CAN THE DEFINITION OF EACH SPEAKER BE EXPECTED TO COME FROM THE LABORATORY IN THE NEXT DECADES?

Francis Nolan

Department of Linguistics, University of Cambridge, UK

ABSTRACT

The symposium of which this paper is a part address questions that arise from speaker identification in forensics. The focus is on what can and cannot be expected of forensic speaker identification, and on directions for future research. The first three sections of this paper set the background for the symposium, and the subsequent sections suggest possible innovations. The overall theme is the need to explore new methods of imposing structure on the data used in speaker identification in order to understand variation. These methods include alternative phonological models, and a more explicit role for articulatory modelling.

1 INTRODUCTION

The primary concerns of phonetics have been to do with the realisation of language in the sound medium, but the scope of phonetics is much wider. A broad view of phonetics might see it as the discipline which answers the questions 'what can we tell when a person speaks, and how?' As soon as someone speaks, listeners are able to infer a wide variety of information other than that contained in the linguistically encoded 'message'. Much of that information is about the producer of the message. Listeners can infer (with a fair degree of reliability) the sex of the speaker, they can induce information about his or her health, and they can often identify the speaker as a person previously heard.

This last ability, the inference of identity, must lead us to assume that information about individual identity is convolved with the other information in the speech signal. This conclusion emerges too from other areas of the phonetic sciences: the difficulties of creating a reliable speech recognition system which is speaker-independent demonstrate that significant speaker-specific information is blended into the acoustic speech signal.

For automatic speech recognition, this speaker-specific information is unwanted noise, to be neutralised if at all possible. But in another domain, that of *speaker* recognition, it is the raw material, the structured variability and underlying regularity of which need to be determined, just as phonetics has done for the linguistically determined aspects of the speech signal. Applications of knowledge about speaker-characterising features of speech include Automatic Speaker Verification, which a massive market awaits in fields such as telephone banking, and, more controversially, forensic speaker identification. The latter provides the focus for this session.

2 DEFINING EACH SPEAKER

In 1934 Twaddell [1] cited Bloomfield as saying 'The physical (acoustic) definition of each phoneme of any given dialect can be expected to come from the laboratory within the next decades'. With the hindsight of six decades of acoustic speech analysis such a statement, if intended literally, might be seen as betraying a certain naivety, not least about the relation between phonological categories and the physical signal.

On the other hand there is a real, practical sense in which Bloomfield's prophecy has been fulfilled. Not only do we have, thanks both to extensive acoustic analysis and to advances in the acoustic theory of speech production, a very good understanding of the acoustic properties which realise phonemes of different types, but we also have advanced statistical models (such as HMMs) which, in some cases speaker-independently, can learn to recognise the realisations of each phoneme in the speech signal.

Whatever the correct assessment of Bloomfield's statement, it may be the case that we are in a similar position *vis-à-vis* individual speaker quality today as Bloomfield was in the 1930s in relation to phonemic quality. We have an analytic construct, speaker quality, for which (if we adopt an appropriately 1930s terminology) we have behavioural evidence, in the ability of listeners to identify speakers. We even have a fairly well worked out phonetic model, parallel to that provided by traditional phonetic analysis for the phoneme, of at least part of speaker quality: Laver's (1980) framework for the analysis of voice quality [2], for instance, can be seen as a model of that part of speaker quality which is under the speaker's control. But we do not have a comprehensive answer to the question: 'What defines an individual in the acoustic signal?'

As a starting point for this session on forensic speaker identification then we can therefore re-phrase Bloomfield's dictum, and debate the proposition that 'The definition of each speaker can be expected to come from the laboratory in the next decades'.

3 TWO STRANDS

There are perhaps two strands to consider in this proposition. The first is the nature of 'speaker quality'. What dimensions are involved? How much variation does an individual exhibit? And, most crucially, does each individual human being occupy a unique location in acoustic space? Or is there instead a significant degree of 'overlapping', by which an individual shares part, or indeed all, of his or her location with others, rather as the English phonemes /ɛ/ and /æ/ may share the phonetic realisation [æ] in words such as *well* and *gag* respectively, as a result of contextually induced allophonic variation? The answers which emerge to questions such as these about speaker quality will inform the issue of what we might mean by 'the definition of a speaker'.

The second strand to the proposition is the implication that it is specifically in the laboratory that progress will be made towards finding the definition of a speaker. Of course, if consideration of the first strand results in the conclusion that we have no viable theory of speaker quality, and if we take a somewhat purist view of empirical science to the effect that measurements and experiments cannot usefully be carried out in the absence of testable hypotheses generated by a theory, then there is no point in going into the laboratory. But it seems unlikely that both these negative conditions would hold. We probably do have the beginnings of a theory of speaker quality; and even if not, it may be that what we most need in this field are large-scale, pre-theoretical, 'taxonomic' studies of between- and within-speaker variation. If we accept that work in the laboratory is appropriate, we can then indulge in informed speculation about the kind of analyses and methodological developments which are likely to bring greater understanding of speaker quality.

Although Bloomfield's proposition is here newly adapted to speaker recognition, the debate which its adaptation encapsulates is already underway. Baldwin and French [3] address essentially the same proposition. Interestingly, the two authors arrive at diametrically opposed views. In Chapter 3, French writes 'For various theoretical reasons, I cannot forsee a day when phoneticians will be able to identify a speaker with the degree of certainty associated with the matching of fingerprints or DNA profiles' (p.62). Baldwin, in the final chapter, despite having taken throughout the book a generally negative stance towards the present-day contribution of acoustic phonetics to forensic speaker identification, writes more optimistically of the future: '... I positively believe there will one day be a "voiceprint", i.e. a print-out from some sort of, not necessarily electronic, device which will be able uniquely to identify an individual speaker' (p.126).

The fact that the authors hold disparate views on such a fundamental matter is, as the foreword to the book (p.iv) points out, potentially productive if the disagreements are rationalised. The shortcoming is perhaps that so little is said about the grounds for the disagreement that it is not clear what the framework for any discussion might be. It is hoped that this session will help to set out the parameters of such a discussion.

4 FUTURE PROGRESS

The other speakers in this symposium provide clear summaries of problems and

methods in forensic speaker identification, and point to ways of improving the methods. Braun focuses on the phonetician's role, while Broeders discusses how much solutions from Automatic Speaker Verification might contribute to the forensic task. Hollien presents a framework of requirements for objective speaker identification, in which nevertheless the ultimate decision is a human one.

Perhaps, though, because all three speakers are closely involved in the day-to-day work of speaker identification, they have chosen to concentrate on improvements and extensions to current approaches and conceptualisations. What I will try to do in the following sections is to suggest more radical departures from current thinking in the area.

The theme which links this speculative train of thought is the need to find new ways of coping with variability, a problem which Broeders draws attention to in his Section 2. No two utterances are identical, even if they are by the same speaker, and so speaker identification cannot proceed on the basis of rejecting a 'match' every time a difference is detected. Where there is a difference we need to understand what lies behind it. To achieve this understanding it will be argued that we need the most comprehensive account available of how phonetic material is structured by phonology, and we need to take advantage of those constraints on variability which are imposed by an individual's speech production mechanism.

5 PHONOLOGICAL THEORY

Like many areas of applied phonetics, speaker identification might be regarded by many contemporary phonologists as stuck in a time-warp. There is little evidence of any view other than one which assumes that phonological analysis is done in terms of phonemes, which receive a variety of realisations according to segmental context. Dialect spotting, measurement of acoustic values, and even more specialised concerns such as coarticulation, are discussed within this framework. Prosody is generally treated as an optional accessory, to be ignored most of the time, but, if dealt with, then treated not as an aspect of the phonological system but purely as an unstructured physical aspect of the signal, in terms of parameters such as mean fundamental frequency or overall perceived pitch. Conspicuously absent from work in speaker identification are concepts and representations taken from schools of phonology such as Autosegmental, Metrical, Dependency, and Government.

Does this matter? After all, the phonetic stuff is there, and the task is to distil out the speaker-specific essence from the signal; and it is hardly going to be important what phonological model one adheres to. But in fact it may matter, because one's phonological prejudices may influence where and how one looks for the speaker-specific essence.

For instance, if one's phonological model incorporates a prosodic hierarchy, with syllables and feet at the bottom, and intonational phrases at the top, it may lead one to be more choosy as regards which events one treats as phonetically equivalent than if one sees speech as a linear string of (phoneme-sized) beads. English /t/ is simply /t/, but an awareness of prosodic structure might restrain one from treating all the vowels in *debility* /dɪbɪlɪtɪ/ as equivalent. Again, speaker identification must have at its disposal accurate descriptions of dialect or accent differences within a language. Some of these are extremely complex, such as English plosive allophony (glottalisation, 'flapping', etc.; see e.g. [4] and [5] for Metrical and Government accounts), and adequate descriptions may only be possible in models embodying a rich phonological mechanism, including syllables, feet, prominence relations, and so on.

Similarly, without a well worked out model of intonational phonology, potentially speaker-specific phenomena may escape investigation. For instance, a search for differences between speakers in the realisation of prosodic categories only arises if one incorporates some prosodic phenomena into the phonological description. If one's phonological model incorporates an autosegmental-metrical representation of intonation in terms of high (H) and low (L) tones, for instance, as in much recent intonational work, it is more likely that the question will arise as to whether speakers may differ in their preferred alignment of the tones to the segmental material. Or an intonational model which includes the notion of downstep will allow of the question whether some speakers use downstep more than others, and of those who do use it, whether there are differences in the implementation of it.

The general point is that the evolution of phonological theory is driven, at least in part, by imperfection in the fit of previous models to the facts concerning the sound structure of language. In order to understand variation in the speech signal, the forensic phonetician needs the best available model. Good forensic phonetic practice is currently immeasurably better than that of the sound engineers mentioned in Braun's contribution to this symposium who compare waveshapes with no regard to the identity of the vowels those portions of signal are realising. But the possibility of further progress through the adoption of more sophisticated phonological models needs to be explored. In a sense, then, some of the means for progress towards the definition of the speaker lie outside the laboratory.

6 ACCENT ANALYSIS

Much of the contribution of the forensic phonetician today is of a kind which pre-dates instrumental analysis of speech samples. It is, in effect, practical dialectology; and when the question is whether two samples of speech were produced by the same human being, a sensible first step is to see whether they manifest the same linguistic properties by comparing their pronunciation. If the pronunciations are grossly different, the samples are unlikely to get as far as the forensic phonetician – those responsible for the legal side of the case will use their own judgment and conclude that a speaker with a London accent in one sample is unlikely to be the same individual as the Scottish speaker in another. The role of the phonetician will normally be to adjudicate in cases where the samples are already superfically similar. The specialist skills which a traditional phonetic training provides will allow the phonetician to notice, and classify, differences between samples which are more subtle than would be noticed by most untrained listeners. Although there are many problems to do with the linguistic variability which the speech of one person undergoes as a result of factors such as style, speaking context, and accommodation to interlocutors, close phonetic analysis can often reveal patterns of difference between samples which make it unlikely that they come from the same source.

How far one can go in the opposite direction and treat the absence of differences in pronunciation as evidence pointing towards the samples coming from the same speaker is a contentious matter. It rests, ultimately, on the question of how finely the 'isoglosses' of a dialect map can be drawn. The strongest position (see e.g. [6]) is that each individual speaks an 'idiolect'. However, even if it could be demonstrated no two individuals share a complete set of linguistic phonetic properties, it is doubtful whether the finite (and often short) samples available in forensic cases would allow a safe extrapolation from 'sameness of sample' to 'sameness of speaker'. As argued in [7] linguistic phonetic sameness may licence conclusions of 'possibly' the same, but not 'probably'.

Nonetheless linguistic phonetic analysis is an important element in the forensic phonetic approach, and must surely enter into 'the definition of the speaker'. Since such analysis requires a well-trained phonetic ear, surely it is unrealistic to expect progress on this aspect of the speaker's definition to come from the laboratory?

It is in fact far from unrealistic. For one thing, in the everyday practice of forensic phoneticians, acoustic analysis already supplements auditory analysis of what are, in effect, dialect features. But looking at the issue more fundamentally, auditory phonetic descriptions are necessarily abstractions, and rely on categories identified by selected perceptually salient characteristics of a sound. It has been shown in other phonetic areas that impressionistic descriptions may be only partially accurate. Production studies, for instance, have shown (cf. [8]) that segmental 'assimilation' can be an articulatorily gradient phenomenon, contrary to the implication of many segmental descriptions. Similarly, doubt has been cast [9] on speech error work based on

impressionistic observation. EMG monitoring of muscle activity reveals that far from involving only discrete segmental effects, speech errors range along a continuum of muscle activation.

The analogies of these findings as far as accent and dialect are concerned must as yet be a matter of speculation; but perhaps they are to be found in the effects which lie outside the static 'frozen frame' ([10], p.108) on which segmental phonetic description is based, and which implicitly or explicitly focuses on a characteristic 'target' for a segment. So whilst we might traditionally describe two accents as having a 'dark' syllable-initial realisation of /l/, and a close-mid realisation of /e/, the description of the accents might be refined by the discovery of systematically different coarticulatory treatments when the segments are juxtaposed. Fine details of intra- and inter-syllabic timing, elusive to auditory analysis, might be highlighted. And differences in intonational features, such as the alignment of pitch peaks relative to segmental material mentioned in the previous section, might be revealed as accent-specific. There is no evidence that the description of an accent is exhausted by what the ear can hear in the context of a classical phonemic framework.

Such instrumentally-mediated detail would, in a sense, provide a finer 'mesh' for the dialect grid which traditional phonetics imposes on the speech community. Given that the forensic phonetician is likely to be sent samples which are at least fairly similar in accent, the finer that mesh is, the more he or she can add to what the lay person can hear. Different speakers might be discriminable by 'sub-auditory' secondary dialect differences too subtle to be consciously manipulable; and if even the fine mesh fails to separate the samples, the odds against them being from the same individual are shortened (though one must still guard against the temptation to claim they are the same speaker merely because, in some more precise way than before, the samples share the same accent).

Another issue is whether the task of the phonetician in 'dialect spotting' will, in the future, be automated. As far as I am aware, little work has been done towards this goal. In the context of a multi-dialect automatic speech recognition system, however, [11] reports a technique which automatically assigns a speaker to one of four major English regional accent groups on the basis of several pre-determined utterances. These are chosen to contain diagnostics for the different accents. In a sentence containing the words *father*, *path*, and *car*, a similar vowel quality for all three ([ɑ:]) suggests Southern British, different for all three ([ɑ], [æ], [ɑᴵ]) General American, and so on. The crucial events are identified in the input utterance by time-warping it to a segmented reference utterance, but all spectral comparisons are internal to the input signal, so that no normalisation for individual speaker characteristics is needed prior to the accent decision.

This method requires the production of agreed speech material, and so even if its accuracy and discriminatory ability were vastly increased it would not threaten the role of the phonetician, whose knowledge and skill often permit an assessment of dialect similarity or difference on the basis of short samples of differing content. But in future decades a semi-automated and vastly improved version might have a role to play where amounts of material are large. Orthographic transcripts of long recordings could be searched automatically for words with dialect-sensitive vowels. These words could be located automatically in the acoustic signal by ASR techniques, with manual correction if necessary. Acoustic parameters would be extracted, and used firstly for 'sample-internal' dialect spotting, as described above; and secondly for direct comparison with values from another sample.

To suggest a procedure of this kind is not to ignore the difficulties – the effects of prosody, segmental context, and so on – but given the extremely powerful signal processing techniques available even today it is not too early to speculate as to how they might be applied in a phonetically informed way to the problem of speaker identity.

In this section, then, I have suggested that the laboratory should provide new approaches to the definition of accent characteristics, and to the detection of accent, which have up to now been a field predominantly for auditory phonetics.

7 VOCAL TRACTS SHAPE

Current research on defining the speaker involves measuring values such as formant frequencies associated with particular phonological events, and deriving estimates of between- and within-speaker variation. This is a vital kind of data collection, and needs to be pursued on as large a scale as possible. But the work tends to treat the measured values as independent, and as varying in a purely statistical fashion, rather than as varying in a lawful way governed by the nature of their source. Only by referring back to the source can the significance of variation begin to be assessed.

To put it another way, an individual's vocal tract shapes, and imposes strict (though by no means absolute) constraints on, the sound he or she can produce; and by considering measured acoustic values not in isolation but in relation to their source we may gain a more powerful grasp on variability.

It is already possible to estimate vocal tract lengths from formant frequencies, and, using for instance linear prediction, to estimate cross-sectional area functions for particular vowels. We can also use vocal tract synthesis models to compute formant frequencies for different tube shapes, and we can restrict the range of tube shapes broadly to those which are anatomically plausible. Source inference, and articulatorily realistic vocal tract synthesis, may prove powerful tools in the interpretation of variation.

For instance, two tokens of a vowel taken from different recordings turn out to have similar first and second formant frequencies, but a less similar third formant frequency. What is the threshold we use to decide 'different speaker'? Though clearly one would never answer the question 'are the recordings from the same speaker' on the basis of one vowel, the 'threshold' problem arises however many factors are taken into account. At present, the threshold would have to be a purely statistical one: from databases, we might estimate that a speaker's F3 frequency will vary by a given percentage for a particular vowel. But if the first recording contains enough material for a reasonably accurate vocal tract model of the speaker to be derived, it may be possible to say something like 'it is highly unlikely that the source of the first recording could achieve the specific combination of formant frequencies found in the vowel from the second recording.'

In this way acknowledging the mechanism producing the speech would allow us in our decision making to go beyond purely statistical treatments of the variability of acoustic data .

Such progress, if it is made, will come not simply from the laboratory, in the sense of empirical discoveries, but from the application of the acoustic theory of speech production.

8 ARTICULATION MODELS

A greater general awareness of the source of the speech signal may permit other novel insights. Speech does not originate from a tube producing a static set of formant frequencies, but from a dynamic complex of articulators working in close coordination to achieve the phonological requirements of an utterance. Generally we do not assume that every phonetic dimension is crucial at every instant in an utterance. Rather, we hypothesise that some 'target' events are more crucial than others. If this is the case, speakers may evolve individual articulatory strategies for achieving and moving between such targets. Such a view is implicit in studies of coarticulatory idiosyncrasy [12], [13].

But the relation between phonological requirements and articulation is not theory-neutral; nor are potential sources of between- and within-speaker variation totally independent of theoretical assumptions; and so it would be negligent for researchers in speaker identification to ignore theoretical and practical developments in articulation modelling.

Perhaps the most radical current view of the relation between phonological specifications and speech is Articulatory Phonology [14,15], whose phonological primitives are 'gestures' such as 'labial closure'. The notion of a gestures is taken from work on the control of skilled actions in a framework called 'Task Dynamics' [16]. Gestures, unlike features or segments, inherently possess dynamic characteristics, and they permit the computational modelling of articulator movements. A 'gestural score' specifies the relations between gestures needed for

particular utterances, and is seen both as a lexical-phonological representation, and a representation of the implementation of the utterance. It is possible to synthesise the speech signal from the gestural score via the task dynamic modelling of articulatory interaction and a vocal tract synthesiser, making it possible directly to predict the acoustic effect of constellations of gestures.

In Task Dynamics coordination of gestures is not represented straightforwardly in the time domain, but in the 'phase-plane', which depends on viewing articulatory movements as oscillations (damped or undamped). It has been claimed ([16], p.41 ff) that representation in the phase-plane may reveal consistency of gestural organisation across differences of rate and stress which is obscured by representations in the time domain. The phase-plane might, in effect, reduce apparent within-speaker variation in the timing of events. If the phasing turned out to differ across speakers (as some studies have implied, e.g. [17]), a better separation of speakers might be achieved than is possible in purely acoustic data.

Importantly many kinds of phonetic variation associated with changes in rate and style, and which are often modelled as the output of phonological rules, are said to emerge automatically from the gestural account as a result of general processes of increased overlap between gestures (presumably some inter-gestural phasing relations do change) and reduction in magnitude of gestures [14]. In the sentence 'He said *a fan could surprise you*', which might be realised as [...ə fæn kəd səpraɪz jʊ] or, more rapidly, as [...ə fæŋ kəd spraɪʒʊ], the apparent change of the alveolar nasal to a velar would result from the velar gesture for the plosive overlapping the alveolar gesture and masking its acoustic consequences (cf. [18]); the 'deletion' of the first syllable of 'surprise' would be an automatic consequence of the labial closure overlapping the [s]; and the [ʒ] would result from the competing effects of two gestures (for [z] and [j]) overlapping and competing for the same articulator.

Faced with two very different acoustic stimuli, let us say a fast and a slow utterance of '*fan could*', or '*surprise you*', from different recordings, it might be possible to determine whether the different acoustic properties of the second one (changed formant trajectories, durations, etc.) are compatible with it being a speeded up version of the first one, or are the product of a different articulatory mechanism. That is, if we have enough speech at one rate in the first recording to be able to replicate the speaker using articulatory synthesis, and if rate change turns out to involve similar articulatory strategies across speakers, we could change the rate of articulation of the synthesised version of the first and test whether the acoustic properties of the second recording are compatible with it having been produced by the same speaker.

The implementation, let alone practical application, of such a procedure, lies a long way off. In particular the inference of articulatory activity from the acoustic signal, which is a prerequisite to the suggestions above, would require a very sophisticated method probably involving analysis-by-synthesis. But such an approach is not unimaginable, as it would have been until relatively recently, and it is the kind of ambitious goal which might stimulate fundamental laboratory research towards taming the variability problem in speaker identification.

9 TWINS SPEAK

The linchpin of any investigation is control. If we are to understand the ways in which speakers differ, and we assume that the differences can broadly be categorised as dependent on 'organic' and 'learned' factors, it would be useful to be able to control one or other of these factors. Nature provides such a control in the case of identical twins, for whom it is a reasonable hypothesis, though not a certainty, that they will have extremely similar vocal mechanisms. This natural control case must surely figure prominently in future research into the definition of a speaker.

Recently a pilot study of three pairs of university-age identical twins, brought up in shared environments, was carried out in Cambridge by Tomasina Oh. The twins recorded lists of words with /l/ and /r/ before a range of vowels, as in [14], one of the objects being to discover if members of a pair had different coarticulatory strategies. Interestingly, consistent differences of various kinds did emerge between the members of each pair. In the most striking case, /r/ was realised by different articulations: as [ɹ] by one member and as [ʊ] by the other, with consequent acoustic and coarticulatory differences. In another pair, one member consistently palatalised /l/ more than the other. In the third pair, whose words showed in general a high degree of similarity, one twin showed greater fronting of /u:/ after /l/ than the other. On the other hand the prediction that there might be distinct coarticulatory strategies was not borne out in general.

This particular study demonstrates that identical twins do not have identical speech. But, more generally, studies of twins provide the possibility of studying the extent to which speaker characteristics are behavioural rather than anatomical.

10 CONCLUSION

It is certain that progress towards the definition of the speaker will involve the laboratory. What I have argued here is that to tackle the central problem of between- and within-speaker variability, it will not be sufficient (though it will be necessary) to carry out acoustic measurement studies on large populations, and to continue only to apply current techniques of analysis. Rather, theoretical and technical innovations of various kinds are needed; and our goals need to be, perhaps, more ambitious than at present.

In what sense 'the definition of each speaker can be expected to emerge from the laboratory' will have to wait for an answer until the results of such innovations begin to materialise. Whilst I share French's more cautious view that speaker identification will never be like fingerprinting (section 3 above), I believe we are far from having reached the bounds of what is possible in speaker characterisation.

REFERENCES

[1] Twaddell, W.F. (1935) On defining the phoneme. *Language Monograph* 16.
[2] Laver, J. (1980) *The Phonetic Description of Voice Quality*. Cambridge: CUP.
[3] Baldwin, J., and French, P. (1990) *Forensic Phonetics*. London: Pinter.
[4] Gussenhoven, C. (1986) English plosive allophones and ambisyllabicity. *Gramma* 10, 119-41.
[5] Harris, J. & Kaye, J. (1990) A tale of two cities: London glottaling and New York tapping. *Linguistic Review* 7, 251-74.
[6] Baldwin, J. (1979) Phonetics and speaker identification. *Medicine, Science and the Law* 19, 231-32.
[7] Nolan, F. (1991) Forensic phonetics. *Journal of Linguistics* 27, 483-493.
[8] Nolan, F. (1992) The descriptive role of segments: evidence from assimilation. In: G. Docherty and D.R. Ladd (eds.), *Laboratory Phonology 2*, 261-280. Cambridge: CUP.
[9] Mowrey, R.A. & MacKay, I.R.A. (1990) Phonological primitives: Electromyographic speech error evidence. *J. of the Acoustical Society of America* 88 (3), 1299-1312.
[10] Laver, J. (1994) *Principles of Phonetics*. Cambridge: CUP.
[11] W.J. Barry, C.E. Hoequist, and F. Nolan (1989) An approach to the problem of regional accent in automatic speech recognition. *Computer Speech and Language* 3, 355-66.
[12] Su, L-S., Li, K-P., and Fu, K.S. (1974) Identification of speakers by use of nasal coarticulation. *Journal of the Acoustical Society of America* 56, 1876–82.
[13] Nolan, F. (1983) *The Phonetic Bases of Speaker Recognition*. Cambridge: CUP.
[14] Browman, C.P. & Goldstein, L. (1989) Articulatory gestures as phonological units. *Phonology* 6, 201-251.
[15] Browman, C.P. & Goldstein, L. (1992) Articulatory Phonology: an overview. *Phonetica* 49, 155–80.
[16] Kelso, J., Saltzman, E., & Tuller, B. (1986) The dynamical perspective on speech production: data and theory. *Journal of Phonetics* 14, 29-59.
[17] Perkell, J. & Matthies, M. (1992) Temporal measures of anticipatory labial coarticulation for the vowel /u/: within- and cross-subject variability. *JASA* 91(5), 2911–25.
[18] Byrd, D. (1992) Perception of assimilation in consonant clusters: a gestural model. *Phonetica* 49, 1–24.