaign.

[5] Ladefoged, P. 2003. Phonetic fieldwork. *Proc. 15th ICPhS* Barcelona, 203-206.

[6] Бондарко Л. В., Вербицкая Л. А., Гордина М. В. 2000. *Основы общей фонетики.* СПб: СпбГУ.

WORD BOUNDARIES IN NATIVE POLISH SPEECH

Andrzej Porzuczek, Arkadiusz Rojczyk

University of Silesia, Katowice, Poland
ampj24@wp.pl , arkadiusz.rojczyk@us.edu.pl

ABSTRACT

A number of studies have suggested that word boundaries are regularly signalled in speech. Syllable onset consonants in word-initial positions are longer than corresponding word-medial consonants, while word-final rhymes or whole syllables tend to be longer than non-final ones, also in non-phrase-final positions. However, instances of ambiguous phrases or foreign language learners' segmentation problems suggest that these acoustic cues are often insufficient for successful word boundary recognition. Therefore, we hypothesise that in connected speech, word boundaries are not indicated if the speaker does not anticipate comprehension problems on the part of the listener. Consequently, we test segmental duration difference of /a/, /v/ and /a/ in *dodawanie* vs *Doda Wanię*, /v/ in two homophonic Polish phrases: (1) *Lubi dodawanie*? (Does she like addition?) and (2) *Lubi Doda Wanię*? (Does Doda like Wania?) It turns out that the presence of a word boundary induces significant lengthening of the postboundary consonant /v/, slight but significant lengthening of the following vowel /a/, but no significant lengthening of the preboundary vowel /a/.

Keywords: word boundary, prosodic domains, segmental duration.

1. INTRODUCTION

A number of studies have recently been devoted to locus and domain of prosodic processes (cf. [28]). Processes related to domain edges affect segments near unit boundaries. In this way, additional cues are given to the listener concerning the utterance structure. The tendency for fluent speech makes pauses regularly coincide only with higher-level prosodic domains, especially the utterance. The boundaries of lower level units are rarely marked by pauses, but domain-initial and final lengthening are reported to operate at lower levels as well.

Initial articulatory strengthening, correlated with prolonged duration, has been observed by Fougeron and Keating [17], Byrd and Saltzman [5], Fougeron [16], Cho and Keating [11], Keating et al. [19], Bombien et al. [3], Cho and Keating [12] in English, French, Korean and Taiwanese. Oller [22] and Cooper [14] indicated that syllable onsets in word-initial positions are longer than corresponding word-medial consonants thus suggesting that the word forms a domain for the process. For instance, measures of linguopalatal contact have shown that consonants are generally produced with greater articulatory magnitude in word-initial relative to word-medial positions in Korean ([10] in [8], [11]), French ([16]), and English ([20]). Cooper [14] reported that VOT in initial /k/ in /ki'kik/ was longer than VOT of medial /k/ in /kikik/, which was taken as evidence that, under the same stress conditions, initial stops have longer VOT values than medial stops. Turk and Shattuck-Hufnagel [26] measured word-initial lengthening in *tuna choir* vs. *tune acquire* and found that word-initial /k/ in *choir* was longer than word-medial /k/ in *acquire*.

Durational lengthening in word-initial positions has been demonstrated to be an effective cue for perceiving word boundary by the listeners. Shatzman and McQueen [24] used an eye-tracking experiment to show that durational differences between word-final and word-initial /s/ in Dutch sequences such as *eens pot* (once jar) and *een spot* (a spotlight) are perceived and utilized by the listeners to resolve lexical ambiguity. Similar results were reported by Cho et al. [13], who used a cross-modal identity priming to demonstrate that initial strengthening serves as an acoustic cue in segmentation of continuous speech by American English listeners.

The evidence for final lengthening is even more abundant (e.g. [4], [6], [7], [9], [15], [18], [21], [29]). With respect to word, Beckman and Edwards [1], and Wightman et al. [29] found word-final rhymes to be longer than non-final ones, also in non-phrase-final positions. Klatt [21] proved that whole word-final syllables are longer. Turk and White [27] observed that, in sequences such as *shakedown stairs* vs. *shake downstairs*, the first syllable *shake* was longer when followed immedi-

ately by a word boundary than when it was the first syllable in a bisyllabic word.
Word-final lengthening of segments is accompanied by a strengthening effect on both vowels and consonants. Cho [8] notes that the pre-boundary, word-final, phenomena can be thought of as strengthening rather than simply as lengthening. For example, word-final vowels have been found to be produced with greater articulatory magnitude ([8], [17]), and with more extreme articulation evidenced by greater jaw lowering ([15]).

2. CURRENT STUDY

All these domain-edge processes, particularly initial strengthening, enhance segment-specific articulations and aid lexical access. However, instances of ambiguous phrases or foreign language learners' segmentation problems suggest that these acoustic cues are often either insufficient for successful word boundary recognition or non-existent. Therefore, even though a recent study ([23]) has also found durational word-boundary cues in Polish phrases *brat Adama* and *brata dama*, we hypothesise that in connected speech, word boundaries are not indicated if the speaker does not anticipate comprehension problems on the part of the listener.

2.1. Materials

In line with the general hypothesis formulated here, we expect no significant duration difference between /a/ in *dodawanie* and *Doda Wanię* or /v/ in *dodawanie* and *Wanię* in two homophonic Polish phrases: (1) *Lubi dodawanie*? (Does she like addition?) and (2) *Lubi Doda Wanię*? (Does Doda like Wania?) if they are uttered in a disambiguating context. Apart from the two segments, we also measured the duration of /a/ in *dodawanie* and *Wanię* in order to observe if the initial strengthening effects extend beyond the onset consonant.
The target phrases were constructed to be both homophonic and equally natural in communication and thus ambiguous. In order to enhance context disambiguation, each was preceded by three similar sentences, with the focus noun replaced by another lexical item of similar semantic properties (male names and arithmetical operations - see Appendix). A complete test set comprised the four sentences repeated twice and the initial sentence added in order not to leave the tested unit at the end of the list.

2.2. Participants

A total of thirty native speakers of Polish participated in the study. They were recruited from students at University of Silesia in Katowice. None of them reported to have any current speech or hearing disorders nor any history of such. They ranged in age from 19 years to 22 years (Mean 19.6; Median 20). Gender was not balanced; there were 20 females and 10 males.

2.3. Procedure

The participants were presented with a printed version of two sets of sequences. They were instructed to read both sets in a natural conversational style. They were encouraged to avoid pauses between separate words and to maintain their natural tempo of speaking throughout the whole experiment. Each recording took approximately ten minutes.

2.4. Recording

All recordings took place in a quiet room in the Acoustics Laboratory at the Institute of English, University of Silesia. The signal was captured with a headset condenser microphone Sennheiser HME 26-600S, phantom powered and preamplified with USBPre 2 (Sound Devices). The sound was recorded into .wav format with the sampling rate of 48 KHz, 24-bit quantization.

2.5. Measurements

Out of thirty recorded speakers, one speaker was excluded from final analysis due to severe production disfluencies. Final measurements included 29 speakers x 4 sequences x 5 segments, which resulted in 580 tokens analysed.
All measurements were performed in PRAAT ([2]) by inspecting a spectrographic display and waveform. The sequences were segmented into /o/ /d/ /a/ /v/ /a/ using TextGrid implemented in PRAAT. The first vowel /o/ was measured from the onset of a release burst of a preceding /d/ to the end of a vowel signalled by a cessation of the second formant. A decision was made to define vowel duration as an interval beginning at the end of constriction of a preceding stop (see [25]). This choice was motivated by the fact that speakers varied largely in production of /d/-to-/o/ transitions, in that some speakers produced long release bursts while others were characterised by an immediate transition from stop constriction to the following vowel. /d/ was delimited as an interval signalled by a decrease in overall amplitude and cessation of formant energy. The criteria for measuring the first /a/ followed

those defined for /o/, the only difference being that the offset of /a/ coincided here with the onset of frication energy for the following /v/. /v/ was measured as a period of continuous energy in higher frequency ranges. The second /a/ was measured from the offset of frication to the onset of abrupt spectral changes corresponding to the appearance of nasal formants in /n/.

3. RESULTS AND DISCUSSION

Table 1 presents mean durations of the five measured segments in *dodawanie* and *Doda Wanię* as pronounced by 29 Polish speakers.

Table 1: Mean segmental duration (ms) in *dodawanie* and *Doda Wanię* (two repetitions each). Standard deviation in parentheses

context	o	d	**a**	**v**	a	Σ
dodawanie 1	88 (8)	47 (9)	**98 (8)**	**69 (17)**	120 (15)	421 (36)
dodawanie 2	87 (10)	46 (10)	**98 (10)**	**68 (17)**	121 (16)	419 (41)
Doda Wanię 1	91 (9)	49 (11)	**101 (13)**	**84 (20)**	131 (16)	456 (45)
Doda Wanię 2	89 (12)	52 (11)	**102 (17)**	**86 (20)**	128 (16)	456 (50)
dodawanie 1+2	**87 (9)**	**47 (10)**	**98 (9)**	**68 (17)**	**120 (16)**	**420 (38)**
Doda Wanię 1+2	**90 (11)**	**50 (11)**	**101 (15)**	**85 (20)**	**129 (16)**	**456 (47)**
increase	**3.4%**	**6.4%**	**3.1%**	**25%**	**7.5%**	**8.6%**

The mean durations of each segment is almost exactly the same in two repetitions of the same phrase. Each segment before the word boundary in /–oda#va/ is slightly (non-significantly) longer while the ones following the boundary are significantly longer, especially /v/, whose length increases by 25%, but also /a/ (7.5% longer). A paired t-test for individual speakers (cf. Table 2) illustrates the influence of word boundary on the duration of surrounding sounds.

The data presented in Table 2 suggest very clearly that the word is a domain for initial lengthening but not final lengthening. Initial lengthening is conspicuous in the consonant in a #CV structure and, though attenuated, also significant in the following vowel, contrary to many studies (see [12] for review).

Table 2: Paired t-test results for individual speaker's segmental length in a phrase with and without a word boundary. Significant values ($p \leq .01$) in bold

contexts compared	o	d	**a**	**v**	a	Σ
dodawanie 1: Doda Wanię 1	.053	.132	.116	**< .001**	**< .001**	**< .001**
dodawanie 2: Doda Wanię 2	.164	**.002**	.064	**< .001**	**.004**	**< .001**
dodawanie 1: Doda Wanię 2	.305	**.014**	.107	**< .001**	**.002**	**< .001**
dodawanie 2: Doda Wanię 1	.026	.084	.079	**< .001**	**< .001**	**< .001**

As expected, no statistically significant difference in individual speaker's segmental duration in two repetitions of the same phrase has been indicated by the same statistical tool (Table 3).

Table 3: Paired t-test results for individual speaker's segmental length in two repetitions of a phrase with or without a word boundary

contexts compared	o	d	**a**	**v**	a	Σ
dodawanie1: dodawanie2	.321	.244	.375	.184	.274	.265
Doda Wanię1: Doda Wanię 2	.194	.08	.326	.33	.029	.48

It seems that initial lengthening/strengthening should form a sufficient cue for the listener to recognise the beginning of a new word in a speech chain. However, despite the significant mean duration differences, the temporal domain-edge cues are not evident in all speakers and their cumulative character is not reflected in clear separation of the same prosodic levels. It is noteworthy that the word-initial /v/ was not made longer than the word-medial consonant by as many as 5 (17%) and 7 (24%) speakers, respectively in the first and second repetition of *Doda Wanię*. Needless to say, similar variability was observed in the duration of the other measured segments.

4. CONCLUSION

The study showed, with some reservations connected with natural language variability, that the tendency to lengthen the word initial fricative consonant is quite strong (25% difference) in native Polish. This adds to the evidence provided for English, French. Korean and Taiwanese by studies discussed in the introduction. Additionally, initial lengthening also affects the following vowel, which becomes 7.5% longer. The pre-boundary syllable segments, /d/ and /a/ in the tested word *Doda* are only slightly longer than in the no-boundary context, which suggests that in the tested

sentence the word does not constitute a domain for final lengthening.

Considering the above observations, we reject the hypothesis that word boundaries are not signalled in fluent Polish speech. Although some speakers do not give durational cues to lexical segmentation, the tendency to lengthen word-initial consonants (and syllables) appears a regular phenomenon.

5. ACKNOWLEDGEMENTS

This work has been partly supported by a grant from Polish Ministry of Science and Higher Education (0576/B/H03/2010/38) to both authors.

6. REFERENCES

[1] Beckman, M. E., Edwards, J. 1990. Lengthenings and shortenings and the nature of prosodic constituency. In: Kingston, J., Beckman, M. E. (eds.). *Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech*. Cambridge University Press, 152-178.

[2] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International* 10, 341-345.

[3] Bombien, L., Mooshammer, C., Hoole, P., Rathcke, T., Kuhnert, B. 2007. Articulatory strengthening in initial German /kl/ under prosodic variation. *Proceedings of the 16th International Congress of Phonetic Sciences.* Saarbrücken, Germany, 457-460.

[4] Byrd, D. 2000. Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica* 57, 3–16.

[5] Byrd D., Saltzman, E. 1998. Intragestural dynamics of multiple phrasal boundaries. *Journal of Phonetics* 26, 173–199.

[6] Byrd, D., Krivokapić, J., Lee, S. 2006. How far, how long: On the temporal scope of prosodic boundary effects. *Journal of the Acoustical Society of America* 120, 1589–1599.

[7] Cho, T. 2002. *The effects of prosody on articulation in English.* New York: Routledge.

[8] Cho, T. 2004. Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics* 32, 141-176.

[9] Cho, T. 2006. Manifestation of prosodic structure in articulation: Evidence from lip kinematics in English. In: Goldstein, L.M., Whalen, D.H., Best, C.T. (eds.) *Laboratory Phonology,Vol.8: Varieties of Phonological Competence.* Berlin/NewYork: Mouton de Gruyter, 519–548.

[10] Cho, T., Jun, S. 2000. Domain-initial strengthening as featural enhancement: Aerodynamic evidence from Korean. *Chicago Linguistics Society* 36, 31-44.

[11] Cho, T., Keating, P. 2001. Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics* 29, 155-190.

[12] Cho, T., Keating, P. 2009. Effects of initial position versus prominence in English. *Journal of Phonetics* 37, 466-485.

[13] Cho, T., McQueen, J. M., Cox, E. A. 2007. Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics* 35, 210-243.

[14] Cooper, A. M. 1991. Laryngeal and oral gestures in English /p, t, k/. *Proceedings of the XIIth International Congress of Phonetic Sciences, Aix-en-Provence*, volume 2, 50-53.

[15] Edwards, J. E., Beckman, M. E., Fletcher, J. 1991. The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America* 89, 369–382.

[16] Fougeron, C. 2001. Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics* 29, 109-136.

[17] Fougeron, C., Keating, P. A. 1997. Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America* 101 (6), 3728-3740.

[18] Gussenhoven, C., Rietveld, A. 1992. Intonation contours, prosodic structure and preboundary lengthening. *Journal of Phonetics* 20, 283-303.

[19] Keating, P. A., Cho, T., Fougeron, C., Hsu, C.-S. 2003. Domain-initial articulatory strengthening in four languages. In: Local, J., Ogden, R., Temple, R. (eds.) *Papers in Laboratory Phonology* VI. Cambridge: Cambridge University Press, 145-163.

[20] Keating, P. A., Wright, R., Zhang, J. 1999. Word-level asymmetries in consonant articulation. *UCLA Working Papers in Phonetics* 97, 157-173.

[21] Klatt, D. 1975. Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics* 3, 129–140.

[22] Oller, D. K. 1973. The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America* 54, 1235-1247.

[23] Rojczyk, A. 2009. Brat Adama vs. brata dama: Temporal phonetic parameters signalling word boundaries in Polish. *Linguistica Silesiana* 30, 59-73.

[24] Shatzman, K. B., McQueen, J. M. 2006. Segment duration as a cue to word boundaries in spoken-word recognition. *Perception and Psychophysic*s 68, 1-16.

[25] Turk, A., Nakai, S., Sugahara, M. 2006. Acoustic segment durations in prosodic research: A practical guide. In: Sudhoff, S., Lenertov, D., Meyer, R., Pappert, S., Augurzky P., Mleinek, I., Richter, N., Schliesser, J. (eds.) *Methods in Empirical Prosody Research*. Berlin, New York: De Gruyter, 1-28.

[26] Turk, A., Shattuck-Hufnagel, A. 2000. Word-boundary related duration patterns in English. *Journal of Phonetics* 28, 379-440.

[27] Turk, A., White, L. 1999. Structural influences on accentual lengthening in English. *Journal of Phonetics* 27, 171-206.

[28] White, L. 2002. *English Speech Timing: A Domain and Locus Approach.* Ph.D. dissertation. The University of Edinburgh.

[29] Wightman, C.W., Shattuck-Hufnagel, S., Ostendorf, M., Price, P. J. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91, 1707–1717.

7. APPENDIX

Two sets of sentences used in the experiment. All sentences were printed in one column on one sheet and the targets were not highlighted in the original research sheet.

Set 1: Sequences without a word boundary in a carrier phrase. Targets highlighted in bold.

Lubi dzielenie?
Lubi mnożenie?

Lubi odejmowanie?
Lubi dodawanie?
Lubi dzielenie?
Lubi mnożenie?
Lubi odejmowanie?
Lubi dodawanie?
Lubi dzielenie?

Set 2: Sequences with a word boundary in a carrier phrase. Targets highlighted in bold.

Lubi Doda Saszę?
Lubi Doda Miszę?
Lubi Doda Hanię?
Lubi Doda Wanię?
Lubi Doda Saszę?
Lubi Doda Miszę?
Lubi Doda Hanię?
Lubi Doda Wanię?
Lubi Doda Saszę?

SUBJECTIVE EVALUATION OF THE PHONETIC REPRESENTATION OF SOME NATIONAL AND REGIONAL VARIETIES OF THE ENGLISH LANGUAGE

Elena Shamina

St.-Petersburg State University, Russia
elena_shamina@hotmail.com

ABSTRACT

The article deals with the results of a sociolinguistic experiment where the subjects (native speakers of different national and regional varieties of English) were to define the place of residence, social class, level of education and occupation of other English speakers after listening to their speech samples. The results show that English speakers with no linguistic training are able to do that with a high degree of correctness. Besides, some speech samples can incur in some listeners emotional reactions of various types. This leads to a conclusion that in the English speaking cultures pronunciation continues to be a decisive factor in sociolinguistic placement of communicators.

Keywords: sociolinguistics, dialect, pronunciation, emotional reaction.

1. INTRODUCTION

For some decades now there has been general understanding among sociolinguists that in probably all English-speaking countries there exists a close connection between language and social class. J. C. Wells, for example, claims that association of speech differences with social differences is something well-known to every member of the speech community [Wells, 1982: 15]. He refers to the results of the 1972 National Opinion Polls survey which included the question about the most important factors in being able to tell which class a person is. The informants (a random sample of the British public) placed the factor "The way they speak" higher than all other factors, which shows that speech is regarded as more indicative of social class than occupation, education or income. J. C. Wells believes that by "The way they speak" the respondents mostly meant accents, or phonetic properties of speech.

Since the publication in 1966 of W. Labov's work "The social stratification of English in New York City" [Labov, 1966] there have been numerous studies focusing on the social dimension of accent variation in English. It has been shown that although some national varieties of English may be characterized by a whole range of local dialects (e.g., British English) and others – by very little of regional variation (e.g., New Zealand English), there is still some correlation between the specific pronunciation features and such non-linguistic parameters as socio-economic class, ethnicity, education, occupation, gender, age, etc.

The so called sociolinguistic indexicalities – relationships between certain speech forms and social meanings – can be accessed in different ways and against different theoretical assumptions. The classical work by P. Trudgill [Trudgill, 2000] reflects on the interrelation between geographical and social accents, while a recent study by N. Coupland and H. Bishop [Coupland, Bishop, 2007] aims to capture patterns of attributed social meaning by informants who are asked to produce responses to linguistic labels alone (such as "Afro-Caribbean", "Queen's English", "Black Country" or "French", depicting various English accents). It may be said that the former deals with the objective situation in a language community, while the latter is concerned with the subjective reaction of English speakers to their own ideological beliefs about English accents with respect to prestige and social attractiveness.

It can be concluded at this stage that phonetic variation in English-speaking cultures does have a sociolinguistic value that native speakers are very well aware of. They are even ready to ascribe certain social properties to non-indigenous English accents, as in I. Abramova's study [Абрамова, 2009] where Russian speakers of English with strong foreign accents were supposed by the British informants to have different personal traits and professional careers from those who spoke English with a weak or almost non-existent foreign accent.

2. CURRENT STUDY

The experiment the article focuses on was designed to establish if native English speakers with no linguistic training are actually able to recognize various English accents and if they attribute cer-

tain social properties to speech samples with any degree of accuracy. It may be said to follow the tradition of subjective evaluation of sociolinguistic data.

2.1. Corpus

The speech samples were taken from the free International Dialects of English Archive website and modified to make them from 11 to 15 seconds long. The recorded texts contained no personal information or indication of the speakers' social status (except the gender or probably age inherent in the voice properties). The 15 samples chosen for the study were representative of the following national varieties and regional dialects:
- British accents: Received Pronunciation (RP), Estuary English (2 varieties: middle class and working class), South-West (Wiltshire), North-East (so called Geordie, Newcastle), Yorkshire (Leeds), Cockney (North London), Scottish (Edinburgh), Northern Irish (Belfast), Welsh (Cardiff),
- American accents (the USA): New England (New Jersey), Southern (Georgia),
- General Canadian,
- General Australian,
- New Zealand.

2.2. Subjects

The 50 respondents in the survey were college-educated native speakers of English (28 men and 22 women) from the UK (the total of 20 people with 5 residents of each of the regions: England, Northern Ireland, Wales and Scotland), the USA (10 respondents), Canada (10 respondents) and Australia (10 respondents) in the age span from 21 to 60 (the majority being 21 to 35 years old). They were contacted directly either in personal interaction (face to face or by Skype) or by e-mail. They were asked to listen to the speech samples and answer some questions defining, among others, such parameters as the place of residence, social class, education and occupation of the speakers.

2.3. Results and discussion

In complete agreement with what J. C. Wells says about the accuracy / inaccuracy of the perceptions of accents other than those of one's own locality ("The closer we get to home, the more refined are our perceptions) [Wells, 1982: 33], the respondents guessed the speakers' place of residence with a varying degree of correctness. The accents of residents of England were described as "English" or "British" by the whole body of informants (98% of responses), although it is not quite clear if all the respondents differentiated between the two, but the respondents from the UK themselves were much more precise in their placement of the accents. For example, the accent of the Yorkshire speaker was defined as "Northern English" by 20% of English and Scottish respondents and 60% of Welsh respondents, and as "Yorkshire" by 80% of English and Scottish speakers and 40% of Welsh and Irish speakers. Similarly, accents of the American speakers were easily recognized as such (in 98% of cases) by all the participants of the experiment, but about 30% of respondents from the USA labeled the speech of one American speaker as "New Jersey" and were right in doing that. All the local accents were correctly recognized by the people from the corresponding region (in 98 - 100% of the cases).

Overall figures of correct recognition of accents are very high for English and American varieties (about 98%) and rather high for Irish, Australian and New Zealand varieties (65%, 65% and 60% respectively). Accents from Canada and Wales turned out to be the least recognizable (with 40% and 35% of correct recognition respectively) with the former being mostly labeled by a wider term "American". The reported accurate geographical placement of speakers only on the grounds of their pronunciation seems even more surprising if the fact is taken into consideration that in some instances the respondents actually experienced difficulties in understanding the speech or even thought they were hearing foreign speech (in 7 cases). The most difficult accents to understand in the present survey were Cockney (with 12% of informants having difficulties), Scottish English (16%), Estuary English in its lower class variety (22%), Northern Irish English (24%) and Yorkshire English (40%). The fact may serve as another reminder of great phonetic differences between the varieties of English.

As to social class, this complex notion closely connected with education, occupation and other social and economic features was attributed to the speech samples in the experiment not as easily as regional characteristics. The respondents' judgments were accurate for the higher class in 40% of cases, for the middle class in about 70% of cases and for the lower class in 56% of cases. The perception of the social dimension of the speech samples may be influenced by the respondents' own class: the majority belonged to middle class and identified the familiar variety more readily than other social dialects. Besides, in this age of political correctness the idea of universal equality predominates over all other social attitudes and the respondents may have

felt restricted in their judgments, especially in direct contact with the researcher. What is of special interest here, though, is that no informant refused to answer the question about the class: presumably, they all were sure they were able to do that.

In these circumstances the question about the speakers' occupation may be considered as much more differentiating and revealing. As the informants in the survey were not offered any suggested responses to the questions set to them, their answers about the speakers' occupation were most varied. Besides, in some instances they chose not to give any answers at all. On the other hand, there was a significant agreement between the respondents in attributing to the speakers similar occupations and positions in accordance with the attributed social class. For example, a speaker of working class, who was a plumber, was ascribed such occupations as "plumber" (surprisingly, by 46% of respondents), "carpet fitter", "carpenter", "manual worker", "unemployed", "in service", "industrial worker", and another speaker, a well-educated professional singer of higher middle class, was described as "banker", "actor", "doctor", "businessman", "politician", "lawyer", "executive manager", "country estate", "company owner", "MP", "professor", "pensioner" and "intellectual", although nobody guessed what he actually did.

The informants' answers to the question about their emotional reaction to the speech samples were also supposedly influenced by the political correctness notion and practices. As D. Crystal states in one of his works [Crystal, 2004: 249-250], complaint about the phonetic change and, as a consequence, "incorrect" pronunciation has been going on in the British culture since the middle of the 16th century. In view of that, more of negative reactions to the pronunciation different from that of the informants could have been expected, but in fact in 34 % of the cases they had a positive reaction to what they were hearing and only in 12 % of the cases the respondents confessed to a negative feeling towards the speech samples.

It may be worthy of notice that the most "likable" accents in the survey were those of Scotland, Australia, the USA (New Jersey) and New Zealand, while Estuary English in its lower class variety got most (about 50%) of the negative responses. The RP still seems to enjoy its traditional high standing: about 50% of the respondents reacted to it positively, although some British informants demonstrated negative attitude to it.

3. CONCLUSION

The question about the personal attitude to some language phenomena is traditional for sociolinguistic studies and may be put to respondents in different ways: some researchers concentrate on a kind of objective evaluation, for example, in terms of "social attractiveness" [Coupland, Bishop, 2007], others appeal to subjective personal reaction, like "friendship" or "fight" [Labov, 1972]. For the present study the latter approach was chosen: it seems really important that phonetic properties of speech can incur in the listeners emotional response (in slightly less than a half of the cases in the survey) and thus make the perception of the sociolinguistic indexicalities even more relevant in the English language discourse. The observation correlates with the data gathered in the study of pronunciation of English literary characters [Шамина, 2010].

The findings reported in the article seem to prove that pronunciation has great evaluative significance for the language-ideological patterns existing in the English-speaking cultures.

4. ACKNOWLEDGEMENTS

The author is most grateful to Elena Luzina for her help in conducting the survey and collecting the data.

5. REFERENCES

[1] Coupland N., Bishop H. 2007. Ideologised values for British accents. *Journal of Linguistics*, 11 (1), 74-93.

[2] Crystal, D. 2004. *The stories of English*. London, Penguin Books.

[3] Labov, W. 1966. *The social stratification of English in New York City*. Washington, D.C.: Center of Applied Linguistics.

[4] Labov, W. 1972. The Study of Language in its Social *Context Studium Generale.* 23 (1).

[5] Trudgill, P. 2000. *Sociolinguistics: an Introduction to language and society*. 4-th ed. London: Penguin Books.

[6] Wells, J. C. 1982. *Accents of English*. vol. I. Cambridge: Cambridge University Press.

[7] Абрамова, И. Е. 2009. *Фонетическая вариативность вне естественной языковой среды*. Петрозаводск: Изд-во ПетрГУ.

[8] Шамина, Е. А. 2010. Произносительные характеристики персонажей англоязычной литературы. Материалы XXL Международной филологической конференции. СПб: Изд-во СПбГУ, (In press).

PERCEPTUAL CHARACTERISTICS OF ENGLISH VOWELS

Olga Morozova, Natalya Shelevaya

The Amur State University, Russia
natalya-shelevaya@mail.ru

ABSTRACT

The paper focuses on the perception of Canadian stressed vowels. The purpose of the perception experiments is to reveal how the perception of a vowel depends on phonetic context and how much perceptual and acoustic characteristics interact. Another aim was to define the difference between the perception of isolated words and words segmented from spontaneous speech.

Keywords: vowel, coarticulation, perception.

1. INTRODUCTION

Speech perception is a very complicated process which involves different fields of science (linguistics, psychology, physiology, etc.). The research of speech perception is a study of what operations are done to go to a certain symbolic form from acoustic speech signal [Венцов, Касевич, 1994]. Studying perceptual phonetics linguists put the following problems: 1. What acoustic characteristics of speech sounds does a native speaker use to decide that the given sound is a realization of a certain phoneme? 2. What acoustic features are used by native speakers to define the amount of phonemes in a certain sound sequence? 3. How does a native speaker cope with extremely great variability of acoustic characteristics and filter only those ones necessary for the message identification? 4. How are the sounds of an unknown language perceived? 5. What is the volume of acoustic information necessary for a person to identify the meaning of the message [Бондарко, 1998]?

The purpose of the research is to reveal the perceptual peculiarities of English vowels. The first task was to figure out how the perception of a vowel depends on phonetic context and whether perceptual and acoustic characteristics of the given vowels coincide. The second task was to define if there is conformity between the perception of isolated words and the words segmented from connected speech.

METHOD

The material of the experiment consisted of 27 stimuli produced by Canadian speakers. These stimuli were divided into two blocks. The first block contained isolated words.

The second series consisted of vowel stimuli segmented from spontaneous speech.

The record of the given series was presented to 7 speakers of British English (BE), American English (AE) and Canadian English (CE). The stimuli of the words contained open vowels /æ, ai, ai, aʊ, ɒ/ and the front close-mid vowel /e/. The choice of the given Canadian vowels was conditioned by their wide acoustic variety revealed while carrying out auditory and instrumental analysis.

During the experiment the native speakers were offered to listen every word two times and write them down orthographically. They were notified that they would probably be unable to recognize the word. In this case they were asked to mark the stimulus according to the rules of English spelling.

In the course of the experiment the listeners marked the stimuli: firstly, by the word read by the speaker; secondly, by the word with correctly recognized vowel, but in the other consonant context; thirdly, for marking the stimulus they offered the word with another vowel, sometimes in the other consonant context.

It wasn't difficult to interpret the orthographic version of the stimuli, because the cases of ambiguous interpretation of the written word occurred very seldom. While processing the results of the experiment we examined only the perception of the vowel without considering right or wrong recognition of the words. The mistakes in the consonant context were not taken into account.

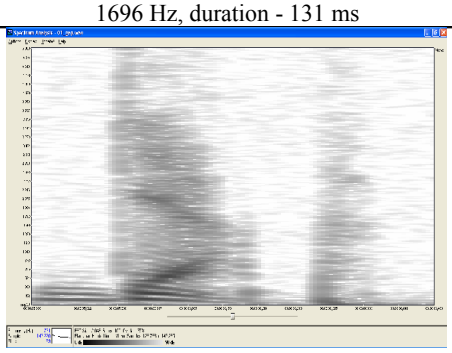
2. RESULTS AND DISCUSSION

The acoustic and perceptual characteristics of the front open vowel /æ/ are shown on Figure 1.

Figure 1: Acoustic and perceptual characteristics of the monophthong /æ/ (isolated words)

Acoustic characteristics

Figure 1a. The sonogram of the stimulus № 1[ˈgˈæp] from the word *gap*, formant values FI-FII = 715 – 1696 Hz, duration - 131 ms



Perceptual characteristics

Figure 1b. The identification of the stimulus № 1 [ˈgˈæp] from the word *gap* as [e] – 71 %, as [æ] – 29 %

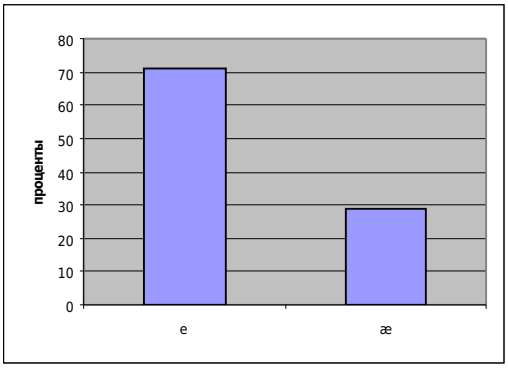
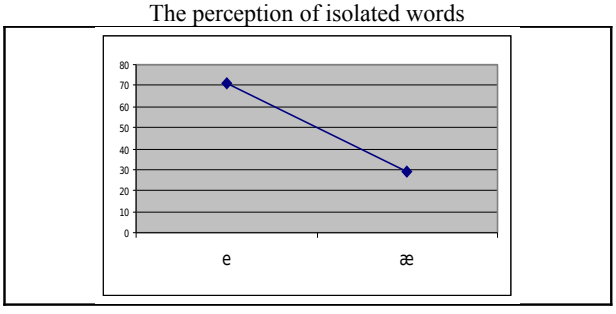
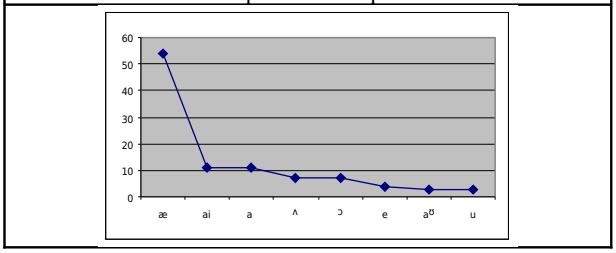


Figure 2: Perceptual characteristics of the phoneme /æ/ (on the whole experiment)

The perception of isolated words



The perception of the words segmented from spontaneous speech



While analyzing perceptual reactions of stimulus №1 [ˈgˈæp] from the separately pronounced word *gap* it turned out that the vowel was perceived as quite close. 71% of the listeners identified it as front mid-open monophthong [e], 29% of the listeners recognized it as front open vowel [æ].

Uncertain perception of the monophthong and its substitute for [e] (in 5 cases out of 7) perhaps was an attempt to recognize the more frequent word *get*. On the other hand it is common knowledge that during the perception native speakers frequently offer the variant perceptually close to the stimulus. The given fact is proved by foreign researchers who also note the mixing of perceptual reactions of the vowels of neighboring tongue heights and frontings of the vocalic trapezoid, in this case [æ] for [e].

The vowel of the given stimulus is also interesting for its acoustic characteristics, that is, glide component which occurs as a result of coarticulation effect of the occlusive voiced backlingual consonant on the front vowel. You can see the spectrum of the vowel on Figure 1a. Occlusive backlingual consonant conditions steep trajectories of movement of FI up and FII down to quasi-stationary part of the vowel, formant value of which is 715 – 1696 Hz, [I]–like element in the spectrum of the vowel has the value of 465 – 2089 Hz.

We would like to note the informativity of ɪ-like transition in the spectrum of the vowel, which is a peculiar marker in coarticulation of backlingual consonants and front vowels. Thus, while presenting the second series of the experiment which contained the stimuli of the syllables segmented from prosodically marked parts of spontaneous speech, it turned out that the listeners are able to restore the cut occlusive backlingual consonant by the first transition of ɪ-like part of the vowel. For example, in the position after occlusive backlingual consonant /g/ (table 2, stimulus №1) monophthong /e/ was correctly recognized in 5 cases out of 7. What's more in the given 5 cases the listeners three times noted the diphthongal realization of [ˈɛ] which was marked during the acoustic analysis. For graphical marking of the diphthongal realization of [ˈɛ] the listeners offered the combination of the words – ye –. So we can conclude that some listeners note well enough [ɪ]-like faint concurrent sound in the beginning of the vowel. Even in the case of the wrong perception of /e/ in the given position the listeners noted [ɪ]-like

beginning of the vowel. The described case conform to the results received on the material of the Russian language [Bondarko L.V., 1998, 136]. Thus in the position "occlusive backlingual + front vowel" the transition part of the vowel is informative enough (we mean the difference between backlingual consonant and front vowel according to the place of obstruction in which in the process of coarticulation influence we observe a slight softening of the preceding backlingual consonant and [ɪ]-like beginning of the vowel), that's why the listeners don't have difficulties in its marking.

The recognition of the phoneme /æ/ by the listeners was more than 50% in the whole experiment (Figure 2). The given monophthong in the position after labial consonants was recognized in the following way. In the combination "labio-dental fricative consonant + vowel /æ/" in the word *fat* the monophthong was recognized correctly by all the listeners (table 2, stimulus 4). In the combination "plosive voiceless labial consonant + /æ/" in the word *pan* the vowel was recognized in 6 cases out of 7. Only in one case the vowel was identified as /e/ (Table 2, stimulus 5).

In the other cases the identification of the given vowel wasn't so successful. In the case if a voiceless affricative /tʃ/ followed the combination "plosive voiced labial consonant + vowel /æ/" (Table 2, stimulus 2), the perception of the vowel was greatly obstructed. The vowel /æ/ was recognized as a central open monophthong [a] by 43% of the listeners, as a front open monophthong [æ] by 29% of the listeners, as a central mid-open monophthong of a broad variation [ʌ] also by 29% of the listeners. However, if we generalize the received reactions it becomes clear that the listeners hear a-like vowel in the given stimulus, what is explained by the acoustic characteristics of the monophthong. The formant characteristics of the vowel segment correspond to quite open central monophthong rather than front one, and that probably caused such perception of the vowel.

The monophthong /æ/ was recognized considerably worse in the preposition to the velar /l/. In this position the monophthong was marked by the vowels /ai/, /aʊ/, /ɔ/, /u/. Such perception of the monophthong can be explained by the influence of the velar /l/ following after the vowel, which provoked identification of /æ/ by the native speakers as a not front vowel.

Comparing the perceptual characteristics of the front open phoneme /æ/ from the stimuli of isolated words and the stimuli segmented from spontaneous speech (Figure 2) it is necessary to note the following. Perceiving isolated words the listeners offer perceptually and acoustically close variants /æ/- /e/.While perceiving the stimuli from spontaneous speech native speakers offered up to 8 variants, though in most cases the perception of the vowel was certain enough (57%).

As far as the front mid-open monophthong /e/ is concerned, some foreign researchers note the change of its quality in Canadian English. In their opinion this monophthong acquires the characteristics of the front-retracted vowel in the speech of Canadians. This fact isn't proved by the results of our acoustic experiment. On the whole in the speech of three Canadian speakers the vowel /e/ in the words like *pen, ten, bed* has the qualities of the front mid-open vowel of a broad variation like in British English and American English (Table 3).

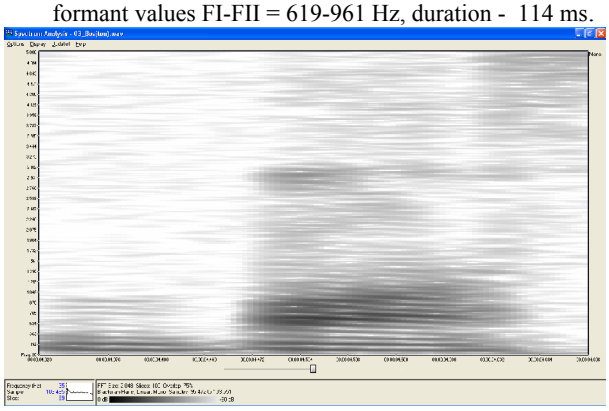
Table 3: Average meaning of FI and FII of the monophthongs in CE, GA, RP

phoneme	CE		GA		RP	
	FI	FII	FI	FII	FI	FII
e	544	1497	536	1591	559	1563

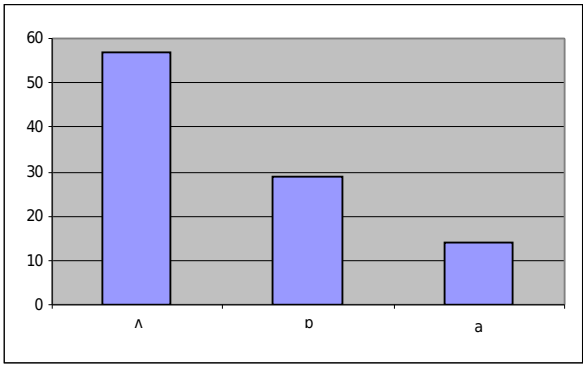
Speaking about the perceptual characteristics of the given monophthong the auditory experiment shows that this monophthong /e/ is recognized with enough confidence. While perceiving the isolated word *ten*, the monophthong was correctly identified by all the listeners (table 1, stimulus 10). Perceiving the syllables segmented from spontaneous speech the listeners were not so unanimous in their answers (Table 2, stimulus 6 [**sev]** from the word seven and stimulus 7 [**pep]** from the word pepper). If in the position after the front vowel in the stimulus [**sev]** the vowel was interpreted as the monophthong and the front mid-open diphthong: /e/ and /ei/ respectively, then in the vicinity of voiceless plosive labial consonants for marking the vowel the listeners offered the variants of both front mid-open monophthong /e/ and front open monophthong /æ/. Meanwhile, most of the listeners interpreted the vowel as more open. In this case we deal with a mixture of perceptual reactions of the vowels of neighboring tongue heights and frontings of the vocalic trapezoid, in this case [e] for [æ].

Figure 3:Acoustic and perceptual characteristics of the monophthong /ɒ/

a. Acoustic characteristics. The sonogram of the stimulus № 3 [bɒs] from the word *Boston*, formant values FI-FII = 619-961 Hz, duration - 114 ms.



b. Perceptual characteristics.
The identification of the stimulus № 3 [bɒs] from the word Boston as [ʌ] – 57 %, as [ɒ] – 29 %, as[a] – 14%.



If we analyze the received reactions of the listeners during the perceiving of the phoneme /ɒ/ in the given positions, the results will be the following. We can see on Figures 3a,b and 4a,b that the vowel in the words pronounced in isolation (the words like *Boston, sock*) will likely be identified with the back open monophthong /ɒ/.

Figure 4a. The sonogram of the stimulus № 11 [sɒk] from the word *sock*, formant values FI-FII = 650-1171 Hz, duration - 80 ms.

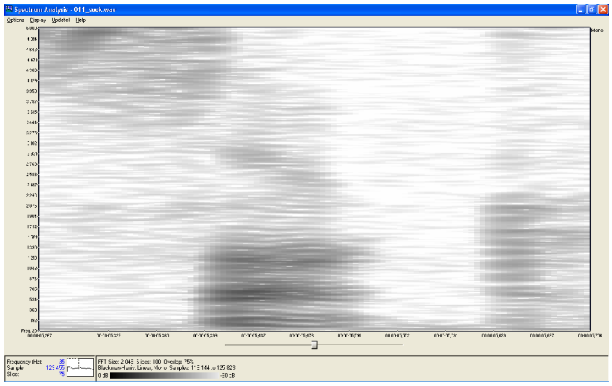


Figure 4b: The identification of the stimulus № 11 [sɒk] from the word *sock* as [ɒ] – 72 %, as [ʌ] – 14 %, as [æ] – 14%.

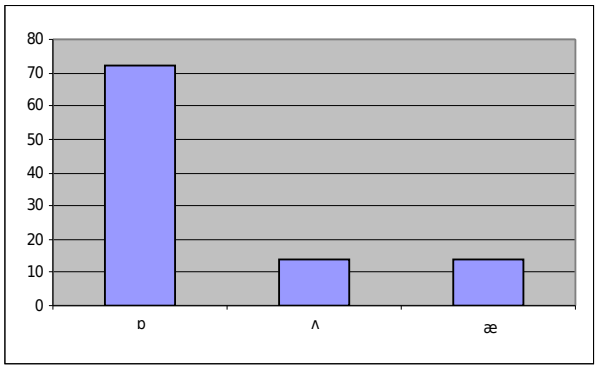


Figure 5a: The sonogram of the stimulus № 8 [sɒs] from the word *sausage*, formant values FI-FII = 619-1187 Hz, duration - 125 ms.

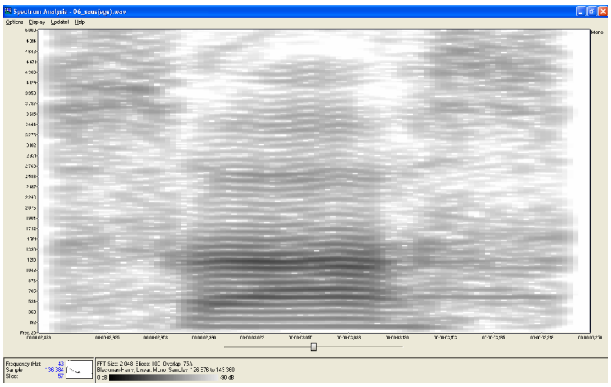
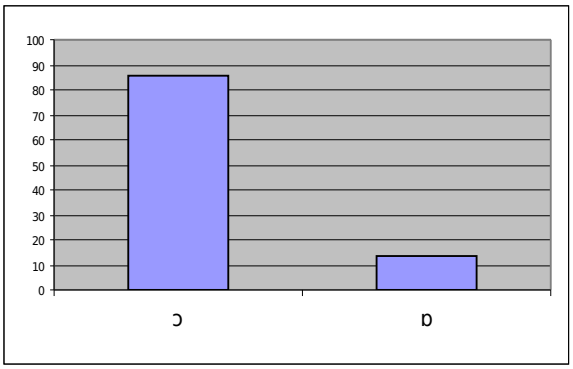


Figure 5b: The identification of the stimulus № 8 [sɒs] from the word *sausage* as [ɔ] – 86 %, as[ɒ] – 14 %.



The given perceptual evaluation amounted to 50% of reactions. In 36% of cases the listeners identified the phoneme /ɒ/ in the analyzed words (surrounded by labial and forelingual context) as the central mid-open monophthong of a broad variation [ʌ]. The interpretation of the vowel both as a central open monophthong [a] and a front central monophthong [æ] amounts to 7%.

Figure 6a: The sonogram of the stimulus № 9 [hɒtp] from the word combination *hot pan*, formant values FI-FII = 634-117 Hz, duration - 74 ms.

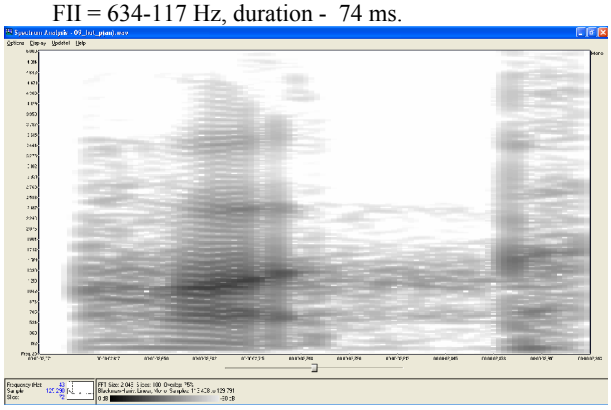
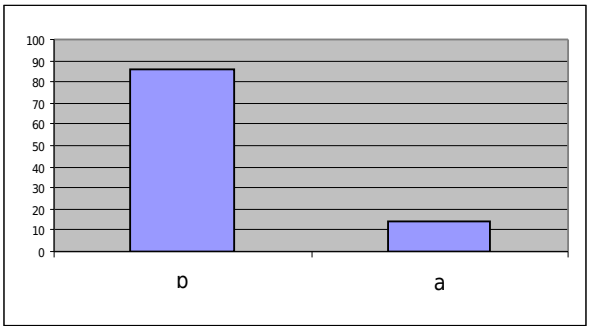
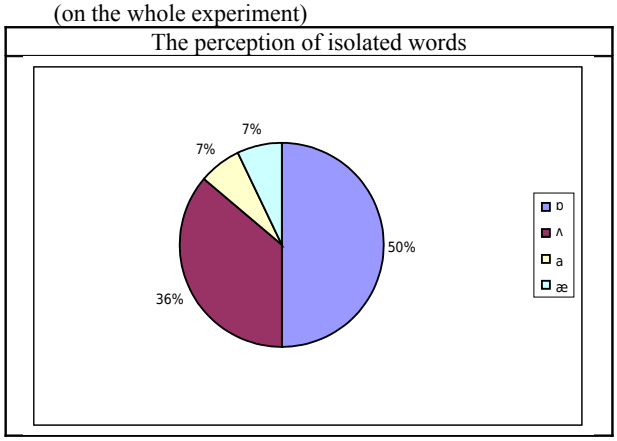


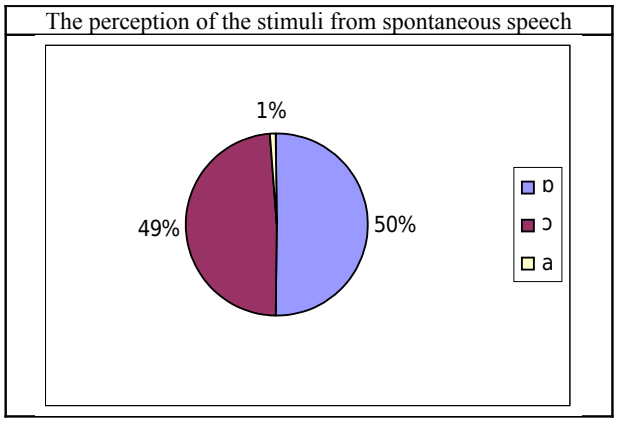
Figure 6b: The identification of the stimulus № 9 [hɒtp] from the word combination *hot pan* as [ɒ] – 86 %, as [a] – 14 %.



It is interesting to note that while presenting the words pronounced in isolation we deal with more varied perceptional reactions of the given vowel (Figure 7), what is in our opinion a pleasant exception proving the rule.

Figure 7:Perceptual characteristics of the phoneme /ɒ/ (on the whole experiment)





If we compare perceptual characteristics of the diphthongs /ai/ and /aʊ/ and the analyzed monophthongs from the isolated words and the words segmented from spontaneous speech, it appears that native speakers have difficulties and therefore show more differences in interpretation of the segments-stimuli from spontaneous speech.

While examining perceptual reactions of complex vowels it is necessary to note that the diphthong /ai/ was recognized by the listeners quite successfully (Table 1, stimuli 2, 4, 5, 7, and table 2, stimuli 10, 11, 12, 13). We can see absolutely different picture while analyzing perceptual characteristics of the diphthong /aʊ/ (Table 1, stimuli 6, 8, table 2, stimuli 14, 15, 16, 17). As it is known the diphthong /aʊ/ in the system of Canadian English vowels is notable for broad variation during realization in spontaneous speech. According to the position the given diphthong is realized as allophones [aʊ], [a̟ʊ], [ʌʊ], [ʌ̟ʊ], and even [oʊ]. According to the data of Chambers and Hardwick, the allophone [oʊ] is more frequent in a

weak phrasal position in the words *out, about*. In the course of our experiment native English speakers evaluated the stimuli containing the given diphthong. They offered the following ways of marking: /aʊ/, /ɔʊ/, /ɔ/, /ɒ/, /ʌ/, /u/. Meanwhile, while perceiving the diphthong in the stimuli from spontaneous speech the listeners couldn't correctly identify the vowel in all the cases.

4. CONCLUSION

Thus, the results of the auditory experiments are the following.

- It is necessary to note the mixture of perceptual reactions of the vowels of neighboring tongue heights and frontings of the vocalic trapezoid.
- We should draw a conclusion about the direct dependence of a word perception success and the frequency of its use. The more frequent is the word the more confident is its perception by the listeners. And vice versa, less frequent words are often substituted by the listeners for more frequent ones.
- The importance of syllable contrasts for the perception by native speakers is proved. In particular, the listener successfully restores the preceding consonant by its place of obstruction while perceiving the information contained in the transition parts.
- And finally, while presenting the stimuli from spontaneous speech we noted a great variety of perceptual reactions in comparison to the reactions of the listeners to the stimuli from the isolated words.

5. REFERENCES

[1] Бондарко, Л. В. 1998. *Фонетика современного русского языка.* СПб: СПбГУ.

[2] Чугаева, Т. Н. 2007. *Перцептивный аспект звукового строя английского языка*. Пермь.

[3] Венцов А. В., Касевич В. Б. 1994. *Проблемы восприятия речи.* СПб: СПбГУ.

Table 1: The matrix of the listeners' answers (perceived from the isolated words).

№	Word	A1 (BE)	A2 (BE)	A3 (BE)	A4 (AE)	A5 (AE)	A6 (AE)	A7 (CE)
1.	g**a**p	gɐt	gɐt	gæp	gɐt	gɐt	gæp	gɐt
2.	p**ie**	pai	pai	pai	pai	pai	pai	pai
3.	B**o**s(ton)	bʌs	bʌs	bɒs	pas	pɒst	bʌs	bʌs
4.	g**uy**	gai	gai	gai	gai	gai	gai	gai
5.	sl**i**ced	slais	saiz	slaist	slais	slais	slais	slais
6.	(a)b**ou**t	bɔt	bɔt	baʊt	bɔt	bɔt	but	bɔt
7.	k**i**te	skeit	skeit	kait	keit	keit	skeit	eit
8.	h**ou**se	haʊs	haʊs	haʊs	haʊs	haʊs	haʊs	haʊs
9.	s**o**ck	sæk	sɒck	sʌk	sɒck	sɒck	sɒck	sɒck
10.	t**e**n	ten	ten	ten	ˈtempɔʊ	ten	ten	ten

Table 2: The matrix of the listeners' answers (perceived from the segmented words).

№	Word	A1 (BE)	A2 (BE)	A3 (BE)	A4 (AE)	A5 (AE)	A6 (AE)	A7 (CE)
1.	(g)ɛt	nɪɛt	fəˈgɛt	jʌp	jɪp	gɛt	vɪɛt	jɛp
2.	b**a**ch(elor)	dʌʧ	laʤ	bæk	klætʃ	aʧ	bʌʧ	aʧ
3.	s**a**l(ad)	said	said	saʊθ	said	sɔlt	sut	sɔlt
4.	(much) f**a**t	fækt	fækt	fæt	fæt	fæt	fæt	fæt
5.	p**a**n	pæn	pæn	pen	pæn	pæn	pæn	pæn
6.	s**e**v(en)	sevn	sevn	sed	sevn	sevn	seif	seif
7.	pepp(er)	gæp	gæp	pep	kæp	kæp	pep	pet
8.	saus(age)	sɒs	sɔs	sɔs	sɔs	sɔs	sɔs	sɔs
9.	h**o**t (pan)	hɒt	hɒt	pɑk	hɒt	hɒt	hɒt	hɒt
10.	l**i**ke (to)	ˈlʌki	ˈlʌki	lʌik tu	ˈlʌki	lʌik	ˈlʌki	lʌikt
11.	sl**i**ce	lais	lais	pleit	slais	slais	lais	slais
12.	h**i**gh s(chool)	ais	hai	hais	ais	hai	hai	hai
13.	(com)b**i**ned	wai	bai	baind	brain	bein	bai	bain
14.	(a)b**ou**t it1	bɔdɜr	bɔdɜr	bɔt ɪt	bɔrdɜr	əˈbʌv	əˈbʌv	əˈbʌv
15.	(a)b**ou**t it 2	gɒd	lɒt	rʌd	lɒt	lɒt	but	but
16.	(a)b**ou**t 3	lɒt	lɒt	blʌd	lɒt	lɒt	bʌt	əˈbʌv
17.	(a)b**ou**t 4	mɔʊt	ɒp	ˈmɔnɪŋ	ˈmɔʊtɪv	mɔʊ	bʌt	əˈbʌv

TWO-NUCLEI INTONATION PHRASE IN THE RUSSIAN TV PRESENTERS' SPEECH

Ekaterina Shumovskaya

St.-Petersburg State University, Russia
sshumowski@yahoo.com

ABSTRACT

This study is focusing upon one of the specific prosodic patterns found in speech of TV presenters. It provides the analysis of 'two-nuclei' intonation phrases, i.e. phrases with two intonational centres, from the perspective of the intonation phrasing. These phrases are non-final phrases with a fall in the first center and a rise in the second. If in some of the examples the additional falling tone can be interpreted as a way to emphasize the word, in other examples this emphasis is not motivated semantically. Nevertheless, the melodic pattern is the same. That rises the question of phrasing the utterance.

Keywords: intonation phrasing, 'two-nuclei' intonation phrases, phrase boundary signals

1. BASIS FOR THE RESEARCH

The basis for this analysis is the presenters' speech recordings of the Russian TV news programmes. The overall volume of recordings is 18.5 minutes.
This research started with the auditory analysis of the recordings with the aim to create the generalized intonational transcription (identifying tones movement in the main stressed syllable of the intonational centre, the prevalence of the 'psychological' and physical pauses, and of the logical and/or emphatic accent. The choice over the phrases of two intonational centres was made on the basis of the transcription. Out of 700 phrases having been analyzed, the 'two-nuclei' phrases constituted 12.3%.

2. TERMINOLOGY AND THEORETICAL FOUNDATION

'Two-nuclei' phrases are non-final phrases which seem to have two nuclei: the first one is characterized by a falling tone and the second one, by the rising tone.
If some of the examples demonstrate the additional falling tone movement which can be interpreted as a means of emphasizing a particular word ("*poli**ti**cheskij" sezon* – '*political season*' - because prior to that details of the military campaign were provided; *zajavljal o svoih namerenijah* …- '*declared his intentions'* - as it was **already** had been done etc. – see Figures 1, 2); other examples do not demonstrate the necessity of such stressing from the semantic perspective of the text (…*v etom dages**ta**nskom gorode*… - *'in this Dagestan city'*, … *p**re**zidentu Putinu*… - *'to President Putin'*).

Figure 1: *Movement of the F0 in the phrase ' Novij poli**ti**cheksij sez**on**…'*

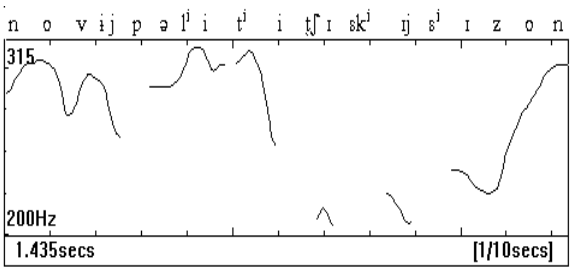
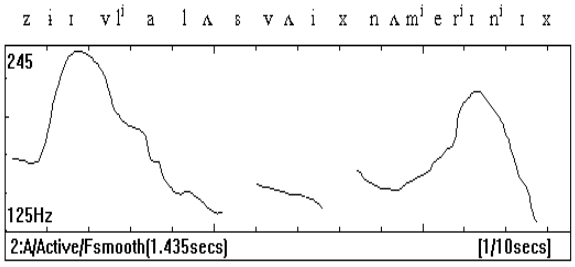


Figure 2: *Movement of the F0 in the phrase '…zajavlj**a**l o svoih na**me**renijah…'*



Nevertheless, the melodic pattern used by the presenters is the same in both cases.
The spoken speech provides instances of the changes in the 'classical' structure of the phrase. It was identified by N. D. Svetozarova [Svetozarova, 1982: 106] that in the 'classical' structure "there are instances of two equally strong phrase stresses which carry different meanings, i.e. the emphatic and the neutral one; these cases are viewed as inherent to the natural spontaneous speech acts" [Bondarko et al., 1981: 164].
Just like the separation of the last word in a phrase, which, according to T. Nikolajeva [Nikolajeva,

1997] is "the result of the special intention of the speaker towards the perception of the listener, thus stimulating his attention", 'two-nuclei' phrases are designed to maintain contact with the listeners, attracting and keeping their attention. It is remarkable that, as it was demonstrated in supporting analysis, the additional intonation centre (the second nucleus) occurs not only as a result of semantic requirements, but also without any of them. The latter cases provide no foundations for these stresses' functions to be classified as different, i.e., as an emphatic and a neutral phrase accent.

3. PROBLEM OF THE INTONATION PHRASING

Within the perspective of the described intonation pattern, there arises a question of the principles underlying the phrasing of the utterance. If that is done on the basis of the phonetic criterion – that is, the intonational unity of a phrase, – then we have to consider the intonation phrases with two centers as containing *two* intonation phrases. D. Hirst [Hirst, 1997: 38] suggested that the number of *phonological phrases* in an utterance is defined by the number of intonational centres, i.e. there is one and only one intonational centre in a phrase. If, however, the phrasing is done on the basis of the semantic criterion, then, against the principle of intonational unity of a phrase, one has to conclude that, paradoxically, the existence of two intonational centres does not contradict the definition of an intonation phrase.

It is known that neither semantic nor syntactic criteria are not formalized and allow the significant level of the ambiguity and subjectivity. That, however, can be also ascribed to the phonetic criterion as well. Is it always the case that the presence of the hesitation pause define the phrase boundary? In other words: can there be a pause within a phrase? What is 'unity' of an intonation phrase in the perspective of the melodic contour? If the melodic contour changes sharply (like in the 'two-nuclei' phrases), would that be a sufficient reason to define the phrase boundary, or a sudden change of the melodic contour is possible also *within* a phrase? Answers to some of these questions were provided as a result of the auditory experiment.

4. MATERIALS AND METHODOLOGY OF THE PERCEPTUAL STUDY

4.1. Corpus

21 phrases with two intonational centres were selected as a result of the perceptual and acoustic analysis. The material of the perceptual experiment included 40 phrases: 21 'two-nuclei' and 19 'regular' tokens. The tokens were randomized in order to avoid listeners getting accustomed to the monotonous melodic contour. Each phrase was repeated twice with a 4 seconds pause.

The purpose of the perception experiment was to clarify how such phrases are perceived by the native speakers: whether a sharp change of the melodic contour was of substantial significance, if so, would it be a sufficient phrase boundary signal.

4.2. Subjects and procedure

38 listeners took part in this experiment (29 of them being students of the philology of the first and second year and 9 of them having no affiliation to philology). The task for the listeners was 1. to identify a word or words accentuated by the presenter and 2. to identify the location of a pause (if it is to occur).

4.3. Results and discussion

As it can be seen from the Table 1, in the 19 out of 21 cases of the 'two-nuclei' phrases, the majority of the listeners (71.1% to 94.7%) have pointed out the additional accent (see Table 1, '*two nuclei*'). Unlike it was initially assumed, the listeners did not identify pauses within the 'two-nuclei' phrases (in cases of the sharp changes of the melodic contour) (see Table 1, '*pause position*'). More so, in the phrase '*chto/mirnij dogovor'*, where the presenter did make a physical pause (see Picture 2), its existence was confirmed only by 11 out of 38 listeners.

An additional experiment (14 participants of the key experiment) followed the key experiment. The listeners were asked not only to identify the accented words but also to identify the phrases carrying a specific connotation, which could have been interpreted as 'right here' (something took place) or 'specifically he/she' (who has said something). The following results were acquired: for the phrases NN 2, 6, 10, 12, 13, 19, 20, 21 the number of positive responses was over 70%, for all other phrases this number did not carry any statistical significance (see Table 1, '*contrastive accent'*).

The experiments have demonstrated that the majority of the listeners do hear the additional accent in the majority of the 'two-nuclei' syntagms, but only in eight cases out of them, the majority of the listeners interpreted such phrases as possessing contrastive meaning.

Table 1: Results of the auditory experiment

N	intonation phrase	two nuclei, %	pause position (|)	contrastive accent, %
1	...v etom dagestanskom gorode	68,4	etom |	28,6
2	...vo vremja osetino-ingushskogo konflikta …	94,7		71,4
3	...prezidentu Vladimiru Putinu …	50,0	prezidentu |	21,4
4	...sozdajot pravovuju osnovu …	65,8		35,7
5	...novij politicheskij sezon …	76,3		35,7
6	...osenju devjanosto vtorogo goda	94,7		64,3
7	...s zhurnalistami vstretilis advokati ljudej …	89,5	vstretilis |	42,9
8	...spetzialnja lingvisticheskaja komissija...	78,9		50
9	...u rossijsko-gruzinskoj granitzi …	78,9		50
10	... chto | mirnij dogovor	76,3	chto |	78,6
11	...chto chast paketa aktzij…	89,5	chast | chto |	57,1
12	... sovmestnogo hozjastvennogo osvojenija …	71,1		71,4
13	...ministr pechati Michail Lesin	81,6	pechati |	71,4
14	..segodnja rukovodtsvo Chechni…	73,7	segodnja|	28,6
15	... zakluchit mirnij dogovor …	71,1	zakluchit |	14,3
16	...zajavljal o svoih namerenijah …	86,8		42,9
17	...chto sovetnika prezidenta что …	73,7	chto | sovetnika |	35,7
18	...provesti rjad terroresticheskih aktov …	73,7		21,4
19	..o segodnjashnih sobitijah v Rjazani …	84,2	sobitijah |	71,4
20	...segodnja v Moskve s zhurnalistami …	86,8	Moskve |	78,6
21	...eto mestnij konflikt	78,9	eto |	78,6

5. CONCLUSIONS

To summarize: the same melodic contour of a 'two-nuclei' phrase can carry different meanings for the listener. In some cases, it means a contrastive accent and it corresponds the additional meaning; in that case the presenter's choice of this particular contour is semantically justified. In the majority of cases, however, the additional accent is perceived as one of the possible melodic contours of the pre-nuclear part of a phrase, carrying no semantic specification, and – as it was pointed by the listeners – characteristic for this style of speech.
Sometimes it is difficult to identify the criteria which defined the listeners' choice and decisions. It is apparent, however, that in that case the most important was the semantic one.

6. REFERENCES

[1] Gordina M.V. 1997. *Phonetika franzuzskogo jasika*. St. Petersburg.
[2] Nikolajeva, T. M. 1997. *Frazovaja intonatzija slavjanskih jazikov*. Moscow.
[3] Svetozarova N.D. 1982. *Intonatzionnaja sistema russkogo jazika*. Leningrad.
[4] Bondarko L. V., Verbitskaya L. A., Zinder L. R. et al. 1988. *Fonetika spontannoj rechi*. Moscow.
[5] Hirst, D. 1997. *Intonative Features: A Syntactic Approach to English Intonation*. Mouton Publishers. P.38

RUSSIAN INTONATION: NEW ASPECTS OF ACOUSTIC AND PERCEPTUAL FEATURES INTERACTION

Pavel Skrelin

St.-Petersburg State University, Russia
skrelin@phonetics.pu.ru

ABSTRACT

In 2010 the Department of Phonetics completed a project dealing with automatic interpretation of sentence prosody . The results obtained indicate that the data we have on the relationship between acoustic and functional properties of the intonation units is insufficient, which affects the prediction of the intonation model based on the automatic analysis of the acoustic data. In this respect - slope of F0 changes and ways of its calculation - acoustic properties of high level tone (perceived as "rising" even when presented without a context) deserve special attention. Of great interest here is a considerable mismatch between acoustic (rising or falling tone) and perceived properties of the intonation units, suggesting "an analogy decision" in favour of their functional properties (question, enumeration, etc.).

Keywords: intonation, prosody, acoustics, perception.

1. INTRODUCTION

In 2010 the Department of Phonetics completed a project dealing with automatic interpretation of sentence prosody. The experimental material had been recorded earlier for the project “High Quality Russian Speech Synthesis using Unit Selection Method” and is known as CORPRES – a fully annotated COrpus of Russian Professionally REad Speech [Skrelin et al., 2010]. The corpus meets all the requirements to the databases of this kind and may be used both for the purposes of development and scientific research. It is large enough for statistical machine learning (60 hours of continuous speech) and has six annotation levels including prosodic annotation, rule-based canonical phonetic transcription and a manual transcription reflecting the actual sounds pronounced by the speakers.

The aim of the corpus was to provide a large sample of Standard Russian continuous speech. It was originally intended for use in unit-selection TTS synthesis, however, with the idea that it might be suitable for a wider range of phonetic research and development.

Firstly, the sample represents a number of speaking styles. As the corpus includes only read speech, different styles of texts were selected for recording with specific characteristics of those styles in mind:

Text A - an action-oriented fiction narrative resembling conversational speech in sentence length and word choice;

Text B - a fiction narrative of a more descriptive nature containing longer sentences and very little conversation;

Text C - a play containing a large number of conversational remarks and emotionally expressive dialogues and monologues;

Text D - purely informational neutral texts on IT;

Text E - purely informational neutral texts on politics and economy issues. Both informational texts contain terminology, geographical and proper names, numerals, acronyms and abbreviations.

The distribution of the texts over speakers is presented in fig.1 in the Appendix.

The diversity of the selected texts served our other goal of making the corpus phonetically and prosodically rich, i.e. containing a large number of all Russian phonemes in all possible contexts and a wide range of prosodic structures. The corpus is composed of 60 hours of speech recorded from 8 speakers (7.5 hours from each speaker).

Thirdly, the corpus was intended as a sample of Standard Russian (St. Petersburg pronunciation variant); dialect variation was not accounted for. However, the recordings were made from eight speakers, four men and four women, in order to cover a certain degree of variation within the St. Petersburg pronunciation variant.

Fourthly, it was necessary to ensure high quality of all data both in terms of technical characteristics and voice quality. The latter objective was achieved by recording professional speakers: some of them worked in radio broadcasting; others were professional actors or television newsmen. In addition to voice training, pleasantness of voice and clear articulation were considered.

The fully annotated part of the corpus covers all speaking styles included in the corpus and all speakers. General corpus statistics is presented in Table 2 in the Appendix.
The corpus has six annotation levels including prosodic annotation, rule-based canonical phonetic transcription and manual transcription reflecting the actual sounds pronounced by the speakers.

2. ACOUSTIC ANALYSIS

The fundamental frequency periods were detected automatically. A linear combination of the following methods was used for this purpose: autocorrelation, analysis-by-synthesis, spectral domain analysis, estimation of the signal peak energy, estimation of the ratio of lengths and correlation of neighboring periods. The efficiency of automatic pitch detection and pitch period labeling was about 98%. The results of the automatic procedure were checked and corrected manually.

3. PROSODIC ANNOTATION

Prosodic annotation was performed by expert phoneticians on the basis of auditory and acoustic analysis of the speech data and presented in a text file containing orthographic transcription. Prosodic labels were later automatically transferred from the text file to the annotation files to coincide with the phonetic transcription levels. Orthographic transcription was stored on the corresponding tier (Level 5), which contains words boundaries and word labels. The prosodically prominent words were marked with special symbols.
Prosodic information was stored on the next tier (Level 6), it contains tone unit boundaries, pauses and their labels. The distribution and frequency of intonation models in the analyzed material is shown in Table 3 in the Appendix.
The set of symbols for pauses and tone units and the principles behind the labeling procedure are described in detail in our earlier publication [Skrelin, Volskaya, 2009].

4. AUTOMATIC INTERPRETATION OF SENTENCE PROSODY

One of the most important aspects of this project is that it revealed the lack of data concerning the relationship between the acoustic and functional properties of the intonation units, which affects the prediction of the intonation model based on the automatic analysis of the acoustic data.

This article deals with the results of the analysis of errors in automatic interpretation of the intonation model [Скрелин, Кочаров, 2009]. It outlines some aspects of prosodic data analysis that need further investigation:

- rate of F0 changes;
- discrepancy between the acoustic properties and perceptual evaluation of static high level tones deserve special attention (some of them perceived as "rising" even when presented without any context);
- a considerable mismatch between the acoustic and perceived properties of the melodic patterns (rising or falling) of the intonation units, suggesting "an analogy decision" in favor of their functional properties (question, enumeration, etc.) is of great interest here.

4.1. F0 slope

At first sight this parameter seems to be properly investigated by M. Rossi [Rossi, 1971], P. Mertens [Mertens, 1995]: the calculation of its perceptual limits is included in the MOMEL stylization for INTSINT. Nevertheless it's role and possible function is far to be clear.
Sometimes it is difficult to interpret whether we are dealing with static or kinetic tones.
One may ask why I am speaking about the F0 slope and show the picture of F0 transition – a jump - from one melodic level to another via a voiceless plosive (see Fig. 2 in the Appendix). How does a listener decide whether he perceives a kinetic tone or a transition from one pitch level to another?
This problem was approached by several investigators. The experiments performed by Rossi and others were concerned with the determination of the melodic rising and falling glissando thresholds. A fundamental frequency variation that takes place during a given time interval will be perceived as a pitch movement if the rate of F0 change exceeds some minimal amount; this amount depends on the duration of the transition: the shorter the stimulus, the larger the required frequency change. Frequency variations below this threshold are perceived without pitch change, that is, they are perceived as static tones. Strictly speaking, the calculated limits are different among researchers (see, for example, Rossi and Mertens) .
Formally, a similar situation we can see in the Fig. 2. The perception tests show that listeners describe this transition as a rising movement.
From the physiological point of view the transition from one F0 level to another one via a voiceless

consonant or a micro-pause is not different from a movement (glissando) from one F0 value to another: the vocal folds may be compressing or stretching, but in case of glissando they are vibrating.

From the **technological** point of view nobody has any doubts about the consistency of the INTSINT-based re-synthesis: the transition from one F0 target point to another is always performed as a tonal movement, not as a jump from one pitch level to another. The rate of hidden "F0 change" (Fig.2) demonstrates that the transition from one F0 level to another is enough for it to be perceived as a rise (here expressing non-finality). Apparently, only the degree of F0 slope allows us to differentiate the continuous movement of the fundamental frequency from a melodic contour break. It is very important for the automatic analysis of prosodic data, especially of spontaneous speech where a lot of micro-pauses are added. The main problem with the F0 slope is to decide what speech signal sequences (units) are characterized by this parameter. It is not an easy task. For example, it is obvious that the listeners are insensitive to the F_0 changes on voiced consonants, except for their function as a domain for a "hidden" F0 movement. Thus it is important to know under which conditions a change in F0 slope is perceived. The critical slope change is known as the differential glissando threshold.

4.2. High and low tones, levels and configurations etc.

In dealing with tone specification we face two problems:

- How does the listener know that the tone is high if the speaker is unfamiliar to him?
- What are the acoustic features that could ensure the discrimination between falling tones (finality) and rising-falling (question)?

I have no answer to the first question. I presume that the reaction of the listener is based on spectral features of the vowel. It may be a spectral tilt, a spectral poverty of the vowel (the lack of expectable harmonics) due to high (for a particular speaker) fundamental frequency. Of course this phenomenon is rather exotic because usually we talk to people we know, and we somehow "calculate" highs and lows of their voices. But in phonetic research we are interested in all kind of situations and we need to know the basis for perception of a pitch level of an unfamiliar voice.

As for the second problem, I am much more optimistic. First of all, the need for an acoustic feature (or maybe a derivative of one or more acoustic features) requires a certain amount of speech material. So the F0 value of the presented tone can be compared to the tonal context: to the previous F0 level (the end of the preceding intonation unit or phrase), or to the average F0 (moda, median) value in the material, or even to the absolute F0 maximum in the material. This information can help determine the nature of the tonal movement (rising, rising-falling or a starting point for the falling movement) and, consequently, its function in the intonation unit.

4.3. An analogy decision

Analyzing prosodic annotation of the experimental material, which was performed by expert and trained phoneticians, we realized that in some cases they reacted to the communicative function of the intonation unit (for example, non-finality, as shown in Fig. 4 in the Appendix) and neglected objective prosodic information. In this situation even a qualified phonetician may act as an ordinary speaker (listener). It is a very serious problem for building speech corpora. To fish out this kind of listeners' mistakes it may be useful to refer to the automatic prosodic data interpretation, but this is another very special problem, and another story.

5. CONCLUSION

I hope this brief overview of the problems one may encounter in prosodic data analysis will be useful for those who perform prosodic annotation for their speech corpora.

6. REFERENCES

[1] Skrelin P., Volskaya N., Kocharov D., Evgrafova K., Glotova O., Evdokimova V. 2010. CORPRES – Corpus of Russian Professionally Read Speech. *Proc. of TSD.* Berlin.

[2] Skrelin P., Volskaya N. 2009. Prosodic model for Russian. *Nordic Prosody X*. Frankfurt am Main.

[3] Скрелин П.А., Кочаров Д.А. 2009. Автоматическая обработка просодического оформления высказывания: релевантные просодические признаки для автоматической интерпретации интонационной модели. *Труды третьего междисциплинарного семинара «Анализ разговорной русской речи» (АРЗ-2009).*

[4] Rossi, M. 1971. Le seuil de glissando ou seuil de perception des variations tonales pour la parole. *Phonetica*, 23.

[5] Mertens P., Alessandro Ch. d'. 1995. Pitch contour stylization using a tonal perception model. *Proc. Int. Congr. Phonetic Sciences* 13, 4. Stockholm.

APPENDIX

Table 1: The distribution of the experimental material over speakers

Text type / Speaker	A	B	C	D	E
Female 1	■	■	■		
Female 2	■	■	■		
Male 1	■	■	■		
Male 2	■	■	■		
Female 3	■	■		■	
Male 3	■	■		■	
Female 4	■	■			■
Male 4	■	■			■

Table 2: General corpus statistics

	Fully Annotated Data	Partly Annotated Data	Total Amount
Phonemes	1 048 867	-	1 048 867
Words	211 437	317 021	528 458
Tone Units	64 055	86 546	150 601
Hours	24	36	60

Figure 1: Phonetic transcription and prosodic annotation

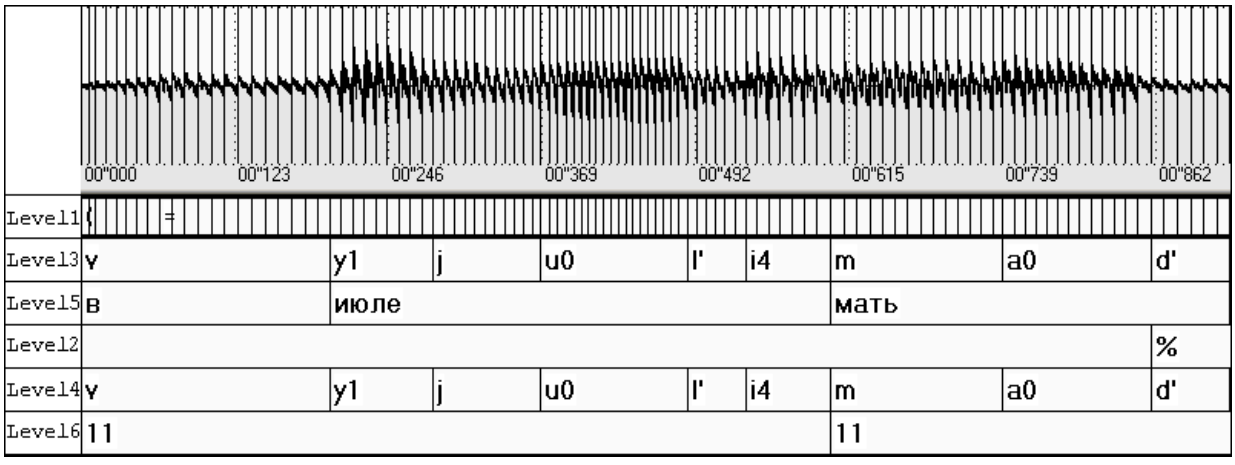


Table 3: Distribution and frequency of intonation models.

	a	%	c	%	g	%	k	%	m	%	o	%	s	%	u	%
1	2922	21,2	5068	28,4	2163	21,5	3443	23,2	4213	31,6	2357	21,6	917	25,3	2302	33,7
2	1702	12,3	2689	15,1	1233	12,2	1991	13,4	1432	10,7	1099	10,1	572	15,8	862	12,6
3	148	1,1	517	2,9	82	0,8	291	2,0	506	3,8	75	0,7	28	0,8	165	2,4
4	469	3,4	1381	7,7	218	2,2	1019	6,9	1292	9,7	219	2,0	103	2,8	413	6,0
5	56	0,4	37	0,2	12	0,1	78	0,5	62	0,5	6	0,1	7	0,2	19	0,3
6	29	0,2	176	1,0	5	0,0	109	0,7	72	0,5	2	0,0	3	0,1	56	0,8
7	176	1,3	788	4,4	85	0,8	438	3,0	589	4,4	84	0,8	34	0,9	251	3,7
8	29	0,2	126	0,7	9	0,1	126	0,8	94	0,7	15	0,1	6	0,2	48	0,7
9	339	2,5	755	4,2	144	1,4	298	2,0	291	2,2	325	3,0	130	3,6	257	3,8
10	1039	7,5	1223	6,9	638	6,3	1	0,0	1099	8,2	1300	11,9	444	12,3	639	9,3
11	4328	31,4	2503	14,0	4034	40,1	5881	39,6	2578	19,3	4088	37,4	1192	32,9	1105	16,2
12	1889	13,7	1821	10,2	1233	12,2	572	3,9	834	6,2	984	9,0	109	3,0	671	9,8
13	663	4,8	756	4,2	213	2,1	589	4,0	280	2,1	378	3,5	75	2,1	48	0,7
Total	13789	100,0	17840	100,0	10067	100,0	14837	100,0	13347	100,0	10932	100,0	3620	100,0	6837	100,0

Figure 2: F0 rise (2 ST) “during” a voiceless plosive. The slope (19,3 ST/sec) exceeds the perceptual limit (15,5 ST/sec)

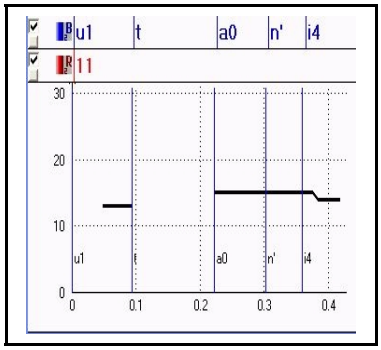


Figure 3: High level tone

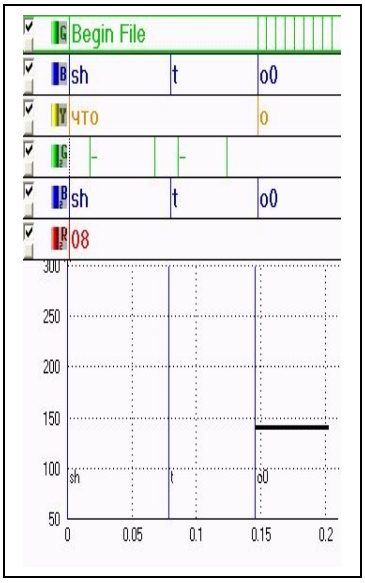
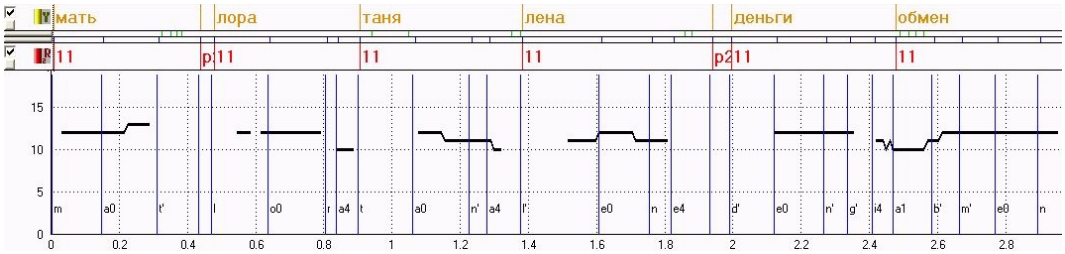


Figure 4: An analogy decision



THE DANGEROUS STATUS OF A LINGUA FRANCA

Svetlana Soboleva

Far Eastern Federal University, Russia
kredimmer787@hotmail.com

ABSTRACT

Among a variety of problems arising from the global spread of the English language there are a number of those which affect the lingua franca most dramatically. The side effects are rooted in the careless and superfluous attitude to English which is now rampant in this country. Articulation fell the first and easiest victim to such an approach. This ought to be expected – foreign articulation is the most tiring and boring aspect in the curriculum. Apart from losing the unique sound of the unique language, we are stripping English of its esthetic function. Unless this trend is reversed, in 10 to 15 years the status of English will be that of a vehicle rather than a language.

Keywords: articulation, problems, lingua franca.

In the recent 20 – 30 years pragmatics has been in vogue with both linguists and methodologists. Strictly speaking, there is little methodology in the pragmatic approach. This is bad enough. But what makes things worse is the fact that a key pragmatic concept of communication has become a justification for a superficial and careless look on the language paradigm. With the advance of the new teaching schemes this look has become especially threatening.

In my opinion the English language is going through a very troublesome period. The lingua franca status involves many negative consequences, the most unpleasant one is that millions of people *are forced* to learn it. Acquiring a foreign language certainly implies acquiring elements of the foreign culture. Strange as it may seem, but it is the European specialists who, for some reasons, appear reluctant to adopt English as the language of global communication, whereas in Asian countries English is given a red carpet welcome. The cultural aspect associated with the language is found especially irritating in Europe. "English has come to represent a specific cultural tradition…including industrial society, commercialism, free-market orientation, individualism, media dominance, and unfortunate colonial histories. While not all these things are necessarily bad (and are in fact emulated), they are not appealing to everyone, especially countries who feel their cultural traditions slipping away under the bombardment of the English language movies, radio, television, music, products, and now the Internet", writes Alan Corre who started creating Lingua Franca Nova, a blend of Romance languages, in 1965 [http://www.lingua-franca-nova.net./lfnappen.html].

Europeans fear very much lest their national identity should be replaced by a newcomer. "Although many native English speakers would rarely consider common speech to be offensive, at times, it would not be suited for an international lingua franca, which should be as culturally neutral as possible…One of the goals would be to remove phrases with inappropriate or culture-specific associations (for example, sports terminology). While this is derided by some as political correctness… this is necessary for a lingua franca, because otherwise some Europeans who adopt the English language as a means of everyday communication would also be forced into adopting the customs, traditions and modes of thought specific to the major English-speaking countries, many of which are embodied in the language", argue ELFE (English as a lingua franca for Europe) activists. [http://www.experincefestival.com/a/English_as_a_lingua_franca_for_Europe].

Another problem is rooted in a rather loose correlation between the sounding and the writing of the English language. "English has one of the worst spelling systems of any language using a western alphabet. Unless it were to dramatically alter its spelling system – not a likely event – it will continue to mystify those who learn it as a second language, not to mention its own native speakers", comments Alan Corre on the de facto lingua franca status of English.

Millions of people prefer English over any other language when deciding on their career. In fact, these people are, to put it mildly, indifferent to the English language and find practicing its grammar and sound system too much of a strain. This state of things cannot be judged in terms of "good" or "bad". This state of things is inevitable because it is natural. A foreign language will always be compared unfavorably to a native tongue by ignorant

learners, in other words, a foreign language will constantly be "weighed in the balance and found wanting". It has always been so since people began learning foreign languages, and it never caused any concern. The schools were well equipped with thick grammar books and strict teachers. Methodology, though in its infancy, was not afraid to call a spade a spade, that is to say, to call a mistake a mistake. Mispronunciations were not permitted automatically, and no allowances were made for the fact that interdental consonants were too difficult for the Russian learners of English. Nobody would have thought of tailoring phonetic courses so as to meet the demands of those who were too busy, or too lazy to project the tip of the tongue between the upper and the lower teeth. Students were given advice on how to avoid or eliminate mistakes. No one was stigmatized, of course, for mispronouncing English words, but everyone knew the difference between "the right" and "the wrong", and everyone was encouraged to pronounce correctly. Good speech habits used to be appreciated because they testified to a high professional and social standing.

The situation has changed dramatically after the English language was placed in the center of linguistic research due to globalization. Globalization produced new doctrines and concepts in social and political science as well as in linguistics. Political correctness, tolerance, and multiculturalism have shaped modern schools of thought and have had a huge impact on the methods of teaching English. Millions of people around the world took to English. They attend English language classes at school and in college, at their work place and after office hours, in their native villages and at local community centers. Their teachers' qualifications vary from quite reputable to non-existent. Educationalists find themselves in a rather awkward situation – while it is all right to say that a *number* of English language learners are apt to mispronounce *some* English sounds, it is politically incorrect to say anything to this effect if the *number* of English language learners mispronouncing *most* English sounds is topping two hundred million. Hence politically correct terms like *transference*, *expanding circle*, and *basilect*. Recent social developments in Europe have convinced us that the value of multiculturalism and political correctness had been considerably overestimated. There is nothing wrong about politically correct terms, though. They are designed to provide theoretical framework for theoretical research. And what does this research reveal? It reveals that there are many more those who speak bad English than those who speak good English. Educationalists are facing an unpleasant dilemma again – which of the two groups of speakers should be relied upon in setting up teaching standards. Willing to be politically correct and tolerant, educationalists have adopted the conformist stance which leads to degraded teaching standards and, in the long run, to the decline of the English language. As John Simon puts it in his sincere book *Paradigms Lost*, "there are things which are worth fighting against".

Millions of people learning English around the world are motivated by prospects of successful career and social prestige. But apart from informing customers about the terms of payment and rates of interest, people need to use the language as the major means of self expression and social integration. This is where the conformist approach fails to reach the goal. Associations of English language teachers are looking for minimum core courses into which to squeeze elementary grammar and survival or professional vocabulary. Phonetic aspect is either neglected or declared irrelevant to the successful communication, because communication under this approach has been reduced to a mere exchange of information. Atrocious accents are regarded as regional varieties and enjoy legitimate status as Received Pronunciation, General American, or Educated Australian. Apart from being unintelligible, these accents erode the phonetic system of native speakers of English. This problem is not new. John Priestley was concerned about it as early as the 1930s. In his well-known novel *Angel Pavement* he provided individual speech characteristics to every personage and emphasized their pronunciation habits. This is how he characterized Mr. Golspie and his daughter Miss Lena Golspie: "... Mr. Golspie spoke with a breadth of vowel sound and roughness of consonants that suggested the toned-down Lowlander or North-country Englishman, whereas his daughter's English did not properly belong to any part of England but seemed to be that international English, of a kind that a clever foreigner might pick up in the Anglo-Saxon colony in Paris and that is sometimes spoken by both English and Americans on the stage, a language without roots and background, a language for 'the talkies'. Indeed, in Lena's company you might have felt you were taking part in a talkie".

A minimum set of internationally acceptable phonetic elements of the English language is what linguists are aiming at now. A list of most unfortunate consonants and vowels has already been prepared by the ELFE phoneticians. The following sounds have been declared *persona non grata*:

1. interdental fricatives, both voiced and voiceless, indicated by letters *th*.
Similar consonants are found only in Spanish, Greek and Icelandic. French speakers replace them with [s] and [z]; Scandinavians and Italians – with [t] / [f] and [d]. For the purpose of international communication ELFE would choose one of these sounds and standardize it (Sic!)
2. all final voiced consonants.
In most other Germanic languages (as well as in the Russian language) consonants in word final position are never voiced, so native speakers of those languages tend to devoice English voiced consonants in this position.
3. all English vowel pairs differentiated by the vowel duration, like in the words "beat" and "bit".

The whole vowel system of the English language doesn't look good to ELFE – very many vowels and very few distinctive features. "The most obvious difficulty
is the large number of vowel sounds in the English language, each one of which has to be learned by listening and training tongue placement."
[http://www.experiencefestival.com/a/English_as_a_lingua_franca_for_Europe]

Shall we expect further suggestions from ELFE on how to "improve" the English language?

The innovations in the teaching approaches are spectacular. In the light of these innovations the history of the English language may roughly be divided into two basic periods – the before-lingua-franca period and the after-lingua-franca period. The *"before" period* teachers were seeking to liquidate mistakes in the pronunciation of difficult English sounds; the *"after" period* teachers are seeking to liquidate these sounds.

Such approach to the English language is spreading rampantly. Its phonetic system fell the first and easiest prey. What will come second? Spelling? Grammar? Unless the concept of the English language as a lingua franca is totally revised, the English we shall have to speak in 30 – 40 years from now, will be a relic vaguely reminding us of its past beauty and power.

REFERENCES

[1] Crystal, D. 2003. *English as a Global Language*. Cambridge.
[2] Crystal, D. Keynote Lecture – "Language Death"
http://compassconference.wordpress.com/2009/10/20/crystal/
[3] Simon, J. 1980. *Paradigms Lost*. New York.
[4] http://www.lingua-franca-nova.net./lfnappend.html.
[5] http://www.experiencefestival.com/a/English_as_a_lingua_franca_for_Europe

ONE RISE-FALL, TWO RISE-FALL...

Nina B. Volskaya

St.Petersburg State University, Russia
volni@phonetics.pu.ru

ABSTRACT

The paper presents the results of acoustic and perceptual study of the Rise-Fall intonation in Russian. Experimental data suggest age-dependent variations in the realisation the Rise-Fall intonation which affect the way it is perceived by Russian subjects. In a number of cases these variations may lead to communicative conflicts between different age groups of native Russian speakers. The comparison of Rise-Fall intonation realisation and functions in Russian and English is provided.

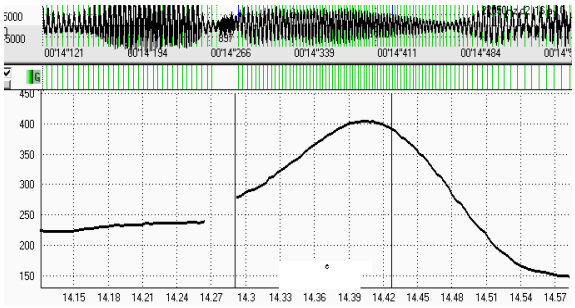
Keywords: Rise-Fall, form and function, perception, age-dependent variation, communication.

PART 1. ONE RISE-FALL...

Observations over intonation patterns of Russian and English languages indicate more similarities than expected. In this study we devote ourselves to some of them.

Russian is known for its specific intonation used to convey a yes-no question: a rise on the tonic syllable followed by a steep fall on the post tonic syllable, if there is any (Fig.1).

Figure1: Typical intonation pattern for Russian general questions: F_0 peak on the tonic vowel



The same pattern although with a somewhat smaller F_0 excursion is observed in non-final units. In neutral general questions and non-final intonation units the F_0 peak should coincide with the tonic vowel (see Intonation Construction # 3 – IC3 – of E.A. Brysgunova [Брызгунова, 1980].

In recent studies of the intonation variation in Russian spontaneous and read speech we came across realizations characterized by shifting of the F0 peak further to the right, so that it is either late in the vowel (or the tonic syllable), or outside the syllable altogether (Figs.2-4) [Вольская, 2008].

Figure 2: Late F0 peak on the post-tonic syllable of the word "**uli**tse" in a non-final unit. Female speaker

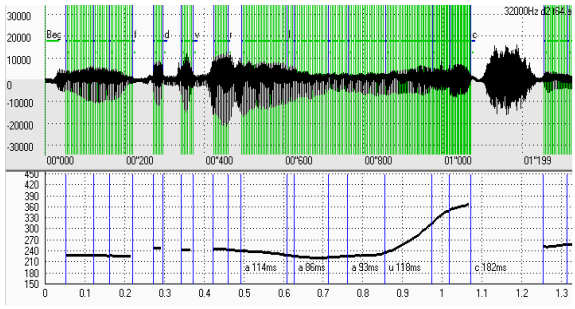


Figure 3: Late F0 peak on the post-tonic vowel of the word "ом**у**тах". Male speaker

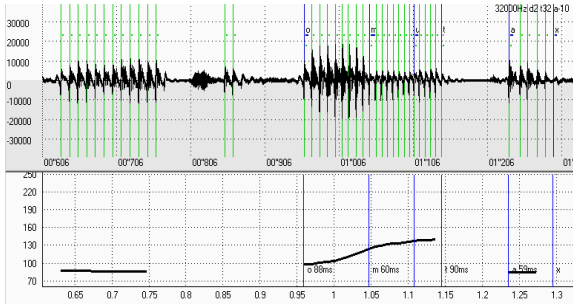
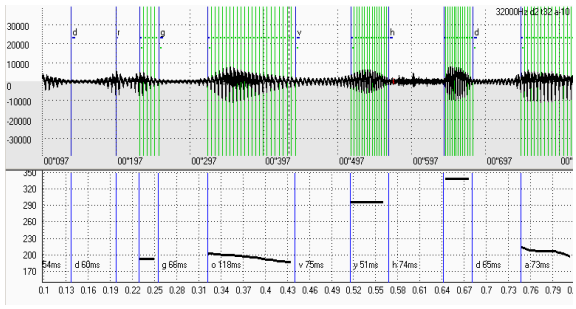


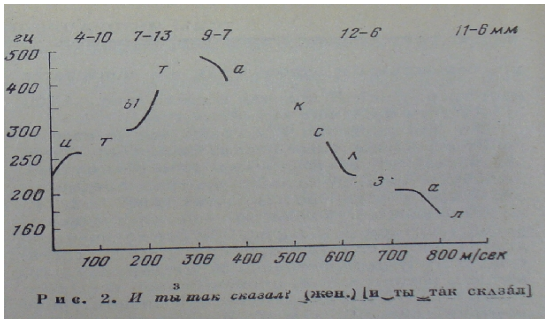
Figure 4: Late F_0 peak on the post-tonic vowel of the word "вы**хо**да". Female speaker



These observations were confirmed by a follow-up study, devoted to the phonetic realizations – late F_0 timing in particular — of the non-final and question intonation in Russian. It showed that they are

common for young Russian speakers mostly [Демидчик, 2009] .
Right-shifting of the F_0 peak in IC3 was mentioned by Bryzgunova (Fig.5) as a possible means for providing special emphasis to a particular word in the question and adding to the utterance note of astonishment, criticism, challenge, etc. to it. [Брызгунова, 1984].

Figure 5: Late F_0 peak on the post-tonic in an emotional question " И **ты** так сказал?" (an example taken from Bryzgunova,1984).



So questions accompanied by a late F_0 peak placement are by no means neutral requests for information. If this question intended by a young speaker as a neutral one, is addressed to an elderly person, whose intonation system requires a different melodic pattern for a neutral question, real feelings and intentions of the speaker may be misinterpreted, and a communication conflict may result.
To confirm or reject this proposition an auditory experiment was carried out using research material specially designed for the purpose [Демидчик, 2009].
90 general questions and non-final intonation units characterized by a displaced F_0 peak were selected from the interviews, recorded from 5 female speakers aged 20-22. They were presented to three groups of listeners: school children aged 12-16, students aged 20-24 and a group of subjects aged 50-60 accordingly. They were requested to answer whether the question they hear is a neutral one, or tick the word from a list indicating a particular emotion if they perceive it in the sentence they hear. The results are presented in Table 1 (see the Appendix).
Though a late F_0 rise followed by a fall seems to be associated with a pronounced emotional reaction in Russian, for two groups of young listeners the effect of late F_0 timing is considerably smaller, than for the third group of subjects (potentially, their parents or grandparents) , who associated it with a particular set of emotions: impressed, happy, boasting on the one hand, and challenging, reproachful, antagonistic etc. on the other.
We may conclude that a lot of misunderstanding between younger and older generations is due to differences in the intonation: there is a mismatch between the
- intention (neutral request for information),
- realization (a displaced F0 peak),
- perception (not immediately pleasant for the listener emotional coloring of the utterance)

and, consequently, misinterpretation of the intended message.

This seems to be an interesting and special case of the intra-language interference.

PART 2. TWO RISE-FALL....

In traditional models of English intonation describing the RP (pronunciation standard) it is widely acknowledged that the rise-fall signals strong emotions, either negative or positive, when the speaker is impressed, either favorably or unfavorably. According to O'Connor& Arnold [O'Connor& Arnold, 1961, 1973], who include the rise-fall in the inventory of English complex falling tones (the Jackknife), it may sound challenging, antagonistic, authoritative, reprimanding [O'Connor& Arnold 1973: 45]. Others report such attitudes as "teachingly reproachful" [Schubiger, 1958], "mocking and impatient; in some cases protest against a false assumption" [Kingdon, 1958: 221]; haughty, dismayed, signaling great annoyance or satisfaction [Crystal,1976: 305]. "Intense surprise" seems to be the most suitable general label.
At the same time, in certain varieties of English, rising-falling intonation is widely used in neutral discourse, as it is not associated with any of the listed meanings: "In fact the Welsh employ the rise-fall …in circumstances where it would not be used in Southern England" [Jones, 1967: 162], e.g. I'm 'going to ^Llanelly.
D. Jones, among others, also mentions a displaced F_0 peak as a means for providing extra emphasis [Jones, 1967: 159].
As it follows from the description, the English rise-fall is similar to the Russian rising-falling intonation described above not only in its phonetic realization, but also in the set of attitudes associated with it.
We face a very curious situation, which in English may lead to the intra-language inter -dialectal prosodic interference. In Russian it resulted in misunderstanding between younger and older genera-

tions: children and their parents seem to speak different languages.

We can easily project this situation on second language acquisition. Would it lead to a communicative conflict?

CONCLUSION

Unfortunately, intonation has been largely ignored in second language acquisition research: in many books on second language learning there is no reference to intonation or prosody. The empirical analysis of native (Russian) and target (English) language intonation may shed light on the processes of prosodic interference and yield important social, sociolinguistic information about the prosody of the speech varieties within one language and across languages. It may also help answer the question: What are the consequences of a young Russian unintentionally using too many rise-falls in his English speech?

REFERENCES

[1] Брызгунова Е.А. Интонация. 1980. *Русская грамматика*. Н.Ю.Шведова (Ed.). Москва: Наука, 96-122.

[2] Брызгунова Е. А. 1984. *Эмоционально-стилистические различия русской звучащей речи*. Москва.

[3] Вольская Н. Б. 2008. Соотношение интонационного типа и вариантов. *М-лы XXXVII Международной филологической конференции. Формальные методы анализа русской речи.* Вып.7. СПб: СПбГУ.

[4] Демидчик Л. В. 2009. *Фонетическая реализация и функциональная нагрузка восходяще-нисходящей интонации в речи русской молодёжи (экспериментально-фонетическое исследование).* Дипломная работа. СПб: СПбГУ.

[5] Crystal D. 1976. *Prosodic Systems and Intonation in English*. CUP.

[6] Jones D. 1967. *The Pronunciation of English*. Cambridge at the University Press, 4th edition.

[7] O'Connor J.D., Arnold G.F. 1961, 1973. *Intonation of Colloquial English*. 1st and 2nd eds. London, Longman.

[8] Kingdon, R. 1958. *The Groundwork of English Intonation*. Longman.

[9] Schubiger M. 1958. *English Intonation. Its Form and Function*. Niemeyer.

APPENDIX

Table 1: Distribution of the types of response in three groups of listeners, %.

Type of response	Subjects, responses ,%		
	Schoolchildren	Students	Grown-Ups
Neutral	55	55	33
Emotional	34	39	67
No answer	11	6	0

RUSSIAN-ASIAN CORPUS OF ENGLISH IN PHONETIC RESEARCH

Victoria Zavyalova

Far Eastern Federal University, Russia
zavyal@mail.ru

ABSTRACT

The paper addresses the issues of applying ELF corpora (Russian-Asian Corpus of English) for developing phonological and perceptual competence among ELF communicators, as well as for conducting research related to phonetic organization of World Englishes. Now that English has become a lingua franca in most spheres of life, the participant of intercultural communication has to be aware of its increased phonetic and phonological variation. English production and perception bases are no more isomorphic. Productive listening comprehension implies a more sophisticated perception base that contains multiple phonological samples of World Englishes. Since communication breakdowns may occur due to unexpected phonological variation, the only way to get prepared for specific ELF talk is to make listener's ear more accustomed to different accents. Russian-Asian Corpus of English provides a rich source of English speech samples recorded in various contexts in the Far East of Russia for educational and research purposes.

Keywords: phonological variation, Asian Englishes, perceptual competence.

Linguists around the world have come to admit that English has become an Asian language [Adamson, 2004; Asian Englishes, 2007; Bolton, 2003; Bondarenko, Zavyalova, 2008; Kachru, 2006; Honna, 2008; Jenkins, 2009; Stanlaw, 2004]. Obviously, the researchers have applied to a bulk of evidence testifying to the fact that English is widely employed as a mandatory lingua franca in inter- and/or intranational communication settings in many countries of East and Southeast Asia. In Southeast Asia, for example, English is used as a sole official language for regional cooperation in ASEAN (Association of Southeast Asian Nations comprising ten states) [Kirkpatrick, 2010].

Asian Englishes is a reality of the globalizing world. The existence of these varieties of English suggests that there should be emerging ELF accents with their phonological specificity determined by the speakers' different mother tongues. Irrespective of the attitudes towards phonetic and phonological variation in the Expanding circle varieties (from treating them as an ethnic sound coloring to recognizing them as deviations from the standard pronunciation model or even errors (see e.g. J. Jenkins and B. B. Kachru [Jenkins, 2009; Kachru, 2005]), ELF interlocutors, in practice, aim at achieving successful communication. Thus, the main communication strategy can be, so to speak, just taking it as it is and making the best of it. In other words, to get prepared for specific ELF talk in Asia and avoid potential breakdowns, one has to learn how to predict the likely pronunciation patterns and to make his/her ear more accustomed to dissimilar regional accents of English.

It must be noted that the research of Asian varieties of English in the sphere of prosodic organization has been conducted for over ten years at School of English Phonetics, Institute of Foreign Languages, Far Eastern Federal University. Significant findings on the prosody of China English, Korean English**,** Japanese English, Vietnamese English, Thai English, and Russian (Far East) English have been discovered [Бондаренко, et al., 2007]**.** However, until recently the FEFU team collected the speech samples somewhat inconsistently, as each researcher had to individually design the experimental ELF material to be further recorded in particular country. Collecting speech material abroad is a rather complicated and costly process. As a consequence, a single researcher was able to gain information necessary for the description of only one variety of Asian Englishes.

The ongoing project of Far Eastern Federal University on collecting ELF corpora in the Far East of Russia and in the neighboring Asian countries (*Building Russian Asian Corpus of English*) is intended to provide an optimization tool for carrying out a thorough comparative study of the regional English varieties and to ensure educational exposure to the diversity of Asian English accents. Its

primary aim is to collect and compare Asian English speech corpora in order to derive a set of core properties common to all varieties of Asian English, as well as to discover features that are particular to individual varieties. The major record fields of the RACE database are designed to include samples of the region's most representative non-native English varieties, namely: Russian English, China English, Japanese English, and Korean English. American and British English fields contain reference point samples that can be used to compare and discuss phonetic variations. Other record fields comprise European Englishes, South-East Asian Englishes, Australian English, and New Zealand English.

Each speech sample is provided with background information which contains the indicators of sex, age group, social status, English-speaking setting, etc. All speech samples are subdivided into spontaneous (or quasi-spontaneous) speech and samples of reading written English texts aloud. The reading section texts were selected with regard to potential complexity for Asian speakers (in terms of consonant clusters realization, lexical stress placement, phrasal accentuation, choice of rhythmic and melodic patterns, macrosegmentation strategies, etc). This criterion for text selection ensures the most problematic zones of the English phonetic system at both segmental and suprasegmental levels for non-native speakers are revealed. At the same time, the texts selected for reading aloud meet another criterion: they all contain a clear beginning, middle, and end for the reader to be able to make natural predictions as well as to provide a contextual framework for comprehension (Goodman 1997).

The analysis of the RACE data shows that samples of identical English texts read by speakers of different mother tongues can be efficiently used for categorization of commonalities and differences in the course of comparative analysis of Asian Englishes' phonetic/phonological systems. Fig. 1 (see the appendix) demonstrates the example of multiple RACE objects that can be simultaneously examined with PRAAT, the software used for speech analysis and phonetic transcription of the RACE samples. Fig. 2-3 (see the appendix) demonstrate segmentation and transcription options of the RACE identical sample phrases in PRAAT.

The findings of the analysis confirm that the RACE data can be successfully applied to optimize the complex multilevel phonological study of Asian Englishes. Parallel texts examination contributes to effective categorization as it allows deriving a set of features common to all represented regional varieties as well as unique to an individual interlanguage system. Below are a few findings that have been discovered so far.

Asian varieties of English are characterized by their unique system of phonetic organization, which is determined by the phonetic systems of the speakers' mother tongues. The blending of genetically and typologically different English and Asian languages (Chinese, Korean, Japanese, etc.) results in the phonetic/phonological transfer. For example, the differences in syllabic structure along with the phonotactic rules in the languages in contact cause an interference phenomenon known as resegmentation, which can be detected in the English speech production by non-native speakers. Asian speakers of English tend to simplify complex English syllable structures. One of the ways is by omitting consonant clusters or consonants that rarely occur or are totally excluded in a syllable final position in Asian languages. Thus Asian speakers of English tend to omit the following final consonants [b d t g f v tʃ l s z]. Another way is by inserting vowels between consonants. The acoustic analysis of speech samples showed that commonly inserted vowels in Asian Englishes are [ə], [ɪ], [ʌ], [ʊ].

Any type of resegmentation (plus-segmentation or minus-segmentation) manifesting itself in the change of the phonemic and thus, syllabic quantity, inevitably leads to alterations in the rhythmic structure of a word and, consequently, a syntagm (see Fig. 2 for the example of plus-segmentation of the English word *first* by the Chinese speaker; see Fig. 3 for the example of minus-segmentation of the English word *language* by the Korean speaker). In many cases simplification of pronunciation efforts by Asian speakers is achieved by pausation at the syllable boundaries within a polysyllabic word (e.g. *needless* [ˊniːd|ˊles], *repetitive* [rɪ ˊpe| tɪtɪv] или [ˊre ˊpe|tɪtɪv], *transferred* [ˊtʃan|ˊfəːrəd], *substitute* [ˊsʌbstɪ|ˊtjuː], *regularity* [ˊregjuː læ|ˊrɪtɪ]. As a consequence of modification of the syllabic and rhythmic structures, the stress in a word is either shifted (see also [Zhang et al., 2008]) or placed redundantly on a normally unstressed syllable. Some linguists argue that extra prominence on normally unstressed syllables in Asian Englishes may facilitate perception [Deterding 2005/2006; Lee et al., 2006].

The described phenomena (along with numerous other phonetic features found in the course of study) are rather consistent as demonstrated by the RACE samples, and, hence, can be characterized as specifically intrinsic to phonetic organization of Asian Englishes as a whole. Some of them do not

greatly affect the perception of Asian English speech, while others may considerably impair comprehensibility, especially if the second ELF interlocutor (the listener) has had no or little prior auditory (and analytical) experience of the kind.

It can be concluded that to be successful in today's ELF communication in Asia it is crucial to be aware of the phonological features peculiar to Asian Englishes as it is known that all other aspects of linguistic and sociocultural competences are intrinsically interrelated with the phonological competence. Since the currently built Russian-Asian Corpus of English will elicit multiple production samples representing the varieties of English spoken in the Far East of Russia and Eastern Asia, it can provide valuable resources for extensive research on phonetic variation in Asian Englishes. The data obtained are further applicable for educational purposes, i.e. for developing phonological and perceptual competence among ELF communicators in the Asia Pacific region. The comprehensive investigation of the core and periphery phonological features with regard to comprehensibility and intelligibility issues of Asian Englishes is also a fundamental issue for the development of modern education as well as spoken language science and technology.

ACKNOWLEDGEMENTS

The project has been supported by the Russian Ministry of Education and Science – GK-P733.

REFERENCES

[1] Бондаренко Л.П., Завьялова В.Л., Уютова Е.В., Пивоварова М.О., Киритова И.А., Белоножко Ю.В. 2007. *Слог и ритм английской речи в странах Восточной и Юго-Восточной Азии (Китай, Республика Корея, Япония, Вьетнам, Таиланд).* Владивосток: Изд-во Дальневост. ун-та.

[2] Adamson, B. 2004. *China's English: A History of English in Chinese Education.* Hong Kong: Hong Kong University Press.

[3] Asian Englishes, 2007. Ed. by Kingsley Bolton; Braj B Kachru. - Vol. 1-5. - London : Routledge.

[4] Bolton, K. 2003. *Chinese Englishes: A Sociolinguistic History.* Cambridge: Cambridge University Press.

[5] Bondarenko L., Zavyalova V. 2008. English as Asian Lingua Franca: a Russian Perspective. *Proc. 14th Conference of the International Association for World Englishes held in Hong Kong: World Englishes and World's languages: Convergence, Enrichment, or Death?* City University of Hong Kong; Department of Chinese, Translation & Linguistics, 76-77.

[6] Deterding, D. 2005/2006. Reduced Vowels In Se Asian English: Should We Be Teaching Them? *SOUTHEAST ASIA: A Multidisciplinary Journal* 6 (1), 71-78.

[7] Goodman, Y. 1997. Reading Diagnosis – Qualitative or quantitative? *The Reading Teacher,* 50, 534-538.

[8] Honna, N. 2008. *English as a multicultural language in Asian contexts: Issues and Ideas.* Tokyo: Kurosio Publishers.

[9] Jenkins, J. 2000. *The Phonology of English as an International language.* Oxford University Press.

[10] Jenkins, J. 2003. *World Englishes: a Resource Book for Students.* Routledge: London & New York.

[11] Jenkins, J. 2009. English as a lingua franca: interpretations and attitudes. *World Englishes*, 28 (2), 200–207.

[12] Kachru, Braj B. 2005. *Asian Englishes: beyond the canon.* Hong Kong: Hong Kong University Press.

[13] Kirkpatrick, A. 2010. *Asian Englishes Today. English as a Lingua Franca in ASEAN. A Multilingual Model.* Hongkong: Hongkong University Press.

[14] Lee B., Guion S. and Harada T. 2006. Acoustic Analysis of the Production of Unstressed English Vowels by Early and Late Korean and Japanese Bilinguals. *Studies in Second Language Acquisition*, 28, 487-513.

[15] Stanlaw, J. 2004. Japanese English: Language and Culture Contact. Hong Kong: Hong Kong University Press.

[16] Zhang Y., Nissen S.L. and Francis A.L. 2008. Acoustic Characteristics of English Lexical Stress produced by Native Mandarin Speakers. *Journal of the Acoustic Society of America*, 123 (6), 4498-4513.

APPENDIX

Figure 1: RACE Samples of identical English phrases read by Chinese, Japanese, Korean and Russian speakers listed as Praat objects.

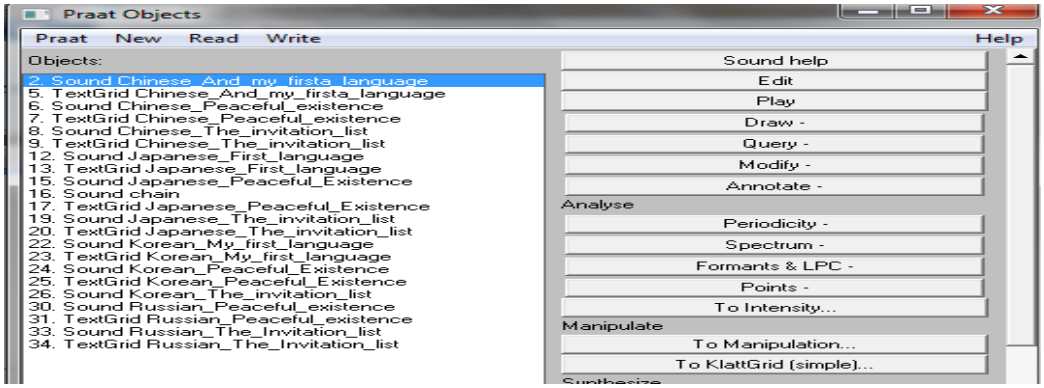


Figure 2: Waveform, spectrogram, phonetic transcription and segmentation at syllables tier of Chinese English phrase (*And my first language is Chinese*).

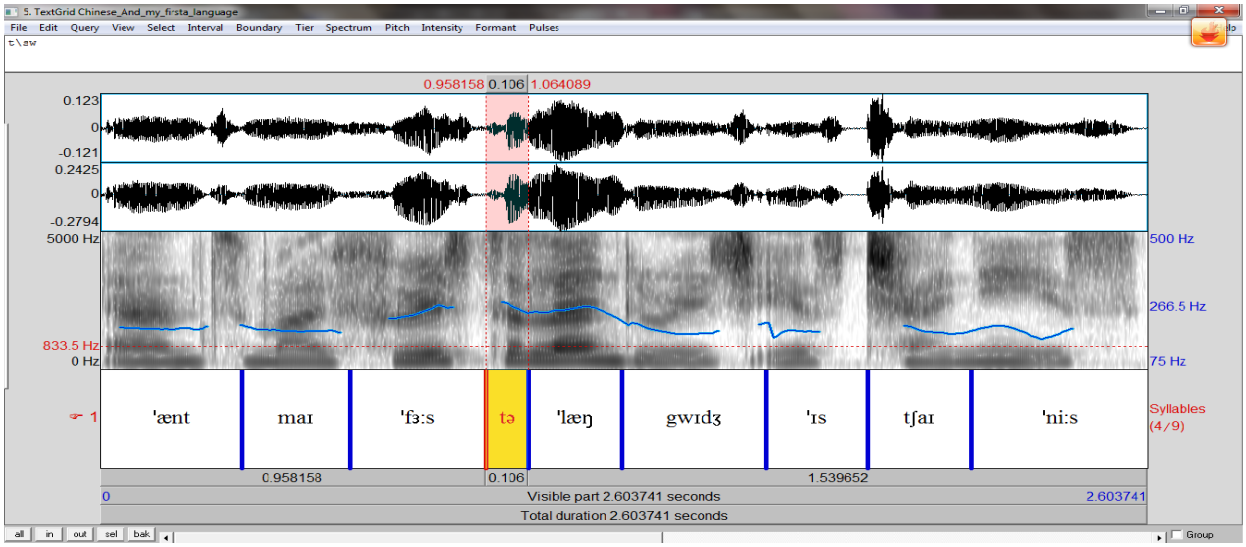
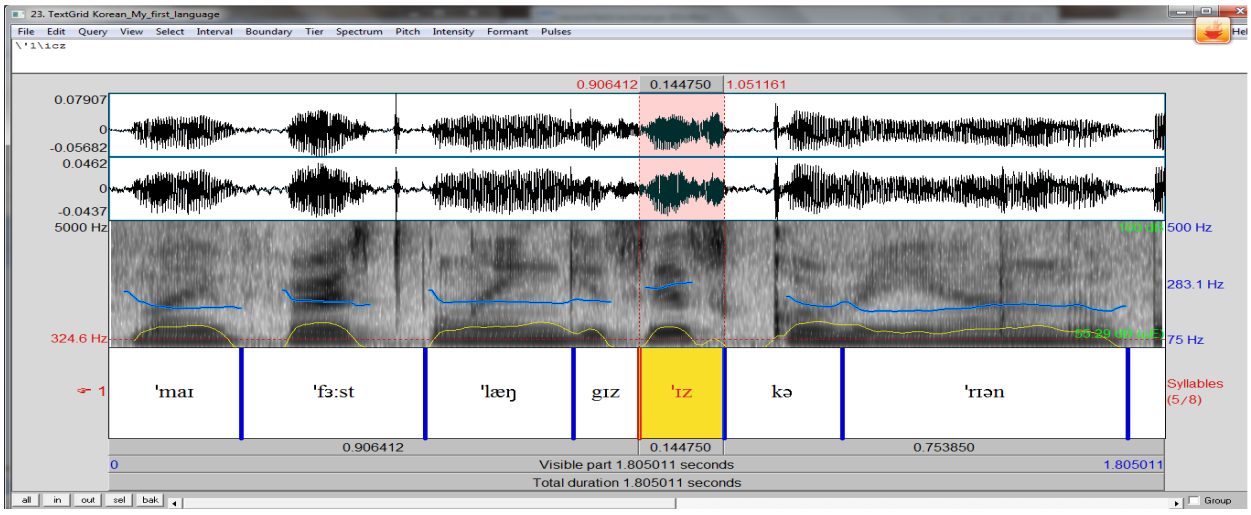


Figure 3: Waveform, spectrogram, phonetic transcription and segmentation at syllables tier of Korean English phrase (M*y first language is Korean*).



THE PROSODIC FEATURES OF BACK CHANNEL ITEMS IN SPONTANEOUS AMERICAN DIALOGUE

Elena, Zharovskaya

The Amur State University, Russia
554454inbox@mail.ru

ABSTRACT

Back-channel feedback is required in order to build responsive spoken dialog systems. It turns out that a low pitch region is a good clue that the speaker is ready for back-channel feedback. This study proves that back-channel feedback has similar prosodic description. The aim was to delineate the prosodic characteristics of back channel feedback which were distributed to three groups in accordance with the fundamental frequency: *Rise, Fall, Rise-Fall*, *Fall-Rise*. The aim was to demonstrate that the prosodic characteristics of back channel items are similar but they have different frequency. The results of the acoustic analysis revealed that some lexically identical words have significant differences in their prosodic parameters.

Keywords: Back-channel items, fundamental frequency, Rise, Fall, Rise-Fall, Fall-Rise

1. INTRODUCTION

Today's typical spoken dialog system produces response until after the speaker finishes an utterance. Humans, in contrast, are very responsive, reacting frequently while the speaker is still talking.

In American English back-channel feedback occurs approximately not only after each utterance but also during them. I hypothesize that in back channels much of the meaning is conveyed by prosody, rather than by the segmental characteristics. Although prosody is one of the major factors for back channel feedback, the pragmatic functions of prosody in back channels which include the class of "confirmation feedback" have not been thoroughly studied yet.

The main idea here is that the speaker provides some clues, which tell the respondent, when back-channel feedback is appropriate. This paper is an attempt to have a closer look at the relationships between prosody and pragmatics in back channels. For this purpose it was necessary to view the pragmatic classification first. However, there turned out to be no universal one. Having studied a number of existing pragmatic classifications [Ward, N., 2003; Ford, C., & Thompson, S., 1996; Schegloff, E. 1999; Schiffrin, D., 1994; Dobrushina N, 2001] I suggest the following one that may be more complete:

1. non-lexical fillers
2. reactive expressions
3. collaborative finishes
4. repetitions
5. resumptive openers
6. elaborations

The second thing to be done was creating another classification of backchannels – the one based on their prosodic features.

2. CORPUS

To look for prosodic cues to backchannels 8 short spontaneous dialogues were recorded digitally in a quiet room at a sampling rate of 44 kHz and 16 bit quantization.

The subjects were 4 volunteer well-educated speakers of American English. The speakers did not report of any speech or hearing disorders. The subjects were asked to talk to each other on suggested issues. The total duration of the recording was 80 minutes.

3. PRELIMINARY ANALYSIS OF PITCH

Acoustic study of pitch was fulfilled using the corresponding PRAAT function. The results run as follows.

1. Almost all collaborative finishes and non-lexical items have a tendency to be lower in pitch.
2. Resumptive openers, however, have a tendency to be higher, perhaps related to the fact that grabbing the turn is a common use of high pitch.
3. Reactive expressions and repetitions, in contrast, are overwhelmingly low in pitch.
4. Elaborations are higher in pitch than non-lexical fillers and collaborative finishes but lower than reactive expressions, repetitions and resumptive openers.

The results of the calculations are shown in Table 1 (see the Appendix).

3.1. Pitch Slope

The study of pitch slope has showed, contrary to the expectation, that the vast majority of back channels have a very flat slope both in short and relatively long words.

3.2. Pitch Contours

It is well known that flat F0 is a distinguishing characteristic of fillers and back-channels also. But the direction of the Pitch contour is the principal way to distinguish the back channels. The back channels were distributed into 4 groups in accordance with their F0 direction (in round brackets the number of occurrence is given):

1. Rise: (21)
The words: Yeah, huh yeah (4), yeah, Okay (5), well (2), yes, huh, yeah (2), etc.
2. Fall (25)
Yeah, Well (3), Ah God, Okay, yes, Oh yeah, I know, It's true, Yes (10), sure, No (4), What oh,
*3. Rise-Fa*ll (22)
Oh yes, yeah(3),Yes, Oh yes, oh that's right, Oh, that was, Well (2), mostly, Yes, No, Sure
4. Fall-Rise (6)
No, Hmuhmu (4), huh, No, Yes, Well.

The biggest group of backchannels is characterized by falling pitch, and they seem to convey decisiveness. Their contour demonstrates tonal and logical completeness. They are pronounced with confidence as the confirmation of the speaker's words. The Pitch of this group varied between 95-156,53 Hz. The difference in pitch was caused by some reasons: rising interest of the topic, speakers' individual traits and their emotional behavior.

Fall

Below there are two examples of backchannels with falling pitch (see fig. 1-2).

Figure 2: The F0 of the word "Yes" (group of collaborative finishes in pragmatic classification)

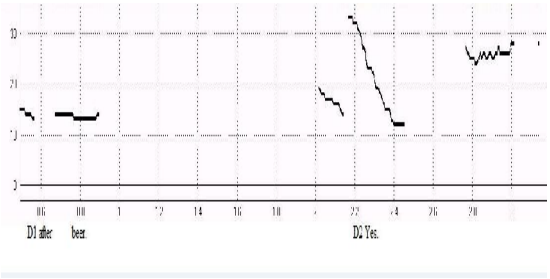
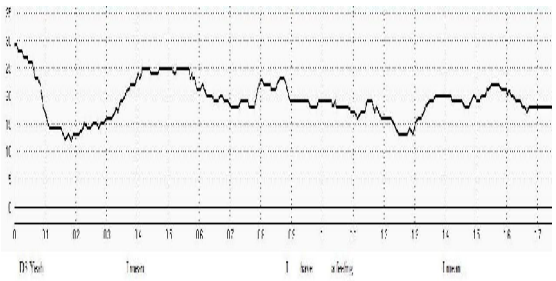


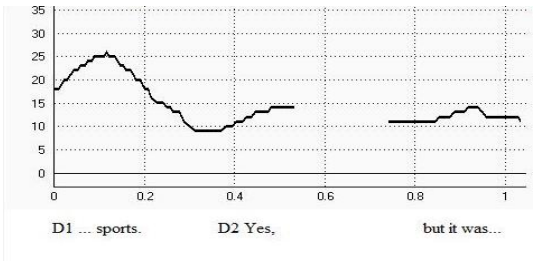
Fig. 2: Fundamental Frequency of the word *Yeah* (group of reactive expressions in pragmatic classification)



Rise

Rising pitch emphasizes the speaker's wish to express his view. He uses the words *Yes, Well, Yeah, Huh* which are applied at the very beginning of the utterance and open a new pronouncement (see fig. 3).

Figure 3: Fundamental Frequency of the word *Yes* group of collaborative finishes in pragmatic classification)



Fall-Rise

Fall-Rise pattern was a rare one in the present research (see fig. 4-5).

Meaning classification

Fall

Back channels seem to convey decisiveness. The contours demonstrate tonal and logical completeness.

Rise

The contours emphasize the speaker's wish to express his view. Open a new pronouncement.

Fall-Rise

Back channels are not emotionally marked out. The direction of their contour is caused by segmental composition.

Rise-Fall

Back channels represent the partner's reaction and demonstrate interest. They are pronounced with emotional stress.

Figure 4: Fundamental Frequency of the word "Hmuhu" (group of non lexical fillers-in in pragmatic classification)

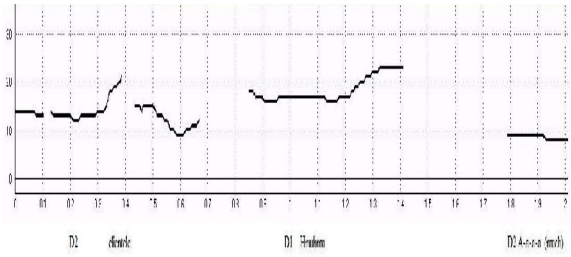
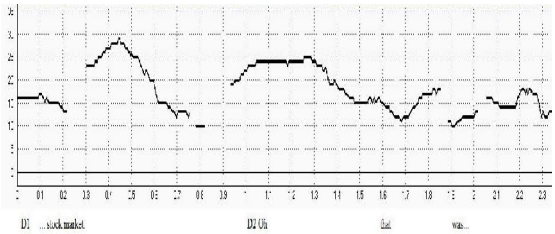


Figure 5: Fundamental Frequency of the word O*h (*group of non-lexical fillers in pragmatic classification)



4. CONCLUSION

1. The pitch contours of non-lexical fillers show that the listener follows the thread of speaker's thoughts and agrees with his words. Falling direction of pitch contour would demonstrate suspicion or hostility.
2. Moreover prosodic features (e.g., fundamental frequency fluctuations, pitch direction, intensity and duration) give the meaning to the words, even to non-lexical.
3. The hypothesis that prosodic characteristics of back channels influence the meaning of the lexically similar words is proven by the fact that all prosodic groups include the lexically similar words.
4. Prosodically backchannels can be easily classified into four groups: the ones with Fall, Rise, Fall-Rise and Rise-Fall tone. No special prosodic pattern for them was discovered: backchannels prosodically do not differ much from other utterances that are classified the same way according to the type of nucleous in the tone unit – Fall, Rise, Fall-Rise, Rise-Fall. In spite of being non-lexical they seem to encompass definite sense.
5. The acoustic study enables to say that none of 92 words was realized with flat pitch contour. The reason might be that the speakers were captivated by the topics.
6. In addition, the present experiments also provide some evidence against grouping back channels according to their pragmatic function only. Back channels of the same pragmatic group might have different prosodic arrangement while back channels belonging to different pragmatic groups might be characterized by the same prosodic arrangement.

5. REFERENCES

[1] Bolinger, D. 1989. *Intonation and Its Uses.* Stanford University Press.

[2] Ward, N. 2003. Non-lexical Conversational Sounds in American English. manuscript.

[3] Sacks H., Schegloff E., and Jefferson G. 1974. A simplest systematics for the organization of turn-taking in conversation. *Language,* 50 (4), 696-735.

[4] Schiffrin, D. 1987. *Discourse markers.* Cambridge: Cambridge University Press.

[5] Tannen, D. 1989. *Talking Voices: Repetition, Dialogue, and Imagery in Conversational Discourse.* Cambridge: Cambridge U. Press.

APPENDIX

Table 1: Relative value of Pitch

Back channels	Relative value of Pitch
non-lexical fillers	1,1
reactive expressions	1,21
collaborative finishes	1,1
repetitions	1,2
resumptive openers	1,21
elaboration	1,17

CONTENTS