HOW TO HANDLE PRONUNCIATION VARIATION IN ASR: BY STORING EPISODES IN MEMORY?

H. Strik

Centre for Language and Speech Technology (CLST)
Radboud University Nijmegen, the Netherlands

ABSTRACT

Almost all current automatic speech recognition (ASR) systems use a similar paradigm [3, 51, 52], which will be referred to here briefly as the 'invariant approach'. Despite intensive research, ASR performance is still at least an order of magnitude lower than that of human speech recognition (HSR). The difficulties encountered in improving ASR performance, in combination with the awareness that current ASR systems have some shortcomings, have led many to believe that a new paradigm for ASR is needed. In this paper a novel paradigm for ASR is presented.

The invariant approach has also dominated (psycho-) linguistics. However, recent findings that indexical and detailed (sub-phonemic) information influence lexical access, have started a debate in (psycho-)linguistics on how these findings could be incorporated in HSR theories and models. On the basis of these findings episodic theories have been proposed. Although the episodic speech recognition (ESR) model is mainly inspired by HSR research, it is also very interesting and promising for ASR, since it has the potential to resolve some shortcomings of the mainstream ASR approach.

1. INTRODUCTION

One of the main issues in speech recognition is the large amount of variability present in speech [31, 50, 51, 52, 53, 54]. Yet, we perceive acoustically very different realizations of words as instantiations of the same 'entity'. Since a large amount of variability is also found in many other types of stimuli, e.g. visual and tactile stimuli, the problems encountered in speech recognition are also encountered in other areas of perception, cognition and learning. The discrepancy between the apparent physical variation on the one hand, and the perception of the same 'entity' on the other, has raised many questions about how variable stimuli can be mapped onto invariant 'concepts'. This invariance problem is related to questions about perceptual normalization and constancy, the representation of the concepts in memory, and the way these concepts are accessed upon the arrival of physical stimuli.

Although the invariance problem has been studied in many areas of cognitive science, such as linguistics, psycholinguistics, psychology, speech science, neuroscience, and engineering [see, e.g., 14, 24, 44, 49, 56], there is no consensus about which theory describes the observations best. The approaches used so far can be roughly divided into two classes. In the 'invariant approach', it is assumed that invariants are stored, and that the canonical representations are obtained through perceptual normalization. The invariant approach is also called the symbolic approach. A radically different approach is the episodic approach, which is also referred to as multiple trace, exemplar, example-based, instance-based, or template-based approach [13, 14, 15, 18, 24, 44, 52, 56]. In the episodic approach, it is assumed that a large number of episodes (traces) are stored, i.e. single instances of stimuli with many details, instead of the canonical representations stored in the invariant approach.

In order to get a better idea of the differences between the two approaches, some of these differences are listed in Table 1. However, it should be noted that the differences mentioned in Table 1 are extremes, and that many theories take an intermediate stance.

2. ASR

Almost all current ASR systems use a similar paradigm [3, 51, 52] that is characterized by a representation of speech as a sequence of phone(me)s. However, it has long been known that this invariant, 'beads-on-a-string' approach is problematic for ASR [8, 11, 12, 25, 41, 42]. As articulators cannot suddenly jump from one position to another, articulated speech sounds change gradually. Furthermore, it is well known that not all articulators change in synchrony at a certain point in time, i.e. there are no clear boundaries between phonemes [16, 17].

Table 1. The invariant approach versus the episodic approach.

invariant approach	*episodic approach*
complex mapping	simple mapping
parsimonious representation	extensive representation
canonical representations	detailed stimulus information
Quantal	gradual
encode generalities	encode particulars
Categorical	graded
abstract units, often described by symbols	episodes, exemplars
abstractionist view	holistic view
analytic approach	analogical approach
variation is noise, a nuisance	variation contains information useful for perception
strip off variation during normalization	use variation
dissociation of form and content	mutual dependencies
linguistic information (content)	also indexical information (form)

These limitations of the mainstream ASR approach are acknowledged, and there have been attempts to (partially) overcome them, e.g. by using segments larger than the phone(me) and by using models that can better incorporate the dynamic information present in the speech signals [e.g., 8, 11, 12, 25, 41, 42].

Despite intensive research, the performance of ASR is still at least an order of magnitude lower than that of human speech recognition [28, 29, 45]. A large part of the gain in performance has been obtained by using increasing amounts of speech. However, Moore recently concluded that approximating the performance levels of humans (with current ASR paradigms) would require up to 10,000,000 hours of speech, which – he estimated – is about equivalent to 100 human lifetimes of exposure to speech [38, 39]. The fact that almost all ASR systems are based on the same paradigm, and the difficulties encountered in improving the current ASR performance levels, has led many to believe that a new paradigm for ASR is needed [see e.g. 3, 39, 41, 45, 51, 52]. Nevertheless, there have been remarkably few attempts to develop radically new approaches.

3. HSR

In HSR research, different kinds of models of spoken word recognition are used. Although these models differ, the dominating models in HSR, such as Trace [34], Shortlist [40], PARSYN [30], and DCM [10]), are all representatives of the invariant approach. These models are challenged by recent findings on non-verbal information and phonetic variation. These findings are summarized here.

Speech contains two types of information: (1) verbal information and (2) non-verbal (indexical) information. Verbal information is mainly related to the content of the message, while indexical information is more related to the form, such as properties of the speaker (e.g. F0 and speech rate). With respect to indexical information, some interesting findings have been reported in the literature recently. Both indexical and non-indexical properties of speech appear to be stored by humans [6, 14, 44]. Familiarity with a person's voice facilitates recognition of that person's speech [13, 14, 44], and facilitation also occurs for speakers whose speech is similar [13, 14]. Also for visual perception it has been found that familiar patterns are perceived better than unfamiliar ones [20, 21]. Besides these findings on indexical information, experimental results also show that fine-grained acoustic details can influence lexical access [18, 19, 27, 35, 36, 37, 46, 56], such as, e.g., subphonemic differences between realizations of the monosyllabic word "ham" and the first syllable of "hamster" [46]. Because these findings on indexical and detailed information are difficult to explain in current models of spoken word recognition, also in the field of HSR there is a growing belief that new models are needed [see, e.g., 31].

4. INVARIANT VS. EPISODIC

The differences mentioned in Table 1 are very general. In this section a short description is given of some differences between invariant and episodic approaches which are more related to speech recognition. Since almost all current ASR systems use the HMM approach or related approaches, such as the hybrid HMM/ 'artificial neural network' (HMM/ANN) approach [4], we will focus here on comparing these HMM-based approaches to episodic ones.

In the HMM-based approach words are stored in the lexicon in the form of sequences of abstract (usually phonemic) symbols. Often for some words more than one entry per word are present in the lexicon in order to model pronunciation variation, and these pronunciation variants are usually stored as different transcriptions in terms of phonemes. In the episodic model words are not represented as sequences of symbols, but as sequences of (abstract) units that are represented in the form of many episodes (trajectories). In an ESR system the incoming signal is compared to sequences of stored episodes, e.g. sequences of feature vectors are compared. In an HMM-based system, the signal is compared to a sequence of states, and for each state the conditional probability of a frame given that state is calculated by means of the stored probability density functions (pdf's, usually Gaussian mixtures) or the stored artificial neural networks (ANNs) [4]. Probability density functions and artificial neural networks are parametric representations of all feature values observed in a large training corpus.

The difference between comparing two trajectories (two episodes) and comparing a trajectory with a sequence of states is well illustrated in the figures in [43]. Two of these figures are reproduced here (see Figure 3). In these figures trajectories are shown for various realizations of the diphthong /aj/ in 'nine' for two speakers (Jim and David). The HMM for this diphthong consists of three states: the first state 'aj(', the middle state 'ajl', and the final state 'aj)'. In these two figures, the ellipses represent the covariances, and 'X' denotes the beginning of the trajectories. These figures show that trajectories differ a lot, not all of them are described equally well by the sequences of states, and that a great deal of, especially dynamic, information is lost if trajectories are described as sequences of states. It can be observed in Figure 3 that a lot of the information present in the trajectories of David and Jim, and especially the differences between the trajectories for these two persons, is not represented well by the three states of the HMM. The episodes (trajectories) keep the temporal continuity that is difficult to model appropriately in the invariant approach. The continuity is preserved within the units, but it is also possible to optimize the continuity between units by using continuity constraints. Various measures can be employed for this purpose; obvious candidates are indexical properties such as pitch, speech rate, gender, etc. The continuity constraints can make it less likely that during recognition the system jumps from a certain trajectory (e.g. from a fast speaking female with high pitch) to another trajectory with different properties (e.g. from a slowly speaking male with low pitch). On the other hand, in HMM-based systems it is possible that the best path jumps from a Gaussian describing certain signals (e.g. fast, female, high pitch) to another Gaussian describing a very different kind of signals (e.g. slow, male, low pitch).

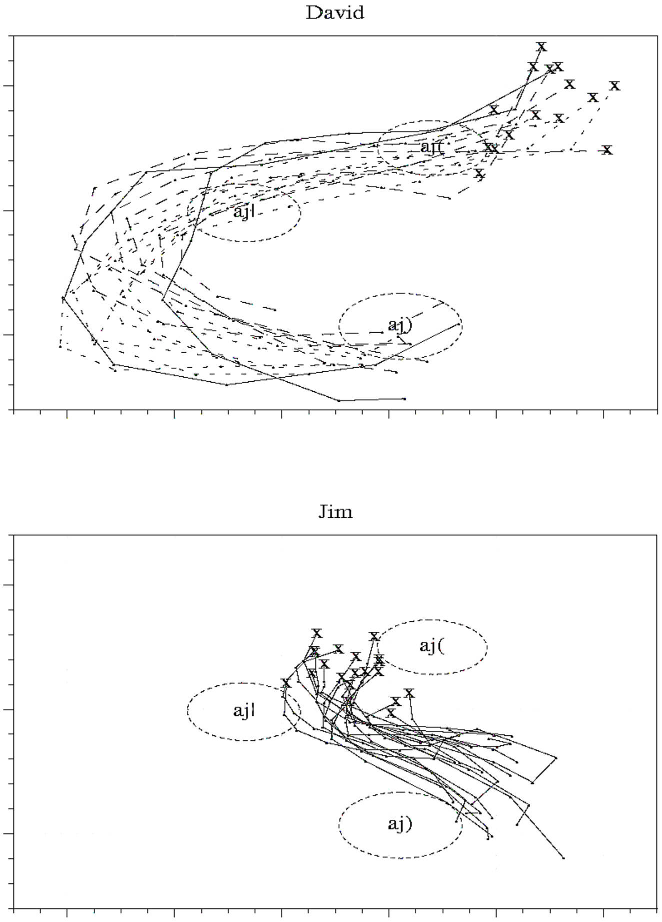

Figure 3. Trajectories of the diphthong /aj/ in 'nine' for two speakers. (Figures 3 and 4c of [43], reproduced with permission of the authors.)

Not only do the ESR and an HMM-based system differ with respect to recognition, but they also differ a lot regarding training. The difference can be shortly described as follows. The starting point in both cases will be a large amount of speech signals, i.e. many trajectories. In an ESR system individual trajectories are stored, while in the HMM-based system parametric descriptions (e.g., probability density functions) are stored.

5. DISCUSSION

When I discussed episodic models for speech recognition and gave presentations on this topic [see, e.g., 52], I received mixed reactions. Some people were enthusiastic, but the majority, especially people from the field of ASR, were skeptical. Up till now, almost all ASR research has focused on recognizing words, the verbal information, while indexical information has hardly been used. Until recently, it was maybe also practically impossible to investigate episodic approaches to speech recognition, simply because computers with sufficiently large memory and computing power were not available. Today, however, the computing power and memory that are needed to investigate the episodic approach to speech recognition are rapidly becoming available. Recently some research has started on using episodic approaches for ASR, and the first results are promising [1, 2, 9, 32, 33].

Using episodic models for ASR offers many interesting possibilities, especially regarding aspects that cannot be handled properly in HMM-based systems. However, there are also many questions to be answered about the application of episodic models in ASR. A few of these issues are mentioned here.

The first issue concerns segmentation and learning. Although in writing words are clearly demarcated by spaces and other punctuation marks, such clear word boundaries are certainly not present in speech. When humans try to decode speech, they thus have to segment the speech signal and recognize words. The correct segmentation can be obtained by finding the correct sequence of words, i.e. segmentation by competition between different sequences of words. However, there are many indications that additional cues are used to find the correct segmentation, such as stress, phonotactic constraints, and allophonic variation [7, 22, 26, 35]. Few or none of these cues are used in most current ASR systems.

It appears that some cues are available earlier than other ones, e.g. stress-based cues are available earlier than phonotactic and allophonic cues [22, 26]. An intriguing aspect is that these additional segmentation cues are related to regularities in the structure of the language, and since this structure is gradually acquired by the language learner, so are the cues that can be used for segmentation. However, in order to acquire structure segments have to be stored in memory. The question then is how this system can bootstrap itself.

This bootstrap issue has been the topic of many studies [see, e.g., the references in 5]. In almost all of these studies transcriptions of speech signals were used as input, usually a phoneme transcription [5]. However, making transcriptions of speech signals requires knowledge, e.g., in the case of phoneme transcriptions, about the phoneme inventory of a language, and how unknown acoustic signals can be transformed into sequences of phonemes. At the beginning (of the bootstrap procedure) the knowledge for making a phoneme transcription is not yet available. How then can the boundaries be identified in the acoustic signals?

In his overview paper about computational models of segmentation and word discovery, Brent distinguishes three strategies [5]. The word-recognition strategy, one of these three strategies, seems to be applicable to acoustic signals too (besides being applicable to transcriptions), and can thus also be applied in ASR systems. In short, it works as follows. Suppose someone says the Dutch utterances: "datiseen" … "datiseenpoes" … "eenpoesje" to an infant, and none of the units already stored by the infant matches part of these utterances. Then first "datiseen" is stored, the second utterance is segmented: "datiseen_poes", "poes" is stored, and finally "een" and "je" of "een_poes_je" will be stored. Thus, if an already stored unit matches part of an utterance, then the remaining parts of the utterance are stored as units. If not, then the whole utterance is stored. Isolated words are not essential for this bootstrapping procedure, as this example shows, but they are certainly helpful (on average, about 7% of the words addressed to infants are isolated words [5, 55]).

Another issue concerns the representations that are stored in memory: what should be stored, and how? What are the units? For instance, should episodes be stored for words, or should words be represented in terms of subword units, e.g. (demi-)syllables, phonemes, or (articulatory) features? Furthermore, what amount of detail should be stored in these representations?

It has been found that indexical properties can facilitate recognition, e.g. familiar voices are recognized faster and better (see section 3). In fact, these findings are one of the main reasons for the increasing interest in episodic models, because such models have the potential of simulating these findings. Which indexical properties should be stored in memory in order to make facilitation possible?

Another interesting question is whether the facilitation effect can be enhanced by using continuity constraints between episodes, i.e. penalizing a sequence of episodes for which the stored (indexical) properties differ substantially (see also section 4). Such continuity constraints have proven to be useful in concatenative speech and musical sound synthesis [47, 48]. These continuity constraints could also be employed to, e.g., simulate the finding that listeners generally expect the voice of the talker to remain constant [23].

And what to do with normalization in episodic models? Speech contains a lot of variation. The amount of variation can be reduced by using (perceptual) normalization. In the literature, various normalization techniques have been proposed for various purposes (e.g., for speaker and channel normalization), and previous research has shown that for invariant ASR models performance can be increased substantially by using appropriate normalization techniques [see e.g. 57]. However, the question is whether normalization should be applied in episodic models. And, if so, what kind of normalization should be applied. A complicating factor is that using normalization, e.g., speaker normalization, will affect the indexical properties, and thus probably also facilitation.

The issues mentioned above are only some of the issues that have to be addressed. Furthermore, to develop complete computational models it also is necessary to make decisions about all details of the models such as features, distance metrics, search algorithm, etc. This requires research to be able to make the optimal decisions. In turn, this requires a lot of effort, which is not unusual when starting a new line of research, like, e.g., acquiring funding which is generally not easy for basic, non-applied, risky research. However, entering new paths may lead to new insights, and, perhaps, in the end to better models. I am confident that this is likely to be more rewarding than if we all keep to the same beaten track.

6. REFERENCES

The references are listed in alphabetic order.

[1] G. Aradilla, J. Vepa, H. Bourlard (2005) Improving Speech Recognition Using a Data-Driven Approach. Proc. of Interspeech, Lisbon, September 2005, pp. 3333-3336.

[2] S.E. Axelrod, B. Maison (2004) Combination of Hidden Markov Models with Dynamic Time Warping for Speech Recognition. Proc. of ICASSP, Montreal, May 2004.

[3] H. Bourlard, H. Hermansky, N. Morgan (1996) Towards Increasing Speech Recognition Error Rates. Speech Communication, vol. 18, no. 3: 205-231.

[4] H. Bourlard, N. Morgan (1994) Connectionist Speech Recognition: A Hybrid Approach. Kluwer Academic Publishers.

[5] M.R. Brent (1999) Speech segmentation and word discovery: A computational perspective. Trends in Cognitive Science, 3: 294-301.

[6] C.G. Clopper, D.B. Pisoni (2002) Perception of Dialect Variation: Some Implications for Current Research and Theory in Speech Perception. In: Research on Spoken Language Processing Progress Report No. 25 (2001-2002), Indiana University: 269-290.

[7] A. Cutler, D. Norris (1988) The role of strong syllables in segmentation for lexical access. Journal of Experimental Psychology: Human Perception and Performance 1988, 14: 113-121.

[8] L. Deng, M. Aksmanovic, D. Sun, C.F.G. Wu (1994) Speech recognition using hidden Markov models with polynomial regression functions as nonstationary states. IEEE Transactions on Speech and Audio Processing, 2: 507-520.

[9] M. De Wachter, K. Demuynck, D. Van Compernolle, P. Wambacq (2003) Data Driven Example Based Continuous Speech Recognition. Proc. Interspeech, Geneva, September 2003, pp. 1133-1136.

[10] M.G. Gaskell, W.D. Marslen-Wilson (1997) Integrating form and meaning: A distributed model of speech perception. Language and Cognitive Processes, 12: 613-656.

[11] O. Ghitza, M.M. Sondhi (1993) Hidden Markov models with templates as nonstationary states: An application to speech recognition. Computer, Speech and Language, 7:101-119.

[12] J.R. Glass (2003) A probabilistic framework for segment-based speech recognition. In: M. Russell, J. Bilmes (Eds.) Special issue on "New Computational Paradigms for Acoustic Modeling in Speech Recognition", Computer, Speech & Language, 17(2-3): 137-152.

[13] S.D. Goldinger (1996) Words and voices: episodic traces in spoken word identification and recognition memory. J. of Experimental Psychology; Learning Memory and Cognition, 33: 1166-1183.

[14] S.D. Goldinger (1997) Words and voices: perception and production in an episodic lexicon. In: K. Johnson & J.W. Mullenix (Eds.), Talker Variability in Speech Processing. Academic Press: 33-66.

[15] S.D. Goldinger (1998) Echoes of echoes? An episodic theory of lexical access. Psychological Review, 105: 251-279.

[16] J. Goldsmith (1979) Autosegmental phonology. PhD thesis, Massachussets Institute of Technology, Cambridge (New York: Garland Press, 1979).

[17] J.A. Goldsmith (1990) Autosegmental and Metrical Phonology. Oxford: Blackwell.

[18] S. Hawkins (2003) Contribution of fine phonetic detail to speech understanding. Proc. of the 15th Int. Congress of Phonetic Sciences (ICPhS-03), Barcelona, Spain: 293-296.

[19] S. Hawkins, R. Smith (2001) Polysp: A polysystemic, phonetically-rich approach to speech understanding. Italian Journal of Linguistics—Rivista di Linguistica, vol. 13: 99-188.

[20] D.L. Hintzman, R. Block, N. Inskeep (1972) Memory for mode of input. J. of Verbal Learning and Verbal Behavior, 11: 741-749.

[21] L. Jacoby, C. Hayman (1987) Specific visual transfer in word identification. J. of Experimental Psychology: Learning, Memory,and Cognition, 13: 456-463.

[22] E.K. Johnson, P.W. Jusczyk (2001) Word segmentation by 8-month-olds: When speech cues count more than statistics, Journal of Memory and Language, 44: 1-20.

[23] K. Johnson (1991) Differential effects of speaker and vowel variability on fricative perception. Language and Speech, 34: 265-279.

[24] K. Johnson, J. Mullennix (Eds.) (1997) Talker Variability in Speech Processing. San Diego: Academic Press.

[25] B. H. Juang, L. Rabiner (1986) Mixture Autoregressive Hidden Markov Models for Speaker Independent Isolated Word Recognition. Proc. of ICASSP-86, Tokyo, Japan: 41-44.

[26] P.W. Jusczyk (1999) How infants begin to extract words from speech. Trends in Cognitive Science, 3: 323-328.

[27] K.J. Kohler (2003) Modelling stylistic variation of speech - Basic research and speech technology application. Proc. of the 15th

Int. Congress of Phonetic Sciences (ICPhS-03), Barcelona, Spain: 223-226.

[28] D.A. van Leeuwen, L.G. van den Berg, H.J.M. Steeneken (1995) Human benchmarks for speaker independent large vocabulary recognition performance. Proc. of Eurospeech-95, Madrid, Spain: 1461-1464.

[29] R.P. Lippmann (1997) Speech recognition by machines and humans. Speech Communication 33: 1-15.

[30] P.A. Luce, S.D. Goldinger, E.T. Auer, M.S. Vitevitch (2000) Phonetic priming, neighborhood activation, and PARSYN. Perception & Pschychophysics, 62: 615-625.

[31] P.A. Luce, C.T. McLennan (2003) Spoken Word Recognition: The Challenge of Variation. In: C. T. McLennan, P. A. Luce, G. Mauner, & J. Charles-Luce (Eds.), University at Buffalo Working Papers on Language and Perception, 2: 203-240.

[32] V. Maier, R.K. Moore (2005) An investigation into a simulation of episodic memory for automatic speech recognition. Proc. of Interspeech-2005, Lisbon, 5-9 September 2005, pp. 1245-1248.

[33] M. Matton, M. De Wachter, D. Van Compernolle, and R. Cools (2005) Maximum Mutual Information Training of Distance Measures for Template Based Speech Recognition. Proc. of SPECOM 2005, October 2005, Patras.

[34] J.L McClelland, J.L. Elman (1986) The TRACE model of speech perception. Cognitive Psychology, 18: 1-86.

[35] J.M. McQueen, A. Cutler (2001) Spoken word access processes: An introduction. Language and Cognitive Processes, 16(5/6): 469-490.

[36] J.M. McQueen, A. Cutler, D. Norris (2003) Flow of information in the spoken word recognition system. Speech Communication 41: 257-270.

[37] J.M. McQueen, D. Dahan, A. Cutler (2003) Continuity and gradednes in speech processing. In: A.S. Meyer & N.O. Schiller (Eds.) Phonetics and phonology in language comprehension and production: Differences and similarities, Berlin: mouton.

[38] R.K. Moore (2001) There's no data like more data: but when will enough be enough? Proc. of the Workshop on Innovations in Speech Processing, UK, Insitute of Acoustics, 23 (3): 19-26.

[39] R.K. Moore, A. Cutler (2001) Constraints on theories of human vs. machine recognition of speech. In: R. Smits, J. Kingston, T.M. Nearey, & R. Zondervan (Eds.), Proc. of SPRAAC (Workshop on Speech Recognition as Pattern Classification), MPI, Nijmegen: 145-150.

[40] D. Norris (1994) Shortlist: a connectionist model of continuous speech recognition Cognition 52: 189-234.

[41] M. Ostendorf (1999) Moving beyond the 'beads-on-a-string' model of speech. Proc. of the IEEE Automatic Speech Recognition and Understanding (ASRU) Workshop, Keystone, Colorado, USA.

[42] M. Ostendorf, V. Digalakis, O. Kimball (1996) From HMMs to Segment Models: A Unified View of Stochastic Modeling for Speech Recognition. IEEE Transactions on Speech and Audio Processing, 4: 360-378.

[43] S.D. Peters, P. Stubley (1998) Visualizing speech trajectories. In: [54]: 97-102.

[44] D.B. Pisoni (1997) Some thoughts on "Normalization" in speech perception. In: K. Johnson & J.W. Mullennix (Eds.), Talker Variability in Speech Processing. Academic Press: 9-32.

[45] L. C. W. Pols (1999) Flexible, robust, and efficient human speech processing versus present-day speech technology. Proc. of the 14th Int. Congress of Phonetic Sciences (ICPhS-99), San Francisco, USA: 9-16.

[46] A.P. Salverda, D. Dahan, J.M. Mcqueen (2003) The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. Cognition, 90: 51-89.

[47] J.P.H. van Santen, R.W. Sproat, J.P. Olive, J. Hirschberg (Eds.) (1996) Progress in Speech Synthesis. Berlin, Germany: Springer-Verlag, 1996.

[48] D. Schwarz (2000) A System for Data-Driven Concatenative Sound Synthesis. Proc. of the COST G-6 Conference on Digital Audio Effects (DAFx), Verona, Italy: 97-102.

[49] K.N. Stevens (2002) Toward a model for lexical access based on acoustic landmarks and distinctive features. Journal of the Acoustical Society of America, vol. 111: 1872-1891.

[50] H. Strik (Ed.) (1999) Special issue of Speech Communication about 'Modeling pronunciation variation for automatic speech recognition'. Speech Communication 29, 166 pages.

[51] H. Strik (2001) Pronunciation adaptation at the lexical level. In: J-C. Juncqua, C. Wellekens (Eds.) Proc. of the ITRW 'Adaptation Methods For Speech Recognition', Sophia-Antipolis, France: 123-131.

[52] H. Strik (2003) Speech is like a box of chocolates. Proc. of the 15th Int. Congress of Phonetic Sciences (ICPhS-03), Barcelona, Spain: 227-230.

[53] H. Strik, C. Cucchiarini (1999) Modeling pronunciation variation for ASR: a survey of the literature. Speech Communication 29 (2-4): 335-246.

[54] H. Strik, J.M. Kessens, M. Wester (Eds.) (1998) Proc. of the ESCA 'Rolduc' Workshop 'Modeling Pronunciation Variation for Automatic Speech Recognition', Rolduc.

[55] J. van de Weijer (1998) Language input for word discovery. PhD thesis, University of Nijmegen.

[56] J.F. Werker (2003) The Acquisition of Language Specific Phonetic Categories in Infancy. Proc. of the 15th Int. Congress of Phonetic Sciences (ICPhS-03), Barcelona, Spain: 21-26.

[57] F. de Wet (2003) Automatic Speech Recognition in Adverse Acoustic Conditions. PhD thesis, University of Nijmegen.