The Big Australian Speech Corpus (The Big ASC)

Michael Wagner[1], Dat Tran[1], Roberto Togneri[2], Phil Rose[3], David Powers[4], Mark Onslow[5], Debbie Loakes[6], Trent Lewis[4], Takaaki Kuratate[7], Yuko Kinoshita[1], Nenagh Kemp[8], Shunichi Ishihara[3], John Ingram[9], John Hajek[6], David Grayden[6], Roland Göcke[1], Janet Fletcher[6], Dominique Estival[7], Julien Epps[10], Robert Dale[11], Anne Cutler[7], Felicity Cox[11], Girija Chetty[1], Steve Cassidy[11], Andy Butcher[4], Denis Burnham[7], Steven Bird[6], Cathi Best[7], Mohammed Bennamoun[2], Joanne Arciuli[5], Eliathamby Ambikairajah[10]

[1]University of Canberra, Australia
[2]University of Western Australia, Australia
[3]Australian National University, Australia
[4]Flinders University, Australia
[5]University of Sydney, Australia
[6]University of Melbourne, Australia
[7]University of Western Sydney, Australia
[8]University of Tasmania, Australia
[9]University of Queensland, Australia
[10]University of New South Wales, Australia
[11]Macquarie University, Australia

bigasc@list.uws.edu.au

Abstract

Under an ARC Linkage Infrastructure, Equipment and Facilities (LIEF) grant, speech science and technology experts from across Australia have joined forces to organise the recording of audio-visual (AV) speech data from representative speakers of Australian English in all capital cities and some regional centres. The Big Australian Speech Corpus (the Big ASC) will provide a standard recording setup and a collaboratively-designed elicitation protocol to create a corpus of AV speech data incorporating annotations and metadata, accessible via a centralised storage facility. The Big ASC infrastructure will provide a significant boost to research in speech science and human communication in Australia.

Index Terms: speech corpus, audio-visual data, Australian English, annotation, data storage.

1. Introduction

Speech Science, in addition to its wider linguistic contribution beyond phonetics, such as to sociolinguistics and psycholinguistics, underpins applications in Speech Technology, such as Text-To-Speech (TTS) synthesis, Automatic Speech Recognition (ASR), speaker recognition and forensic identification, talking head interfaces and hearing prostheses. Paramount among the basic infrastructure needs of Speech Science is a large speech corpus collected and annotated to accommodate multiple applications, and stored to allow easy and immediate access. As yet, no such corpus exists for Australian English, mainly because such speech corpora are difficult and expensive to create. Although data for a variety of languages can be obtained from various sites such as the LDC, these corpora are either not large enough or not systematically designed to serve the interests of a variety of users. On the other hand, while large systematically designed speech corpora cannot be bought off the shelf and they are extremely expensive, labour-intensive and time-consuming to construct, their collection does not constitute research in itself, so the usual funding avenues are unavailable.

Recognising the importance of such a corpus, speech science and technology experts from across Australia banded together and in 2010, MARCS Auditory Laboratories, with partners Macquarie University, Flinders University, University of Melbourne, University of Canberra, University of New South Wales, University of Queensland, University of Sydney, University of Western Australia, Australian National University, University of Tasmania, the Max Planck Institute for Psycholinguistics and the Australasian Speech Science and Technology Association were awarded an Australian Research Council Linkage Infrastructure, Equipment and Facilities grant for a project entitled '*The Big Australian Speech Corpus: An audio-visual speech corpus of Australian English*' [1].

The Big ASC will establish the infrastructure for the collection of large quantities of AV speech data from many locations across multiple sessions per speaker. It will provide standardised recording equipment and protocols for data collection and annotation and will provide access to the annotated data and the meta-data via a centralised storage facility. The Big ASC will, as far as possible, take on board lessons learned in other large-scale corpus collection exercises such as those undertaken within the EU CLARIN and US American National Corpus projects. However the Big ASC is not only a large Australian speech corpus, it is also the only Australian corpus of audio-visual speech in the public domain, and it is designed to cater to various clients with a variety of interests.

2. Aims and goals of the Big ASC

The only publicly available Australian speech corpus is the 14-year-old ANDOSL database [2], which, although still relevant for research, is now outmoded due to its small number of speakers, non-representative sampling, and a single recording session per speaker. In addition, it is audio-only and has limited coverage of Australian English (AusE) variation. With the Big ASC, we are now in a position to establish a much

ISBN 978-0-9581946-3-1
Accepted after peer review of full paper

larger and more representative corpus of speakers from all over Australia, using more modern technology, and resulting in higher quality data. Over a period of time, 30 of the top speech scientists and technologists from 11 Australian universities have participated in the definition and design of the project, which builds on pre-existing collaboration across a range of networks, projects and associations. The data collection will be conducted at each of the 11 sites, with additional data collection in major regional centres.

The Big ASC infrastructure will provide a significant boost to speech research in Australia now and well into the future because it incorporates contemporary and rigorous design features. With a projected research lifespan of at least two decades, the Big ASC, once established, will have the potential to engender and enhance Australian efforts in a range of human communication and speech science areas:

- *Phonetics and linguistics*

A corpus such as the Big ASC is essential in order to examine variation in AusE over geographical areas [3], [4], [5], [6], [7]; ethnocultural and social background, and speech style [8]; changes to the language since the collection of the ANDOSL database [9]; and to provide greater access to information on speech production [10].

- *Psycholinguistics*

The Big ASC will have applications in projects on psycholinguistic models for word processing [11], [12]; young children's perception of phonetic variability and dialectal variation in spoken words [13]; the effect of pronunciation on written language [14]; and hearing training programs for children and adult users of cochlear implants [15], [16].

- *Engineering – Spoken Language Processing*

The Big ASC will support research projects in AV/ASR [17], [18], [19] and AV separation [20], [21], [22]; the development of real-time visual biometric systems and more robust systems for authentication or identification [23]; improved perception for cochlear implant sound processing [24], [25]; emotion recognition and/or detection for automated user interfaces and for talking heads [26], [27], [28]; and auditory-visual TTS synthesis [29]. The mapping between the auditory and visual components of the Big ASC will contribute to the development of smarter ASR and talking heads.

- *Language technology and computer science*

With data from the Big ASC, a range of interfaces will be enabled: ASR tailored for Australian English and its variety of accents and emotional tones/textures/expressions [30], speech dialogue management [31], [32]; AV user-centric/context-aware/ask-once/ask-nonce information retrieval and monitoring [33]; as well as web search and training products and guides based on grounded speech understanding [34], [35], [36].

- *Speech pathology*

While the current Big ASC will only collect a small set of disordered speech from stutterers, corpora of representative Australian speech are critical to disordered speech research generally, as typical speech provides a framework against which atypical speech can be investigated. Such corpora can contribute to the understanding of disorders and to the development of intervention strategies and devices [3], [38].

- *Forensic speech science*

Speech data collected across multiple sessions allows estimation of between- and within-speaker variability. This allows estimation of the strength of evidence with a Likelihood Ratio using Bayes theorem [39]. The Big ASC will be of great use in testing forensic speaker recognition approaches and conducting real-world casework, as well as identifying individuality in speaker behaviour [40], [41] and will work with a recently awarded Linkage Project grant (LP100200142) on reliable forensic voice comparison.

3. Infrastructure

The legacy of the Big ASC project will not only be the extensive set of annotated AV data, but also the recording infrastructure and elicitation protocols that will form the basis of further standardised data collection for Speech Science and Technology research in Australia and overseas.

3.1. Data Storage

A single server on the web will provide both a web-based view of the data and a 'shared drive' view allowing the data to be accessed from a desktop.

3.2. Hardware / Software

Standardisation in equipment and data collection procedures is essential to the success of the Big ASC. A Standard Speech Science Infrastructure Black Box (SSSIBB) and a Standard Speech Collection Protocol (SSCP) are to be used at each collection site by all research assistants to ensure that speech collection conditions are carefully controlled and documented.

- *Standard Speech Science Infrastructure Black Box*

The Standard Speech Science Infrastructure Black Box (SSSIBB) is designed to ensure consistency of data capture across all sites. The recording equipment comprises a headworn and desktop microphones, digital audio acquisition device, and a stereo camera. The data is all digitally captured to a computer. For the Map Task (see below) an additional headworn microphone and stereo camera are used. All these items are packaged together in a specially designed piece of hardware which serves as a reinforced box for transportation and folds out into a table with integrated shelves for recording. The SSSIBB ensures the same equipment is used at all locations, in the same configuration and with the same setup and spacing. The portable design allows the data collection to be conducted in any location and to be extended to new locations as needed.

- *Standard Speech Collection Protocol*

The Standard Speech Collection Protocol (SSCP) covers the requirements for software on the SSSIBB. It not only includes the drivers for all relevant devices but also a single application to control recording of audio and video data, storage to disk and management of sessions. At a minimum, the protocol will provide the following functions:

- collect session meta-data (subject id, location, etc.);
- perform calibration, to ensure that settings and conditions are known when recordings start, and to ensure that all components are set up correctly;
- synchronise capture of audio and video devices, writing data to disk in appropriate format to minimise risk of data loss;

- manage the recording session, allowing the RA to control timing, restart tasks, and store all recorded data in defined file structures;
- provide randomised prompts for read speech tasks, and record order of presentation with the recorded data;
- carry out backup copying of data and register session meta-data with central server at the end of each session.

3.3. Meta-data

A central meta-data store receives the meta-data entered as part of the recording process and associates it with audio and video data in the main file store. This will provide the main entry point for browsing/searching the corpus data, via a web interface to the meta-data store.

3.4. Annotations

The audio and video data will be annotated to varying degrees, one imperative of the project being that the management of the annotation data is key to making this corpus as useful as possible.

For recordings that are read (digits, words, read sentences, etc.) this consists of storing the start and end offsets of each word, while for longer unscripted recordings annotations, this will be a transcript of what is said aligned at the phrase or sentence level.

Basic level annotation, together with more detailed annotation, will be conducted by a central Annotation Team, following consistent principles and protocols for annotation. The Annotation Team will mark up aspects of dialogue, intonational, syntactic and rhetorical structure as appropriate, using variants of the Emu and ELAN tools to be interfaced with the shared annotation server, thus ensuring that the corpus is built collaboratively and consistently.

The annotation task will be automated as far as possible, first by performing forced alignment for the read speech, and second by using annotation tools allowing checking and inter-annotator comparison.

The annotation store will:
- be independent of any annotation tool or format;
- be able to accept and deliver annotation data in many formats;
- provide web based browsing and searching.

4. Data Collection

This will be the first ***Australian*** speech corpus to meet the demands of modern speech science. It will sample widely and appropriately from AusE, an increasingly important consideration given that AusE is becoming an influential new standard in East Asia [42]. Since audio-visual data have now become essential for many applications, e.g., ASR and speaker recognition, biometric applications, automated user interfaces, all the data will be AV-recorded.

4.1. Protocol

A variety of tasks appropriate for different applications will be completed by all speakers across three separate recording sessions. As literacy cannot be assumed, some variation of the protocol is necessary: sentences and word lists may be orally prompted, and the map task may be replaced by an alternative task such as storytelling. However, importantly, all the word-level and natural sentence-level material will be retained.

4.2. Type of Data

- *Demographics*

A comprehensive set of demographic, family and historical data will be collected from all speakers to document the regional, ethnocultural, educational, language, speech and hearing, and employment backgrounds

- *Isolated Words (IW)*

The IW data consists of 325 words comprising 3 word types:

a) a monosyllabic standard word set comprising the stressed vowels of AusE in the following standard contexts: hVd, hVt, hV, hVl, hVn;
b) an additional set of words selected to address specific AusE characteristics;
c) a set of polysyllabic words to address stress patterning.

- *Read Speech (RS)*

The RS data comprises a set of sentences that deliver connected speech in a standard format from every informant in the corpus. Informal speech is sacrificed for standardisation and phonetic sampling. However, valuable phonetically comparable cross-corpus data is achieved as the sentences are derived from the ANDOSL additional set of 50 sentences.

- *Digits*

A set of 44 four-digit strings will be provided for people to read. The specific purpose of these is to provide data for background databases for audio password systems

- *Interview*

The aim is to capture spontaneous, engaged, narrative talk – 'story telling' in the 'vernacular' speech style [43]. This style of speech is hard to capture in formal interviews, but can be achieved with skilful or empathic interviewers and with particular attention to facilitating conditions of elicitation. 'Story telling' is a natural task, appropriate for all speakers and speech communities; it is a task that is equally appropriate for all groups included in the scope of the project. Different topics will be suggested for discussion, subject to speakers' preferences.

- *Map Task*

We propose to design a new Map Task [44] for the elicitation of spontaneous interactive speech data. The new map task will be of interest to a broad range of research interests by virtue of careful design at various linguistic levels. By incorporating a range of linguistic features into the map design, we anticipate that the Big ASC Map Task will additionally provide corpus data of interest outside Australia.

- *Text reading and recounting*

A short newspaper style article will be provided for people to read. Then they will be asked to retell the same story but in their own words.

4.3. Amount of Data: Size and Distribution

Large speech corpora are essential in order to cover idiosyncrasies and variation [45]. For the Big ASC, speech will be collected from around 1000 speakers at 17 collection sites from every state and territory of Australia.

- *Multiple sessions (within-speaker variation)*

Each speaker will be recorded on three separate occasions to capture within-speaker variability over time.

- *Diversity (between-speaker variation)*

Representative sampling from 17 different sites and field locations will reflect *regional* (all states & territories) and *demographic* variation (with AusE recorded from speakers whose entire school education has been in Australia, thereby capturing both Australian born and migrants who arrived in Australia before school age). This group will include specific sampling of Indigenous Australians and will also include speakers from a variety of different ethnic backgrounds (e.g. from the Lebanese and Chinese communities)

5. Looking to the future

The Big ASC project is funded for a period of 12 months, and technically will be completed in 2011. During that period, we will design and build the SSSIBB; develop the SSCP; launch the website to advertise the data collection and recruit speakers; train RAs and run the data collection at 17 different sites; collect data from 1000 speakers; establish the data server for storage and access to the data; develop the annotation environment and annotation protocols; annotate half of the data recorded.

However, this is only the first step in building a truly Australian corpus. Available funding and time restrict the kinds of data we can collect for this project, even though the corpus will be much larger in detail and scope than anything that has ever been attempted for this country. Project members are focussed on maintaining collaborative efforts that underpin Big ASC, even after the initial LIEF funding ends. We hope the planned successful completion of the Big ASC in the form outlined above will facilitate national funding in the future that will allow this initial effort to be continued and expanded by collecting other useful data which cannot be covered in this initial phase, such as speech with a variety of *emotions* and degrees of *intactness* (disordered speech), speech in *noise*, and additional varieties of *ethnocultural* backgrounds, including ***indigenous*** creoles (NT and Kimberley Kriol, and Torres Strait Creole).

6. Acknowledgements

The Big ASC project (LE100100211) was developed with support from the Human Communication Science Network (HCSNet) and the Australasian Speech Science and Technology Association (ASSTA). Thanks are due to all the participating institutions (see list of authors) for making this project possible and also to Ms Amanda Reid, who collated a large amount of information for a former version of this paper.

7. References

[1] Burnham D, Ambikairajah E, Arciuli J,Bennamoun M, Best CT, Bird S, Butcher AB, Cassidy C, Chetty G, Cox, FM, Cutler A, Dale R, Epps JR, Fletcher JM, Goecke R, Grayden DB, Hajek JT, Ingram JC, Ishihara S, Kemp N, Kinoshita Y, Kuratate T, Lewis TW, Loakes DE, Onslow M, . Powers DM, Rose P, Togneri R, Tran D, and Wagner M. 2009. 'A Blueprint for a Comprehensive Australian English Auditory-Visual Speech Corpus'. Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus, ed. Michael Haugh et al., 96-107. Somerville, MA: Cascadilla Proceedings Project.

[2] Millar J, Dermody P, Harrington J & Vonwiller J 1990 'A national database of spoken language: concept, design, and implementation'. Proceedings of the International Conference on Spoken Language Processing (ICSLP-90). Kobe, Japan. http://andosl.anu.edu.au/andosl/ANDOSLhome.html.

[3] Butcher AR 2006 'Formant frequencies of /hVd/ vowels in the speech of South Australian females'. Proceedings of the 11th Australasian International Conference on Speech Science & Technology, Pp 449-453.

[4] Butcher AR 2008 'Linguistic aspects of Australian Aboriginal English'. Clinical Linguistics & Phonetics, 22, 625-642.

[5] Cox F & S Palethorpe 1998 'Regional variation in the vowels of female adolescents from Sydney' Proceedings of the 5th International Conference on Spoken Language Processing, ICSLP, Sydney, November 30-December 4.

[6] Cox F & S Palethorpe 2001 'The Changing Face of Australian English Vowels'. In Blair D & Collins P (eds.), Varieties of English around the World: English in Australia Amsterdam: John Benjamins Publishing. Pp. 17-44.

[7] Cox F & S Palethorpe 2004 'The border effect: Vowel differences across the NSW/Victorian border' in Moskovsky, C (ed.), Proc. 2003 Conference, Australian Linguistics Society.

[8] Ingram J 1989 'Connected speech processes in Australian English' in D Bradley, R Sussex, & G Scott (eds.), Studies in Australian English Aust. Linguistic Society. Pp 21-49.

[9] Cox F & S Palethorpe 2008 'Reversal of short front vowel raising in Australian English', Proceedings of Interspeech 2008, 22nd-26th September 2008, Brisbane, 342-345.

[10] Fletcher J, Grabe E & P Warren 2004 'Intonational variation in four dialects of English: The high rising tune' in Sun-Ah Jun (ed.), Prosodic typology Oxford: Oxford University Press. Pp. 390–409.

[11] Cutler A & D Carter 1987 'The predominance of strong initial syllables in the English vocabulary' Computer Speech & Language 2: 133-142.

[12] Cutler A 2005 'The lexical statistics of word recognition problems caused by L2 phonetic confusion' Proc. Eurospeech 2005, Lisbon, 413-416.

[13] Best CT, Tyler MD, Gooding TN, Orlando CB & CA Quann 2009 'Emergent phonology: Toddlers' perception of words spoken in non-native vs native dialects' Psychological Science, 539-542.

[14] Kemp N 2009 'The spelling of vowels in influenced by Australian and British English dialect differences' Scientific Studies of Reading 13: 53-72.

[15] Dawson PW, McKay CM, Busby PA, Grayden DB & GM Clark 2000 'Electrode discrimination and speech perception in young children using cochlear implants', Ear and Hearing 21:597-607.

[16] Mok M, Grayden, DB, Dowell, RC & D Lawrence 2006 'Speech perception for adults who use hearing aids in conjunction with cochlear implants in opposite ears', Journal of Speech, Language, and Hearing Research, 49:338-351.

[17] Lewis TW & DMW Powers 2005 'Distinctive Feature Fusion for Improved Audio-Visual Phoneme Recognition' Paper presented at 8th IEEE International Symposium on Signal Processing and Its Applications (ISSPA) 2005, Pp 62-65, Sydney, Australia, 28-31 August 2005. IEEE Press.

[18] Lewis TW & Powers DMW 2008 Distinctive Feature Fusion for Recognition of Australian English Consonants. Proceedings of Interspeech 2008, Brisbane. Pp 2671-2674

[19] Goecke R & Millar JB. 'Statistical Analysis of the Relationship between Audio and Video Speech Parameters for Australian English'. In J.L. Schwartz, F. Berthommier, M.A. Cathiard, and D. Sodoyer (eds.), *Proceedings of the ISCA Tutorial and Research Workshop on Auditory-Visual Speech Processing AVSP 2003*, pages 133-138, St. Jorioz, France, 4 – 7 September 2003.

[20] Li Y & DMW Powers 2001 'Speech Separation Based on Higher Order Statistics Using Recurrent Neural Networks', pp. 45-56, Proc. International Workshop on Hybrid Intelligent Systems (HIS'01), December 2001; Springer-Verlag ''Advances in Soft Computing' Series.

[21] Tran D, Wagner M, Lau YW & M Gen 2004 'Fuzzy Methods for Voice-Based Person Authentication' IEEJ (Institute of Electrical Engineers of Japan) Transactions on Electronics, Information and Systems, 124(10): 1958-1963.

[22] Tran D & M Wagner 2002 'A Fuzzy Approach to Speaker Verification' International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI), 16(7): 913-925
[23] Naseem I, Togneri R & M Bennamoun 2009 'Sparse Representation for Video-based Face Recognition'. Proceedings of ICB, June 2009, Alghero, Italy.
[24] Bavin EL, Grayden, DB, Scott K & T Stefanakis 2010. 'Testing auditory processing skills and their associations with language in 4-5 year-olds', Language & Speech, 53: 31-47.
[25] Talaricoa M, Abdillaa G, Aliferisa M, Balazica I, Giaprakisa I, Stefanakis T, Foenander K, Grayden DB & AG Paolini 2007 'Effect of age and cognition on childhood speech in noise perception abilities', Audiology & Neurotology, 12:13-19.
[26] McIntyre G & R Goecke 2007 'Towards affective sensing'. Proceedings of the 12th International Conference on Human-Computer Interaction HCII2007, 3, 411-420.
[27] Yacoub S, Simske S, Lin X & J Burns 2003 'Recognition of emotions in interactive voice response systems' in Proc. Eurospeech, pp. 729-732, September 2003
[28] Vidhyasaharan S, Ambikairajah E & Epps J 2009 'Speaker dependency of spectral features and speech production cues for automatic emotion classification', in Proc. IEEE Int. Conf. on Acoust., Speech and Sig. Proc. (Taipei, Taiwan).
[29] Kuratate K 2008 'Text-to-AV Synthesis System for Thinking Head Project' Proceedings of International Conference of Auditory-Visual Speech Processing 2008, Pp 191-194.
[30] Powers, DMW, Leibbrandt, RE, Pfitzner, D, Luerssen, MH, Lewis, TW, Abrahamyan, A & K Stevens 2008 'Language Teaching in a Mixed Reality Games Environment' The 1st International Conference on Pervasive Technologies Related to Assistive Environments (PETRA) DOI 10.1145/1389586.1389668
[31] Dale R & Viethen J 2009 'Referring Expression Generation through Attribute-Based Heuristics' Proceedings of the 12th European Workshop on Natural Language Generation, 30th-31st March 2009, Athens, Greece.
[32] Viethen J & Dale R 2006 'Algorithms for Generating Referring Expressions: Do They Do What People Do?' Proceedings of the 4th International Conference on Natural Language Generation, 15-16 July, Sydney, Australia. Pp. 63-70.
[33] Powers DM & Leibbrandt RE 2009 'Rough Diamonds in Natural Language Learning'. Proceedings of the 4th International Conference on Rough Sets and Knowledge Technology, Springer Lecture Notes in Computer Science. pp.17-26.
[34] Huang JH & Powers DM 2008 'Suffix-tree-based approach for Chinese information retrieval'. Proceedings of the International Conference on Intelligent Systems Design and Applications (ISDA 2008), Vol. 3, Pp 393-397.
[35] Pfitzner DM, Leibbrandt RE & Powers DM 2008a 'Characterization and Evaluation of Similarity Measures for Pairs of Clusterings', Knowledge and Information Systems: An International Journal, DOI 10.1007/s10115-008-0150-6.
[36] Pfitzner DM, Treharne T & Powers DM 2008b 'User Keyword Preference: the Nwords and Rwords Experiments' International Journal of Internet Protocol Technology 9:149-158 DOI 10.1504/IJIPT.2008.020947
[37] Butcher AR 1996 Levels of representation in the acquisition of phonology: evidence from 'before and after' speech. In B. Dodd, R. Campbell & L Worall (eds): Evaluating Theories of Language: evidence from disordered communication. London: Whurr Publishers, 55-73.
[38] Arciuli J & S McLeod 2008 'Production of /st/ clusters in trochaic and iambic contexts by typically developing children'. Proceedings of the 8th International Seminar on Speech Production (ISSP). Strasbourg, France. pp. 181-184.
[39] Rose P 2002 Forensic Speaker Identification .London: Taylor & Francis.
[40] Butcher AR 2002 'Forensic Phonetics: Issues in speaker identification evidence'. Proceedings of the Inaugural International Conference of the Institute of Forensic Studies: 'Forensic Evidence: Proof and Presentation', Prato, Italy 3-5 July [CD-ROM no page numbers].
[41] Loakes D & K McDougall 2010 'Individual variation in the frication of voiceless plosives in Australian English: a study of twins' speech. Australian Journal of Linguistics 30, 155-181.
[42] Foulkes P 2005 'Sociophonetics'. In Brown K (ed) Encyclopedia of Language and Linguistics (2nd ed). Amsterdam: Elsevier, pp.495-500.
[43] Labov W 1972 Sociolinguistic patterns, Philadelphia: Univ. of Philadelphia Press.
[44] Anderson ., Bader M, Bard E, Boyle E, Doherty GM, Garrod S, Isard S, Kowtko J, McAllister J, Miller J, Sotillo C, Thompson HS &Weinert R. 1991. The HCRC Map Task Corpus. Language and Speech 34, 51-366.
[45] Smits R, Warner N, McQueen, J & Cutler A 2003 'Unfolding of phonetic information over time: A database of Dutch diphone perception' Journal of the Acoustical Society of America 113: 563-574.