# Automatic Transcription Of English Connected Speech Phenomena

Svetlana V. Androsova and Evgenij Yu. Androsov
Department of Foreign Languages
Amur State University
Blagoveshchensk, Russian Federation
androsova_s@mail.ru, eugene_ap@mail.ru

***Abstract*—Many phonetic phenomena that occur in connected speech are classified as phonetic periphery where anything can happen. A well-known convenient way to fix any phonetic phenomenon using certain symbols is transcription. The current paper aims at showing the model of predicting allophones by coordinating a number of factors that determine the choice of a particular allophone and visualizing the result changing certain letters into corresponding IPA symbols. Free Pascal compiler and Geany editor are used for programming purposes. The model is created for American English. It is tested for tap and glottal burst, the latter being one of the three glottalization patterns. The difference of the combination of factors for purely linguistic analysis and for computer programming is explained. We demonstrate (i) the framework for integrating separate blocks each dealing with one phenomenon (ii) a block for tapping which is almost finalized and a part of a block on glottalization, particularly patterns for glottal burst.**

***Keywords—transcription; tap; glottalization; computer modeling; Free Pascal***

## I. Introduction

Since instrumental methods of speech analysis appeared, a lot of research of the flow of speech has been performed, numerous results have been obtained concerning canonical and non-canonical phenomena. A number of connected speech phenomena have been studied, the factors determining them have been more or less described.

A well-known convenient way to fix any phonetic phenomenon using certain symbols is transcription. When data on connected speech phenomena are obtained, a series of logical questions appear: How regular are these phenomena? Can we create a computer model predicting them using e. g. a written text as an input file? How well will this model correspond with reality?

In our paper we tried to answer these questions concerning taps and glottalization in American English. Traditionally in the phonetic literature they are viewed as a part of phonetic periphery where anything can happen [1]. We hypothesize that, no matter being phonetic periphery or not, phenomena like tapping and glottalization demonstrate regular patterns that can be formalized into a computer model. In the current paper we demonstrate (i) the framework for integrating separate blocks each dealing with one phenomenon (ii) a block for tapping which is almost finalized and a part of a block on glottalization, particularly patterns for glottal burst.

## II. Conditions Determining Tapping and Glottalization: Linguistic and Computer View

The choice of the two particular allophones is determined, first by their regular occurrence in the conditions given below (90.1% for taps, 88.9% for glottal burst), second, by the drastic difference of their acoustic characteristics (Praat was used for acoustic analysis [2]) from the ones of the canonical allophone of /t/. Fig. 1–6 demonstrate that difference. In the canonical allophone (see Fig. 1–2) the following well-known phases can be found: occlusion, impulse, friction (always present) and aspiration (appears only before non-front vowels).

Taps (Fig. 3–4) are generally very short, they might or might not have an occlusion; their impulse-like phase is very short and often not localized properly; there is always fundamental frequency (F0) presence and it is often accompanied by considerable intensity drop.

Glottal bursts (Fig. 5–6) vary in duration from quite long to super short, might or might not be accompanied by the previous occlusion, might or might not have F0. Its impulse phase is concentrated in low-mid frequencies and it is very weak for good physiological reason – you cannot produce a strong one with your vocal folds.

Four conditions for /t/ and /d/ tapping in American English can be easily singled out based on phonetic literature analysis:

1. intervocalic ($V_1CV_2$) word-internally in *putting*, *pudding* etc. with any stress pattern of $V_1$ and $V_2$ being only unstressed;
2. intervocalic word-finally in *get in, had a* etc. with any stress pattern of both $V_1$ and $V_2$;
3. before syllabic 'l' in *battle*, *middle* etc.;
4. between rhotic and non-rhotic vowels in words like *party, sort of, harder, heard of* etc.

More information on taps (or flaps) is available in Wolfram and Johnson [3], Orion [4], Mills [5], Laver [6], Language Files [7], Herd et al. [8], Huffman [9], Broadbent [10], and Warner et al. [11].

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

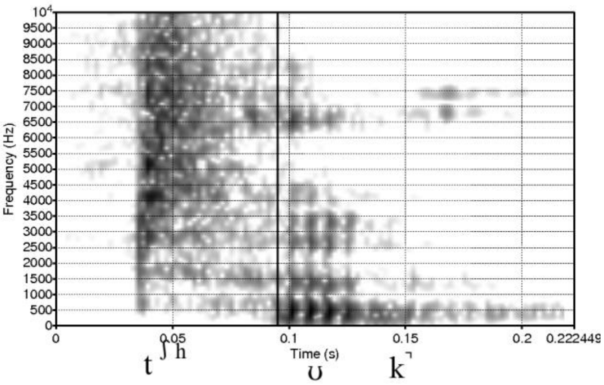


Fig. 1. Canonical /t/ with fricative and aspiration phases in *took*.

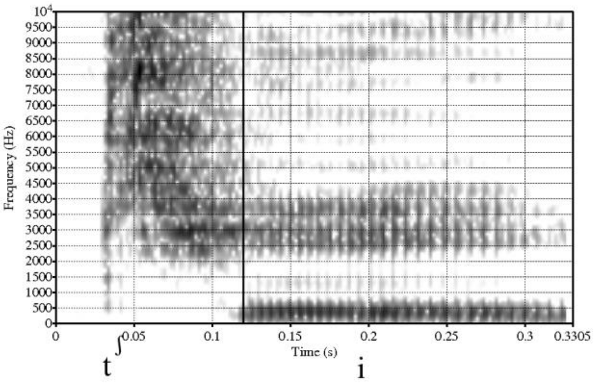


Fig. 2. Canonical /t/ with fricative phase in *tea*.

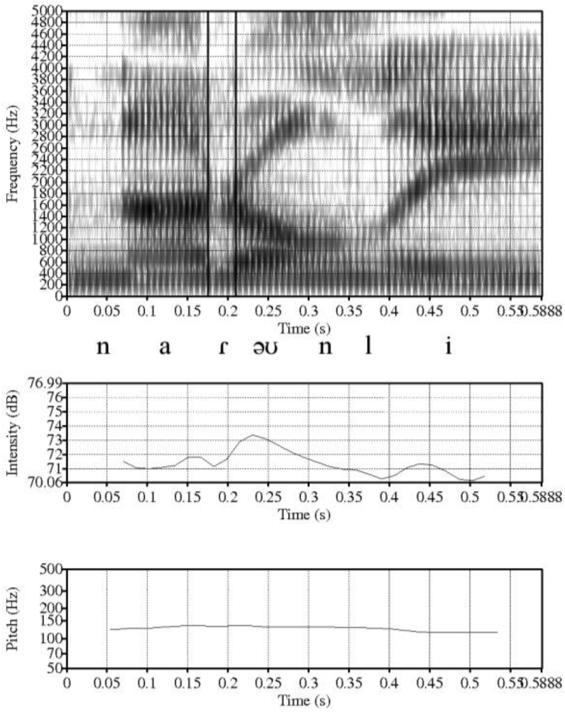


Fig. 3. Tap in *not only*.

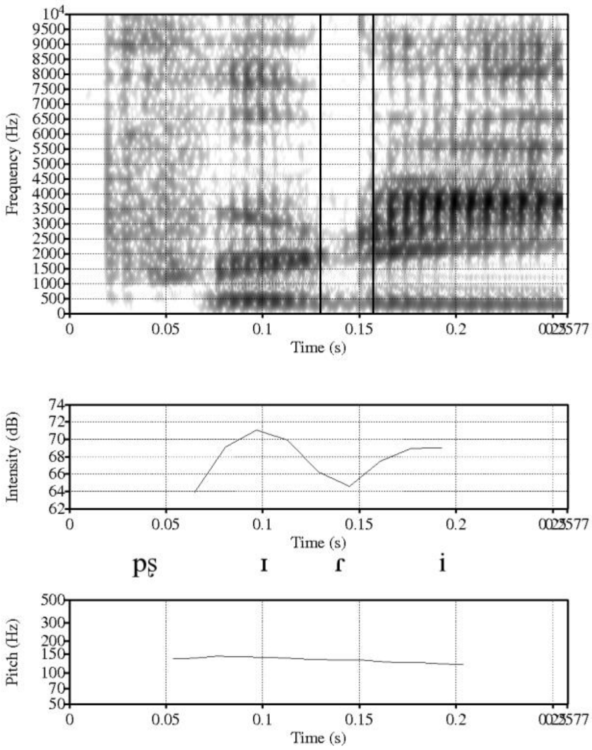


Fig. 4. Tap in *pretty*.

The following conditions make for glottal burst:

- word-internal /t/ + /l/, /n/, /m/;
- word-final /t/ + any sonorant except for /ŋ/.

Glottal burst is a part of glottalization phenomenon which includes: 1) voiced and voiceless implosives (canonical locus is preserves, glottal occlusion is added, no audible release is noticed), 2) glottal stop (canonical locus is substituted by glottal stop, no audible release), 3) glottal burst (canonical locus is substituted by glottal stop which is released) [12]. For more information on glottalization see Firth [13], Cruttenden [14], Laver [15], Ladefoged [16], and Broadbent [10].

These conditions are meant for a user who possesses basic skills of transferring letters into sounds. There is quite a number of issues that are taken for granted and need no explanation for a native American English speaker or an American English learner who can at least read properly. People easily manage a number of things that computer does not naturally possess skills for. The ones crucial for tap and glottal burst are:

- mute word-final “e” in words like *hate, definite* etc.;
- mute word-initial “h” in words like *heir, honor* etc.;
- groups of letters like “ough”, “augh”, “eigh” that contain a consonant letter but make one single vowel sound in words like *bought, daughter, height* etc.;
- mute “ai”, “e” in words like *certain, written* etc.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

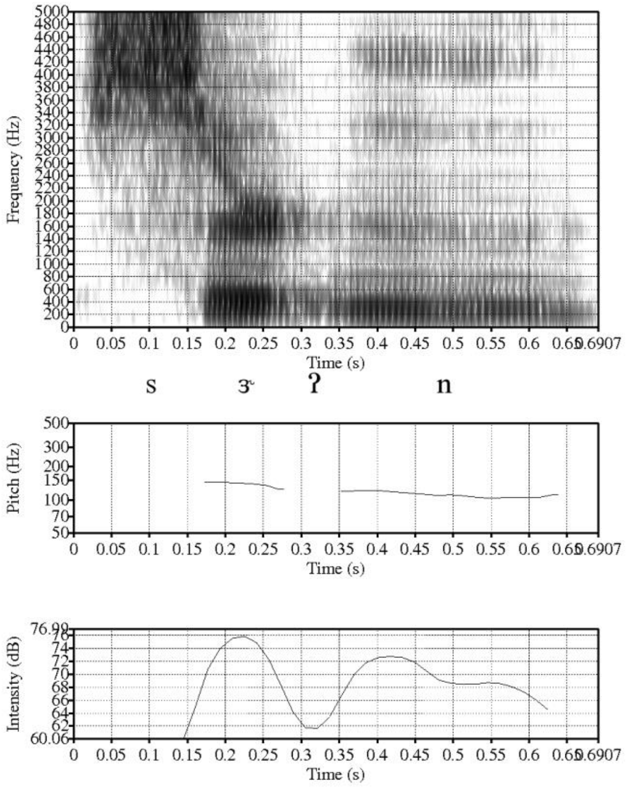


Fig. 5. Glottal burst in *certain*.

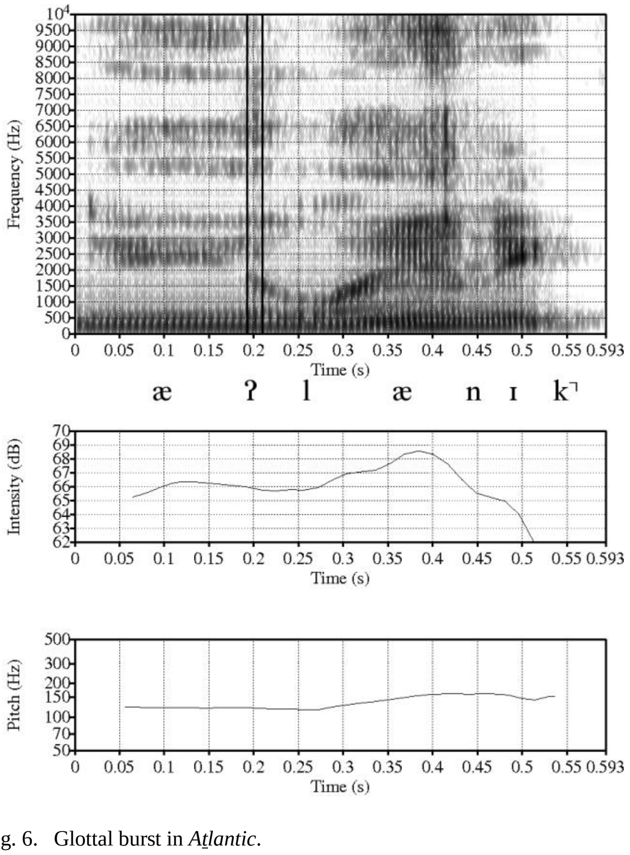


Fig. 6. Glottal burst in *Atlantic*.

To these issues word-stress and sentence-stress should be added as far as a person, who speaks decent English, knows where to put stress and the computer does not.

Issues, like the ones mentioned above, need to be programmed properly otherwise it will lead to inadequate result of the program performance that would distort real-speech pattern.

## III. Experiment

### A. Tools

For transcription the symbols of the International Phonetic Alphabet (IPA) are used. For correct visualization of source text program and the results of its execution a font with unicode support is needed. Therefore Doulos SIL font was used.

For programing Free Pascal is used. The source text was typed in Geany editor. Standard library of Free Pascal was used to work with text variables. Version 2.6.2 (2015) that supports unicode was used [17]. The program written in this version can by compiled in any further versions of Free Pascal. The program compilation and execution were performed in Linux.

### B. The program

#### 1) Framework and conflicting cases

The program operates within the following framework (see Fig. 7) that is very universal and can be suitable for any phonetic modifications, not only for tap and glottal burst.

We view some text where we wish to transcribe certain connected speech phenomena as an input file. As far as a number of phenomena occur on the word boundary, we cannot focus only on what is going on word-internally. Therefore we have to deal with every current word looking inside and out. This includes the characteristics of the previous word, the following word and non-alphabetic symbols between those words, particularly punctuation or symbols for pauses or any other symbols that might be deliberately added during text processing. Words and fillers are variables that are not subjected to any changes. They are analyzed according to certain criteria that have been targeted. For instance, if we target taps then the program needs to find out if there are certain letters in certain positions that presuppose tapping.

There are two things that we need to keep in mind. First, if talking about «t» letter we know that it can be transcribed by one or more than one phonetic symbol, e. g. tap [ɾ], weak voiceless allophone [t], glottal burst [ʔ], glottal stop [ʔ˺], canonical alophone [t$^h$]. Second, there are so called conflicting cases. These are where formally one and the same position might require different allophones, like, for example word-final intervocalic «t» with the next «u» vowel letter: tap in *that up* but glottal burst in *that usual* (see Fig. 8). While processing the word *that* an additional variable *sIsChange* is introduced. Initially it has «-» for every symbol: «----».

After the checking procedure in the final position in this word that has t-letter the program changes «-» into «+». After confirming the correspondence to the condition <Vowel>+t+<end-of-word>+<u><n><i> the variable *sIsChange* will be interpreted as«---+», the variable

ISSN (Online): 2203-1731

*aChangeStr[4]* will be assigned ʔ symbol. After that processing will continue but further processing will not change the symbol that has been previously assigned.

Another example of the same sort would be postvocalic word-final «t» with the next «h» consonant letter that in most instances works for glottal stop but in case of mute «h» (*honor* etc.) makes for tap.

### 2) *Regular expressions*

We use standard regular expressions that are provided with Free Pascal. This enables to decrease the number of if-operators and by this simplify the program structure (see Fig. 9).

### 3) *Two blocks of the program: tap and glottal burst*

The program deals with word-final and word-internal taps separately (earlier we demonstrated a pilot version of this block as a separate program [18]). Fig. 10 shows the algorithm of word-final tap processing. Fig. 11 shows the algorithm of word-internal tap processing.

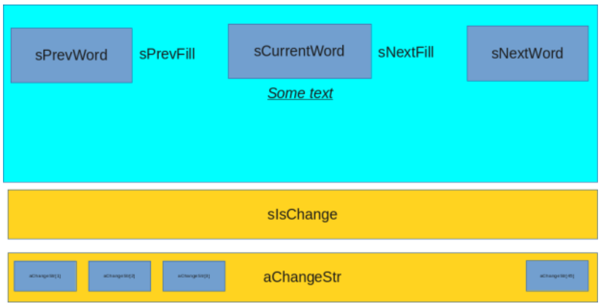


Fig. 7. Framework for automatic transcription of connected speech phenomena.

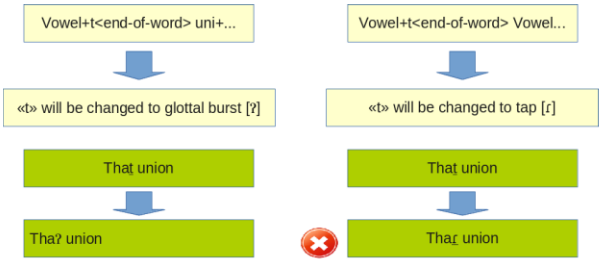
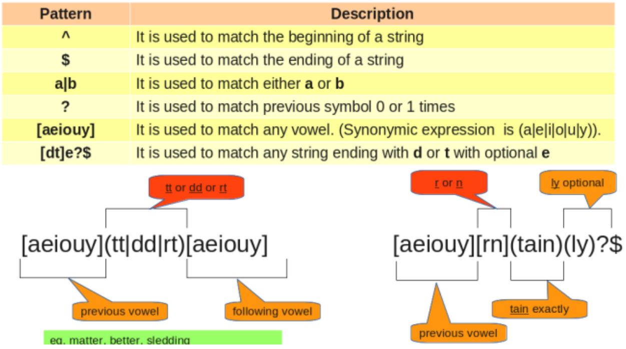

Vowel+t<end-of-word> uni+...
«t» will be changed to glottal burst [ʔ]
That union
Thaʔ union
Vowel+t<end-of-word> Vowel...
«t» will be changed to tap [ɾ]
That union
Thaɾ union

Fig. 8. Conflicting cases: that + «u».



| Pattern | Description |
|---|---|
| **^** | It is used to match the beginning of a string |
| **$** | It is used to match the ending of a string |
| **a\|b** | It is used to match either **a** or **b** |
| **?** | It is used to match previous symbol 0 or 1 times |
| **[aeiouy]** | It is used to match any vowel. (Synonymic expression is (a\|e\|i\|o\|u\|y)). |
| **[dt]e?$** | It is used to match any string ending with **d** or **t** with optional **e** |

Fig. 9. Regular expressions.

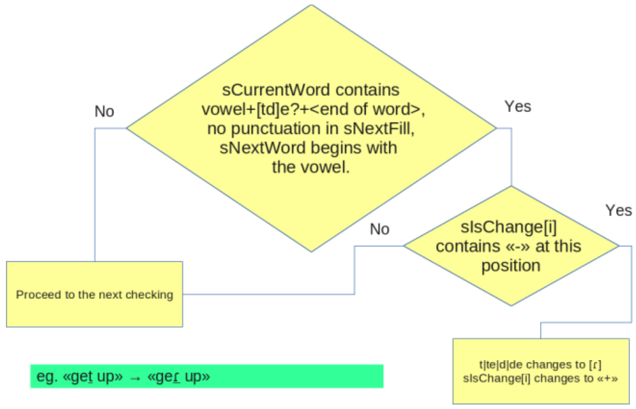

sCurrentWord contains vowel+[td]e?+<end of word>, no punctuation in sNextFill, sNextWord begins with the vowel.
No
Yes
sIsChange[i] contains «-» at this position
No
Yes
Proceed to the next checking
t|te|d|de changes to [ɾ] sIsChange[i] changes to «+»
eg. «get up» → «geɾ up»

Fig. 10. Example of word-final tap processing.

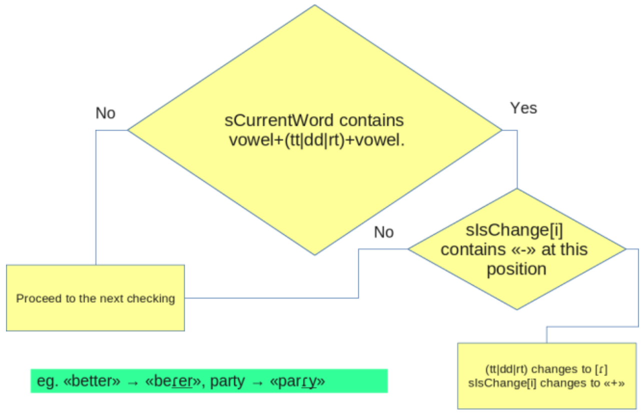

sCurrentWord contains vowel+(tt|dd|rt)+vowel.
No
Yes
sIsChange[i] contains «-» at this position
No
Proceed to the next checking
(tt|dd|rt) changes to [ɾ] sIsChange[i] changes to «+»
eg. «better» → «beɾer», party → «paɾy»

Fig. 11. Example of word-internal tap processing.

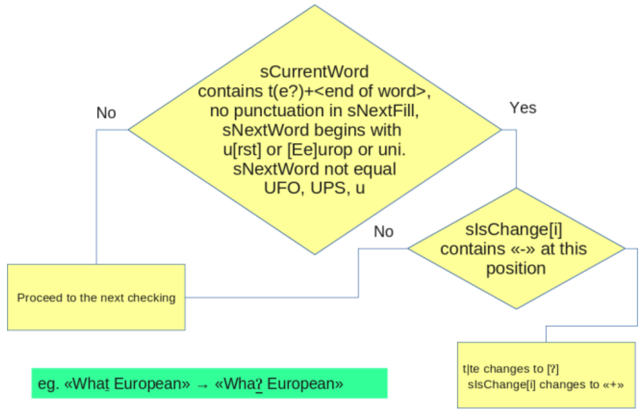

sCurrentWord contains t(e?)+<end of word>, no punctuation in sNextFill, sNextWord begins with u[rst] or [Ee]urop or uni. sNextWord not equal UFO, UPS, u
No
Yes
sIsChange[i] contains «-» at this position
No
Proceed to the next checking
t|te changes to [ʔ] sIsChange[i] changes to «+»
eg. «What European» → «Whaʔ European»

Fig. 12. Example of word-final glottal burst processing.

Figures 12 and 13 demonstrate the algorithm for correspondingly word-final and word-internal glottal burst processing. Both blocks operate within 90% accuracy. In the tap block the unattended issue is word-initial intervocalic tapping that occurs for *to* as particle, preposition, or prefix in word sequences like *so to speak*, *go to college*, *be together* etc. Those cases are not very frequent but quite stable for tapping in American English speech flow and need to be programmed

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

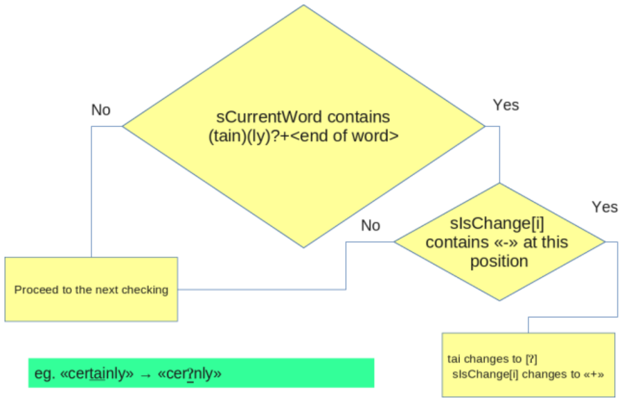

No
sCurrentWord contains (tain)(ly)?+<end of word>
Yes
No
sIsChange[i] contains «-» at this position
Yes
Proceed to the next checking
tai changes to [ʔ]
sIsChange[i] changes to «+»
eg. «certainly» → «cerʔnly»

Fig. 13. Example of word-internal glottal burst processing.

properly. Glottalization block is to be considerably enlarged: glottal stop and implosive parts are to be added to glottal burst part. The first one will present no problem: just a set of char and strings enumerating consonants of a different than fore-lingual articulator are needed: f, v, p, b, k, g and, additionally, ph, qu strings.

The second part is a challenge because it is not easy to accurately predict pauses. Punctuation can be helpful (full stops, exclamatory and question marks, dashes, colons and semicolons, and at times – commas), but a certain number of pauses might be located in places other than punctuation marks. Such pauses must be tagged during text preprocessing.

## IV. Conclusion

In the present paper we pointed out the necessity to make a computer model predicting various connected speech phonetic phenomena and, using the example of taps and glottal bursts that are regular in American English, showed how it might look like if we use Free Pascal language. The number of blocks each corresponding to a certain phonetic phenomenon can be easily enlarged; each block itself can be enlarged as well to reflect the majority of conditions and increase allophone predictability in the output file.

The same kind of programming can be done for Canadian and Australian English where taps are almost as frequent and stable under the conditions described above as in American (about taps/flaps in Australian English see e.g. Malcolm [19], Cox [20]). Unfortunately tapping and glottalization cannot be modeled reliably for British English due to changing standard [21], [22] and high degree of free variation in the condition described above: for instance, one word-final intervocalic position presupposes several allophones with almost no predictability of the pattern in British and only one allophone in American with more than 90% predictability [23].

## Acknowledgment

We express our grateful thanks to ICITA Conference Chair Dr. David Tien and the Rector of the Amur State University Dr. Andrei D. Plutenko who sponsored the delivery of this paper to ICITA (July 2016).

## References

[1] S. Greenberg, H. Carvey, L. Hitchcock, and S. Chang, “Beyond the Phoneme: A Juncture-Accent Model of Spoken language,”. Proc. of the Human Language Technology Conf. (HLT – 2002), pp. 36–43, San Diego, California, 2002.

[2] P. Boersma and D. Weenink,. “Praat: Doing phonetics by computer,” (Version 5.4.15) [Computer Program]. Retrieved November 30, 2015, from <http://www.fon.hum.uva.nl/praat/>.

[3] W. Wolfram and R. Johnson, Phonological Analysis. Focus on American English, Washington D. C. : Prentice Hall Regents, 1982, pp. 19–20.

[4] G. F. Orion, Pronouncing American English. Sounds, Stress and Intonation, New York, 1985, p. 199.

[5] C. Mills, American Grammar. Sound, Form and Meaning, New York, 1990, p. 119.

[6] J. Laver, Principles of Phonetics, New York : Cambridge Univ. Press, 1994, p. 218.

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

[8] W. Herd, A. Jongman, and J. Sereno, “An acoustic and perceptual analysis of /t/ and /d/ flaps in American English,” J. Phonetics, vol. 38, pp. 504–516, 2010.

[9] M. K. Huffman, “Segmental and Prosodic Effects on Coda Glottalization,” J. Phonetics, vol. 33 (3), pp. 335–362, 2005.

[10] J. M. Broadbent, “t-to-r in West Yorkshire English,” English language and linguistics, vol. 12 (1), pp. 141–168, 2008.

[11] N. Warner, A. Fountain, and B. V. Tucker, “Cues to Perception of Reduced Flaps,” JASA, vol. 125 (5), pp. 3317–3327, 2009.

[12] S. V. Androsova, “Allofonnoe var'irovanie soglasnyh fonem v spontannoj rechi (jeksperimental'no-foneticheskoe issledovanie na materiale amerikanskogo varianta anglijskogo jazyka) [Allophonic variation of consonants in spontaneous speech (experimental study based on American English)],” PhD Dis. St.-Petersburg State University, Blagoveshhensk, 2001.

[13] J. R. Firth, Papers in Linguistic 1934–1951, London : Oxford Univ. Press, 1957, p. 124.

[14] A. Cruttenden, Gimson's pronunciation of English, Edward Arnold Limited, 2001, p. 168.

[15] J. Laver, Principles of Phonetics, New York : Cambridge Univ. Press, 1994, p. 171.

[16] P. A Ladefoged, “Course of Phonetics,” 2nd ed. Singapore: Cengage Learning, 2006, p. 60.

[17] Free Pascal. URL: http://www.freepascal.org

[18] S. V. Androsova and E. Yu. Androsov, “Automatic transcription of taps in American English,” Proc. of the 2nd International Conference "Phonetics without borders" [ed. by S. Androsova]. Amur State University. Blagoveshchensk, pp. 10–16, 2015.

[19] K. Malcolm, Phasal Analysis: Analyzing Discourse through Communication Linguistics, New York, 2010.

[20] F. Cox, Australian English pronunciation and transcription, Cambridge University Press, 2012.

[21] A. H. Fabricius, “T-glottaling. Between Stigma and Prestige: A Sociolinguistic Study of Modern RP,”. Copenhagen, 2000.

[22] J. C. Wells, Accents of English, New York : Cambridge University Press, 1982.

[23] S. V. Androsova, V. G. Karavaeva, “Odnoudarnyje allofony v amerikanskom i britanskom variantah anglijskogo jazyka [Flaps / taps in American and British English],” Teoreticheskaja i prikladnaja lingvistika [Theoretical and Applied linguistics], vol. 1(2), pp. 5–20, 2015.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731