2007-5

DITCall-Slow: Slowing Native Speech for Language Learners

Dermot Campbell
Dublin Institute of Technology

Ciaran McDonnell
Dublin Institute of Technology

Yi Wang
Dublin Institute of Technology

Marty Meinardi
Dublin Institute of Technology

Bunny Richardson
Dublin Institute of Technology

See next page for additional authors

2007-01-01

DITCall-Slow: slowing native speech for language learners

Dermot Campbell
Dublin Institute of Technology

Ciaran McDonnell
Dublin Institute of Technology

Yi Wang
Dublin Institute of Technology

Marty Meinardi
Dublin Institute of Technology

Bunny Richardson
Dublin Institute of Technology

See next page for additional authors

Authors
Dermot Campbell, Ciaran McDonnell, Yi Wang, Marty Meinardi, Bunny Richardson, and Charles Pritchard

DITCall-Slow
Slowing Native Speech for Language Learners

EdTech 2007

Dermot Campbell, Ciaran McDonnell, Wang Yi, Marty Meinardi, Bunny Richardson and Charlie Pritchard

Dublin Institute of Technology

Introduction

It is a common experience of many learners of a foreign language that native speakers (NSs) of that language speak too quickly for them to understand or imitate. Slowing down a segment of speech with older technology results in the familiar deepening of the voice as the pitch drops as well. The result is unpleasant and not particularly instructive. The ***DITCall-Slow*** tool slows recorded speech without tonal distortion, so that the learner has – literally – more time to hear what was said by the NS and, especially at slower playback speeds, can attend to the manner in which the sequence was spoken. While the tool is currently being commercialised for the study of English as a foreign language (EFL), it can be applied to any spoken language. The current paper examines why the technology works and its place in modern language learning.

Background

DITCall-Slow was funded in 2001 by Enterprise Ireland under the Informatics Programme. The objective of the project was the development of a digital interactive package to assist foreign learners of English to enhance their listening and speaking skills in self-study. The main feature is a unique variable slowdown facility for speech recordings allowing students to capture details in native, natural spoken English which might otherwise have been lost in the flow of native-to-native speech. While time-scaling algorithms have been in use for music for quite some time, ***DITCall-Slow*** has implemented and optimised an algorithm for use with human speech (Donnellan and Coyle (2004)[1]).

Speech Flow

Unlike writing, speech does not consist of neatly separated individual words, but is rather a continuous flow of vowels – interrupted by consonants – to which the NS listener assigns meaning. The words assigned by the listener are not actually ***in*** the speech, but are labels provided in an attempt to re-construct the speaker's utterance plan.

Figure 1 Written flow = speech flow

In order to better understand what happens in a native speech flow, it is useful to compare spoken communication with written communication. In relaxed dialogue with other native speakers, NSs tend to expend the minimum of energy necessary to ensure their interlocutors can properly understand their utterance. This can be compared with a hastily scrawled message or a signature (see Figure 1). Here the individual letters cannot be made out as the writer is relying on the overall

outline of the written sequence to effect communication rather than each individual letter. This corresponds in speech – seen from the perspective of an idealised utterance – to a flow of phonemes where some of the speech units have been elided, some eroded and the boundaries between some blurred by coarticulation.

Figure 2 More careful handwriting = more articulate speech

If it is necessary for the reader to make out every letter (as in spelling a name), then the writer will make the extra effort to write in such a manner that each letter is intelligible. It is still handwriting which is produced (see Figure 2), but the constituent elements of this written communication are now clearly distinguishable. This is analogous to 'second pass' speech production, where the speaker takes the listener's perspective into consideration and produces an utterance which consciously sets out to be more easily intelligible.

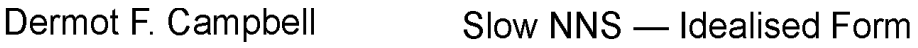

Figure 3 Print = hyper-articulated speech

Finally there is the printed word (Figure 3) whose form has been standardised and is the equivalent of broadcast quality speech designed to reach out to the widest audience possible. Learners of foreign languages tend to have internalised such idealised models, which partially explains their difficulties when they find themselves in a NS-NS environment.

For decades language teaching has avoided the difficulties of teaching pronunciation (Campbell et al (2004)[2]). It is labour intensive, difficult and invasive. DIT's ***DITCall-Slow*** technology aims to overcome these difficulties by making natural, relaxed NS-NS speech available as a learning tool. EFL learners will in future have at their disposal a technology which will allow them to study NS speech production, concentrating on ***how*** something was said rather than merely on ***what*** was said. Most NS recordings in language learning situations are used as comprehension exercises, where the listener is asked content-relevant questions and the native speech itself is not further exploited.

The Need to Use Natural Language

It is essential for learners of English to be exposed to 'real' conversational English. Very often, language learning material does not manage to bridge the gap between 'classroom language and language in use'. The reason why 'real' everyday conversation should be part of language learning material is because it is the type of language where one is not on one's 'best linguistic behaviour' (Crystal (1981)[3]). Learners need to be able to identify with that type of language in order to become part of the target language community. Affectively, authentic input also increases motivation and can serve to overcome the initial cultural strangeness in foreign language learning. It is felt that not exposing learners to authentic NS speech leaves the students in a very vulnerable position should they ever visit the country of the target language, and builds false hope of being able to communicate effectively in the L2. NSs use certain forms of speech that lets the interlocutor know that there is a shared deixis and it is this type of authentic language use which is most frequent and most natural in NS speech and which should therefore be included in language learning material.

Cognitively, authentic materials provide the necessary context for relating form to meaning (decoding) appropriately in the language acquisition process. When students are properly prepared, authentic oral and written materials have a positive perceived effect on comprehension and student satisfaction, and explicit attention to the development of listening skills improves listening

comprehension at all levels of instruction with no negative effect on grammar, vocabulary, or oral skills (Bacon and Finneman (1990)[4] and Herron and Seay (1991)[5]). The pedagogical value of the use of authentic texts is that they become a tool to help the learner come to grips with natural language and its communicative rules, facilitating the learner's inferring skills and improving shared knowledge with NSs (Breen (1985)[6]).

Emulation of authentic spoken language is moreover essential to give the learner the opportunity to acquire some of the idiosyncrasies of native natural speech, in order to better fit in to the culture of the target language. Non-native speakers of English (NNSs) often display none of the false starts, hesitations, grammatical mistakes and unfinished sentences that occur in NS speech. Such unnatural perfection in NNS speech could affect a NS's perception of the learner in a negative way, disabling smooth communication (Brown and Yule (1983) [7]). The use of authentic, NS to NS spoken material attempts to facilitate the following:

- understanding of the English language as it is spoken locally by NSs
- acculturation into the L2 community through the use of authentic audio-visual material which aims to present the learner with background information to the L2 culture where this is appropriate to the learner's needs
- improvement of the learner's speaking ability and language processing skills by providing opportunities for language awareness and emulation.

How Technology can help

In the same way that fast photography makes the subtleties of a tennis serve or a golf swing perceptible to the learner tennis player or golf tyro, ***DITCall-Slow*** will 'allow time to be spent with the signal', to use Richard Cauldwell's term (Cauldwell 2002)[8]. Studying the intonation patterns of slow**ED** speech (as opposed to slow speech) and repeating the model at increasing speeds until a NS rate is reached, enables the learner to acquire a NS delivery, or at the very least render NS speech more intelligible.

DIT internal research (Meinardi 2006[9]) has produced evidence that students prefer a slowed-down speed of 80% as this rate gives the user sufficient time to process the NS speech signal while still retaining a high degree of naturalness in the recording. Slower speeds, however, while sounding 'drunk' or 'exhausted', highlight the natural prosody of the speech act and are useful for the learner to study native-speaker intonation patterns.

The programme is particularly suited to students preparing for an extended stay in a country of the language studied. In the case of English, for example, students could choose to study the NS variety spoken in Australia, Great Britain, Ireland, North America, South Africa etc. All that is required is a sample of recorded speech in WAV (or any other uncompressed) format of the language variety required. Asian learners in particular can benefit from ***DITCall-Slow***, as it allows them to listen to NS recordings at normal speed or slowed down to a practical limit of 40%, steplessly. By imitating the slowed speech at low speeds the learner can more easily follow NS intonation patterns, imitate the model speech and gradually speed up the playback and imitation until native speeds of production are attained. Since this is done on a PC or laptop, practice can be undertaken in an unthreatening and non-intimidating manner. Asian students in particular, who tend to be slow in uttering a sentence of spoken English until they feel they have the pronunciation perfect, appreciate the ability for autonomous and private practice.

Example of Advantages of Slowed Speech - English

English belongs to a minority of language which have a high number of vowels (12 pure vowels, 8 diphthongs) and whose stress patterns dictate that many unstressed syllables experience a centralising effect in relaxed or rapid speech. This poses particular difficulties for learners of English as a foreign language. Spanish or Japanese learners, whose mother tongues have only 5 vowels, find the fine gradations of English vowels difficult to hear and imitate. The most common sound in NS English, for example is 'schwa' – the 'to' sound in 'today' – which is a reduced, centralised, unstressed vowel which does not exist in many languages and which is very difficult to catch when listened to at full speed.

At slower playback speeds some idiosyncrasies of spoken speech become apparent which are not evident at normal speaking rates. In an internal DIT recording the speaker says: ' ... that must be quite difficult'. Only when the recording was slowed to 60% did it become apparent to the speaker herself that she had, in fact, said 'diff-cult', not the perceived 'diffi**c**ult'. The fact that NSs tend to say 'te**M** balloons' instead of 'te**N** balloons' in informal, relaxed speech can only be made obvious by use of the slow-down algorithm.

Example of Advantages of Slowed Speech - Chinese

More than half of all human languages are tonal languages—that is to say that words are distinguished by the tone used on the syllabic vowels. The consonant-vowel combination in Mandarin Chinese 'ma' for example can refer to such disparate things as 'mother', 'hemp', 'horse' or 'to scold' depending on whether the tone is high-steady, rising, dropping-rising or dropping. In order to ensure lexical integrity, it is necessary to maintain the correct, distinguishing tone pattern.

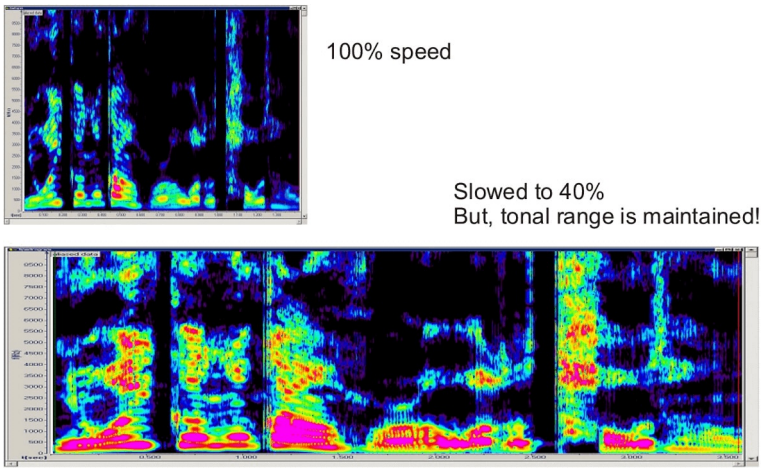

Figure 4 Slow-down makes tonal contours more obvious

DIT has recently introduced the teaching of Chinese as a foreign language to students on its International Business and Language programme. Initial research undertaken by the Digital Media Centre has indicated that ***DITCall-Slow*** could be of invaluable assistance in helping students of Chinese to come to grips with the difficulties of mastering the tone system of that language. At the slowest practical speeds, 40% for example, the listener's attention is drawn away from ***what*** is said to ***how*** the utterance was made. Under normal conditions we listen to speech in order to recover the speakers utterance plan, i.e. ***what*** s/he intended to say. Due to the somewhat disembodying effect of the slower playback speeds, the listener's attention is drawn to the prosody of the speech act—to the rhythm and intonation pattern of the utterance. In English this can highlight grammatical and expressive content. In Chinese it makes the lexical intonation patterns accessible to speakers of languages where intonation patterns operate at phrase level.

DITCall-Slow as part of an Integrated Speech Technology Platform

DITCall-Slow can help students to hear more accurately not only what has been said but how it has been said. Consequently, students who use ***DITCall-Slow*** may be able to learn to make more

accurate utterances in the language they are studying. Having thus made natural speed, native-speaker language available to learners, it would also be helpful if they had a program to guide them on what utterances to make and explicitly to show them how to improve their attempts at imitating these utterances.

The ***Articulate!*** project, funded by Enterprise Ireland under the Proof-of-Concept Program was designed to do just that. The aim of this project was to create an original computer assisted language learning (CALL) tool for students of English as a foreign language, with the focus on teaching the pronunciation of English vowels. The novelty of the program is the incorporated approach of providing automatic, computer-generated feedback on students' attempts to produce vowels as required in the context of the exercises. ***Articulate!*** gives accurate and meaningful graphical feedback to students. The feedback consists of an intuitive animated graphical interface that shows the student where his attempt to produce the vowel is in relation to the target vowel requested.

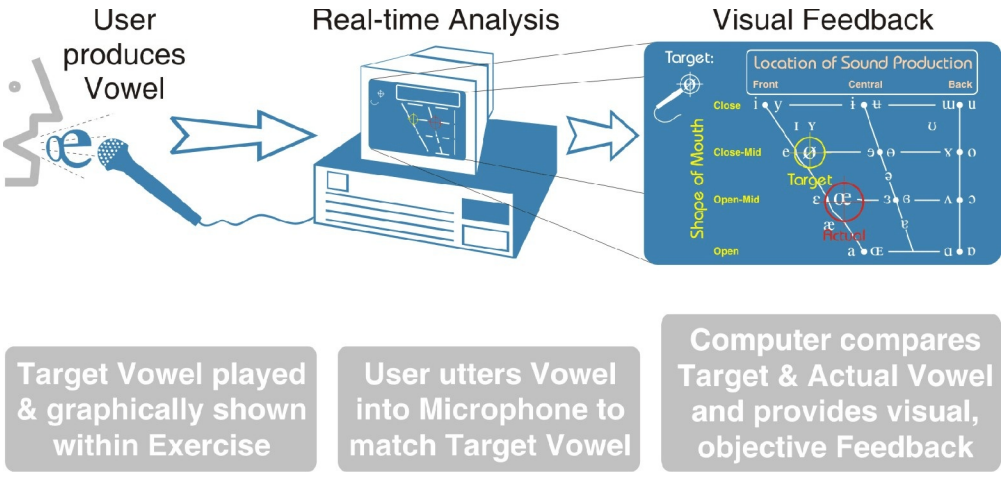

Figure 5 ***Articulate!*** gives graphical feedback on vowel production

Figure 5 gives an overview of the functional structure. The user is asked to say something, such as to pronounce the 'i' in 'bit'. The software captures the user's utterance and maps it in real-time, via a mapping mechanism, to a vowel quadrilateral, thus giving visual feedback to the user. We have experimented with and evaluated a number of potentially suitable mapping mechanisms. There were issues to be overcome regarding the stability and consistency of the vowel uttered. An exemplar exercise for language learning was developed. Other potential applications of the system in language learning have been identified and the system has also been found to have potential use in remedial situations for those with vocal, speech or language disorders.

References

1. **Donnellan, O., Jung, E. and Coyle, E.** 2003. *Speech-Adaptive Time-Scale Modification for Computer-Assisted Language Learning*. ICALT 03, Athens
2. **Campbell, D., Richardson, B. and Meinardi, M. et al.** 2004. *Naturally speaking but slow*. IATEFL PronSIG, Reading, UK
3. **Crystal, D.** 1981. Directions in Applied Linguistics. London: Academic Press (p.90)

4. **Bacon, S. and Finnemann, M.** 1990. A study of the attitudes, motives, and strategies of university foreign language students and their disposition to authentic oral and written input. Modern Language Journal, 74, (pp.459-473)
5. **Herron, C. and Seay, I.** 1991. The effect of authentic aural texts on student listening comprehension in the foreign language classroom. Foreign Language Annals 24, (pp.487-495).
6. **Breen, M.P.** 1985. Authenticity in the Language Classroom. Applied Linguistics, Vol 6, No.1, (pp. 60-70).
7. **Brown, G. and G. Yule**. 1983. Teaching the spoken language. An approach based on the analysis of conversational English. New York: Cambridge University Press (p.21).
8. **Cauldwell, R.** 2002. *Phonology for Listening: Relishing the Messy*. TESOL, Salt Lake City, Utah.
9. **Meinardi, M.** 2006. *The Use of 'Real' English in Language Learning*. PhD Thesis, Dublin Institute of Technology, Ireland