

HUMAN RESOURCES AND MOBILITY (HRM)  
ACTIVITY

Marie Curie Research Training Networks  
(RTN)

Call: FP6-2005-Mobility-1

**S2S**

Sound to Sense

**PROJECT DESCRIPTIONS**

## Table of Contents

<b>1.</b>	<b>RESEARCH THEMES AND QUESTIONS .....</b>	<b>3</b>
<b>2.</b>	<b>THEME I: MULTI-LINGUISTIC AND COMPARATIVE RESEARCH ON FINE PHONETIC DETAIL (FPD) .....</b>	<b>5</b>
	Project 1: Corpora for discovering language-specific and language-general patterns of FPD .....	6
	Project 2: Automatic methods to identify FPD; Corpus-based probabilistic work .....	8
	Project 3: Perceptual salience of FPD 1 - Segmentation of speech into words and phrases .....	9
	Project 4: Perceptual salience of FPD 2 - Cross-linguistic comparison of contextual sensitivity in speech perception ..	10
<b>3.</b>	<b>THEME II: IMPERFECT KNOWLEDGE/IMPERFECT SIGNAL .....</b>	<b>11</b>
	Project 5: Perceptual coherence and the role of FPD in noise .....	12
	Project 6: Low-level processing in L2 speech perception .....	13
<b>4.</b>	<b>THEME III: BEYOND SHORT UNITS OF SPEECH .....</b>	<b>14</b>
	Project 7: Integration of multiple units in computational models .....	15
	Project 8: Prosodic structure and FPD: segmental-suprasegmental interaction .....	16
<b>5.</b>	<b>THEME IV: EXEMPLARS AND ABSTRACTION .....</b>	<b>17</b>
	Project 9: Hybrid episodic-abstract computational modelling: ASR .....	18
	Project 10: Hybrid episodic-abstract computational modelling: HSP .....	19
	Project 11: Lexical decoding of speech using sub-phonemic features .....	20
<b>6.</b>	<b>SUMMARY OF PROJECT RATIONALE AND STRUCTURE .....</b>	<b>21</b>
<b>7.</b>	<b>RELEVANT PUBLICATIONS FROM S2S PARTNERS .....</b>	<b>23</b>

---

## Tables

Table 1.	Summary information for the Projects. RQs = research questions addressed. ....	4
Table 2.	S2S senior project scientists .....	22

## 1. Research themes and questions

### **Theme I: Multilinguistic and comparative research on fine phonetic detail (FPD)**

1. Can we find patterns of FPD that are general across languages, or to languages within one family (Germanic, Romance, Slavic, Finno-Ugric)? Or are all FPD patterns language-specific?
2. Can the discovery of FPD be automated? Conversely, how can FPD be represented in models of HSP, in ASR, and TTS? Answers to either question will inform work on the other.

### **Theme II: Imperfect knowledge/imperfect signal**

3. FPD may provide ‘perceptual coherence’. To what extent can observed patterns of FPD be explained by principles of auditory perceptual organization, such as grouping by onset synchrony and harmonicity, or as a consequence of the physical properties of articulator dynamics?
4. How do the phonetic context, and contextual-pragmatic information (knowledge about the general situation), influence the use of FPD in understanding degraded speech in L1 and L2?
5. Under what circumstances does systematic variation in FPD increase speech robustness and intelligibility, and how does it do so?
6. How does knowledge about FPD in L1 facilitate or hinder learning to understand L2?
7. If some FPD patterns are general, can they be exploited in L2 language learning? If language-specific patterns of FPD are identified and explicitly taught, is communication improved? How?

### **Theme III: Beyond short units of speech**

8. What durations or ‘units’ of speech must be analysed to use information maximally efficiently?
9. When FPD is taken into account, do all traditionally-distinct levels of linguistic analysis (e.g. intonation, phonological words, lexical items, segments) have similar, multiple, time spans?
10. How can computational/psychological models combine units with very different time spans?

### **Theme IV: Exemplars and abstraction**

11. What FPD is retained in memory and what is lost in the process of abstraction?
12. How can exemplar-based and abstract representations of speech be combined?

**Table 1. Summary information for the Projects. RQs = research questions addressed.**

Theme Project number & title		Coordinators	RQs	Indicative hosts (Fellows, person-months)
<b>I. Multi-linguistic &amp; comparative research on FPD</b>		<b>Baayen, Mattys, Ogden</b>		
1	Corpora for discovering language-specific and language-general patterns of FPD	Baayen, Mattys, Local	1	Charles U, Prague (ER, 22) U. York/Sheffield (ESR, 36) NTNU Trondheim (ESR, 36)
2	Automatic methods to identify FPD: Corpus-based probabilistic work	Baayen, Svendsen	2	Radboud U, Nijmegen (ESR, 36)
3	Perceptual salience of FPD 1: Segmentation of speech into words and phrases	Ernestus, Hawkins, Mattys	1, 7	U Bristol (ESR, 36)
4	Perceptual salience of FPD 2: Cross-linguistic contextual sensitivity	Gaskell, Nguyen	1	CNRS/U Provence, Aix, with U Geneva (ER, 20) U York (ER, 14)
<b>II. Imperfect knowledge/signal</b>		<b>Cooke, Lecumberri</b>		
5	Perceptual coherence and the role of FPD in noise	Cooke, Hawkins	3, 4, 5	U Sheffield (ESR, 36) UTCN, Cluj (ESR, 36)
6	Low-level processing in L2 speech perception	Lecumberri, Giurgiu	4, 6, 7	Basque Country (ESR, 36) Charles U, Prague (ER, 12)
<b>III. Beyond short units of speech</b>		<b>Cutugno, Post</b>		
7	Integration of multiple units in computational models	Cutugno, ten Bosch, Barker	8, 9, 10	Naples (ESR, 36)
8	Prosodic structure and FPD: segmental-suprasegmental interaction	Palková, Post, Nguyen	8, 9 (1)	CNRS/U Provence, Aix (ESR, 36) U Cambridge (ESR, 36)
<b>IV. Exemplars and abstraction</b>		<b>Van Compernelle, Hawkins, Moore</b>		
9	Hybrid episodic-abstract computational modelling: ASR	Van Compernelle, Moore	2, 11, 12	Katholieke U, Leuven (ER, 34)
10	Hybrid episodic-abstract computational modelling: HSP	Norris, Ogden	2, 11, 12	U Cambridge (ESR, 36)
11	Lexical decoding of speech using sub-phonemic features	ten Bosch, Moore	2, 11, 12	Radboud U, Nijmegen (ER, 22)

## 2. Theme I: Multi-linguistic and comparative research on fine phonetic detail (FPD)

**Rationale.** The status and use of FPD is fundamental to all projects. We need to identify FPD in as many languages in the RTN as possible (especially FPD that signals non-phonological linguistic structure e.g. grammatical class and discourse function), specify when it does and does not occur, and test its perceptual salience in different situations. Interactional phonetics, which uses Conversation Analysis (CA) to study FPD in interpersonal communication, links this work to the usual use of speech: everyday conversation. But spontaneous conversational speech is the hardest to collect and to generalise from, and clearer speech styles will be preferred for most ASR and TTS applications for the foreseeable future. Since the type(s) of speech needed for FPD research is an open question, we need to determine the most efficient way to collect those that give most insight.

Four inter-dependent projects focus on gathering new data about FPD. Project 1, led by a phonetician, a psychologist, and a mathematical linguist, will build corpora for FPD research for two main purposes: (a) to compare the types and functions of FPD in speech corpora differing in naturalness, from conversations to read sentences; (b) to provide information about FPD in as many languages as possible. Patterns studied will be chosen from the literature or be demand-driven by other S2S projects. Corpus-building for little-studied languages will be emphasized. When conversational data can be obtained, traditional methods and CA will be used to search for FPD. Phoneticians and morphologists will work together on this. Project 2, led by experts in corpora and their statistical analysis, will refine automated techniques for finding FPD, in order to increase the speed and efficiency of FPD research, thereby reducing fragmented research and acquainting engineers with the new potential phonetic knowledge, including FPD, for ASR. Thus Project 2 is also part of other projects, notably 1. Projects 3 and 4, also led by psychologists and phoneticians, test the perceptual salience of the patterns identified by statistical analysis, Project 3 for the influence of FPD on word segmentation, and Project 4 for perception of assimilation.

**Training outcomes.** Fellows who work solely within Theme 1 will choose paths from Projects 1-4 that are intellectually coherent, train in all skills represented, and involve at least two languages. Such Fellows will be skilled in (1) corpus design for FPD research, (2) auditory and acoustic phonetics, (3) non-segmental phonology (4), CA, (5) automatic processing and analyses of corpora, (6) experimental design, testing (production & perception), analysis. Within these constraints, specialisation is possible. E.g. perceptual experiments could use only one paradigm; synthesis techniques may not be acquired. Concentration on automatic processing and corpus design is acceptable if many languages or speech styles are studied. Fellows could combine a narrow focus in Theme 1 with another Theme. ‘Narrower’ Fellows would help a less experienced Fellow in the area(s) in which s/he is strong, to ensure breadth of training and as professional development.

**Intellectual outcomes.** (1) Knowledge about types of FPD found in different varieties of speech. (2) Richer cross-linguistic understanding. (3) New corpora designed to access FPD will complement existing ones (4) The corpora and algorithms for FPD discovery will be made publicly available via the web, to benefit linguistics, psycholinguists, discourse analysis researchers, machine speech research and applications. (5) Publications will encourage speech scientists in general, and psycholinguistics in particular, to consider natural speech when developing theories and collecting supporting evidence. Fellows trained within S2S will be exceptionally well trained to advance this point of view.

Theme I underpins all S2S work, addresses fragmentation, and is strongly interdisciplinary. Project 1 is particularly large, so more space is devoted to its description, and more Fellows requested.

## Project 1: Corpora for discovering language-specific and language-general patterns of FPD

*Coordinators: Baayen (NL), Mattys (UK), Local (UK)*

*Team: Cenoz, Cooke, Cutugno, d'Imperio, van Dommelen, Ernestus, Ford, Frauenfelder, Giurgiu, Gussenhoven, Hawkins, Koreman, Lecumberri, Meunier, Moore, Nguyen, Ogden, Palková, Post, Svendsen, Volín, Wells*

The S2S partners already use many corpora, ranging from spontaneous conversations between two or more people to tightly-controlled read phrases or isolated words, and from large, multi-speaker collections to just a few tokens of a few structural types from one or two speakers. We will use these corpora when appropriate. Project 1 will take S2S further by serving three related functions: infrastructure service provision, corpus-building in a range of languages, and hypothesis-testing.

**Infrastructure.** Project 1 will provide advice and resources for making corpora for the other S2S projects e.g. for a language for which there is no appropriate corpus (e.g. Basque), or for studying a particular type of FPD, such as a particular morphological distinction. The infrastructure service will be offered to all partners from month 1. It will be carried out in collaboration with the project concerned and needs no detailed description. Collaboration of experts in phonetics, psychology, computation, corpus design, and the main S2S languages should assure wise decisions.

**Corpus-building.** The aim is to discover salient FPD in as many languages as practicable (depending on who the Fellows are). S2S will encourage Fellows to build corpora for focused research in little-studied languages. Examples:

- Basque and Spanish for speech varieties and L2 learning.
- Romanian morphological and grammatical properties offer valuable comparisons with Italian, a much more studied Romance language; spread of palatalisation (involved in the singular/plural system) and connected speech processes at word boundaries, and to enhance Romanian ASR and TTS.
- Norwegian: the interaction of word accent (1 and 2) and global intonation contour, relevant to technology and L2 applications.
- Prague already has a large corpus of different varieties of Czech speech, and will begin to analyse for FPD guided by York and Sheffield and the needs of the other S2S projects Prague scientists undertake.

### **Hypothesis-testing by corpus comparisons.**

(1) **Comparisons of speech style** using standard phonetic and automatic analyses will establish which types of FPD can be found in different types of speech. Many partners believe that FPD's most important perceptual role is in casual speech. However, it is extremely challenging to collect enough tokens of casual speech in the same linguistic context for statistical analysis or application of automatic machine methods such as HMMs. S2S partners have approached this challenge in different ways. Nijmegen uses large corpora (often of spontaneous speech) and tries to control for structural variation using appropriate statistics. But the choice of statistical controls involves many a priori assumptions, and there is debate about what insights are lost by lumping the same words together despite different contexts. At the other extreme, York and Sheffield phoneticians use close analysis of individual utterances in spontaneous speech, generalising across tokens grouped together by their interactive function in the conversation. This produces small sample sizes and problems or perceived problems of generality. In an intermediate approach, Cambridge phoneticians record specially-constructed read texts that have been practised until they sound natural. This controls over the linguistic structures, but writing the texts is difficult, there may be limited numbers of repetitions, and, although the speech has many characteristics of casual speech, it is not spontaneous, and it is rarely in dialogue. Another intermediate solution, advocated by Bristol, is the

Map Task. Two people describe routes to each other using maps showing slightly different landmarks, while neither can see the other's map. This produces spontaneous speech, but although words are repeated, they are mainly content words reflecting where the maps differ, whereas much FPD reflects grammatical and discourse structure. Moreover, the syntactic/prosodic structures of the repetitions are uncontrolled, which means that there is limited scope to model the linguistic structures that mediate observed phonetic patterns.

(2) **Cross-language comparisons** will help to identify language-general vs. language-specific FPD. FPD discovery will be informed by (a) observations of expert phoneticians, (b) patterns of FPD in more-studied languages (Dutch, English, French), (c) needs and recommendations from other S2S projects. Examples include:

- morphological distinctions e.g. Dutch singular vs. plural nouns; English past tense forms compared with non-past-tense 'homonyms' (*massed-mast*) and similar patterns (e.g. *row-rowed-road*);
- segmental distinctions e.g. syllable-level properties associated with voicing in segments of a coda; long-domain reflections of particular segments, such as English /r/ and French vowel harmony;
- interactions between prosody and segments e.g. f0 alignment (timing of changes in the f0 contour) relative to the segmental structure of syllables, morphological structure of words, and higher-level prosodic boundaries; rhythmic (spectro-temporal) differences between productive vs non-productive morphemes (*mistimes-mistakes*).

When conversational and other forms of spontaneous speech are available, two topics will be studied in depth using CA, which offers a way to achieve comparability in spontaneous talk:

(i) **turn-taking**. The phonetic details of pausing in various languages will be examined, looking at the segmental and prosodic details of speech. Speakers sometimes use phonetic detail to project the next consonant type, e.g. by producing a velar closure when the word searched for starts with a velar; and in turn, recipients (listeners) can orient to this by providing a candidate word for the word-search which starts with the consonant projected in the prior turn.

(ii) **speech reduction in context**. This study will consider the detail, distribution and function of segmental reduction in conversational speech, by examining reduction in particular conversational structures. It unifies Nijmegen's reduction analyses with York's interactional and sequential analysis, and with lexis and intonation. Guiding questions include: What are the ranges of variability for any given lexical item or combination of lexical items? Does the range interact with conversational function? How do segmental and intonational features relate to one another? How do speakers and listeners orient to such features of talk as it progresses in time?

**Indicative resources.** Total of 3 fellows (2 ESRs and 1 ER to address fragmentation). For CA analyses, one ESR could be shared between Sheffield and York (close commuting distance) or two could start together. For less-studied languages, one ESR will be based at Trondheim and one ER in Prague. Visits: other countries for languages as appropriate (Basque Country, Cluj, Nijmegen), and York for CA and FPD. Backgrounds: phonetics with strong quantitative backgrounds, or possibly computation with a specialisation in language or speech. Morphology might be done by a psycholinguist especially if ER. Multilingualism an advantage.

**Links.** Underpins all other Themes, but especially other Theme I projects.

Will inform and be informed by the automated FPD discovery processes in Project 2.

FPD discoveries will be tested for perceptual salience when possible in Projects 3 & 4, & 5-8.

## **Project 2: Automatic methods to identify FPD; Corpus-based probabilistic work**

*Coordinators: Baayen (NL), Svendsen (NO)*

*Team: ten Bosch, Bowers, Cooke, Corazza, Ernestus, Giurgiu, Hawkins, Johnsen, Local, Moore, Ogden, Mattys, Strik, Svendsen, Volin*

Manual techniques for the identification of FPD are painstakingly slow, and no techniques for automating its discovery are in common use. However, the development of tools to help find FPD is of critical importance for both linguistics and engineering. For linguistic researchers, automation will vastly speed up the process of determining which aspects of FPD are universal and which are language-specific. For engineers, automating the discovery of FPD may lead to the development of better formalisms for both speech recognition and synthesis. In essence, the search for FPD will force engineers to contemplate trainable architectures which can accommodate those aspects of speech communication which have traditionally been ignored or downplayed in the construction of recognizers e.g. long-duration effects and systematic interactions between suprasegmental and segmental levels.

**Methods.** In the initial phase of this project, a range of statistical learning approaches favoured in automatic speech recognition (e.g. HMMs, ANNs) will be employed to determine the capabilities and limitations of current approaches in the discovery of FPD. To evaluate the performance of existing methods, systems will be trained using material where salient FPD is already known to exist, and a comparison will be made between automated and manually-derived indicators of the fine phonetic distinction. Pilot studies at Cambridge and Sheffield using HMMs to estimate systematic acoustic differences between supposed homonyms such as ‘missed-mist’ have demonstrated that certain automatically-derived measures (such as differences in spectral energy distributions) can correlate well with those captured by manual work, but these studies also suggest that other differences (e.g. segmental durations) are less easy to estimate.

The results of this evaluation will help to focus effort on those elements where existing approaches perform poorly. In the second phase of the project, newer and more advanced statistical learning techniques such as dynamic Bayesian networks and multistream architectures will be employed on the same problems. In addition, statistical techniques used at Nijmegen, and ASR techniques at Trondheim, will trace consequences for speech production and comprehension of a word's a priori lexical probability estimated from spoken corpora, its probability re. competitors in the lexicon, and its probability re. its context. These same techniques will trace the presence of FPD in lexical exemplars.

**Indicative resources.** 1 ESR hosted at Nijmegen, with visits to Sheffield, Trondheim (or Naples if long units are studied) for further computational training; and to Sheffield (again), Cambridge or York for FPD training. Can be assisted by the Nijmegen ER associated with Project 11.

**Links.** Will use and be informed by Project 1's newly-collected speech data and FPD discoveries. Any reliable methods for automated FPD discovery will feed back into Project 1 and be tested in Projects 3 and 4. Through the replacement of the optimisation criterion universally employed in ASR (namely, minimisation of segment error rates) with one based on a closer match to perceptually-salient FPD in listeners, this project will contribute to Theme IV on forging closer links between ASR and HSR. This project entails very close working relationships between linguists and engineers.

### **Project 3: Perceptual salience of FPD 1 - Segmentation of speech into words and phrases.**

*Coordinators: Ernestus (NL), Hawkins (UK), Mattys (UK)*

*Team: Ford, Frauenfelder, Gaskell, Local, McQueen, Meunier, Nguyen, Norris, Post, Volin*

Speech-segmentation research investigates how listeners identify word boundaries in connected speech. Many perceptual/linguistic mechanisms supporting speech segmentation are documented, but evidence for their application to real-life speech is virtually non-existent. Having established the distributional validity and reliability of segmentation cues from spontaneous-speech corpora (Project 1), S2S will fill this gap by testing listeners' reliance on such cues when hearing spontaneous speech. The aim is to provide the most ecologically valid, empirically supported account of speech segmentation to date, thereby helping to model efficient speech understanding.

An original new focus will compare the role of stress-based word segmentation in Czech, Dutch, English, and French. Word-initial stress is a useful heuristic for word segmentation in English, presumably because most common English words are stress-initial. At first sight, the contribution of stress to word segmentation should be even greater in Czech because all Czech words are stress-initial. But the phonetics of stress are so distinct in the two languages that, in practice, one would expect many differences in the contribution of stress to segmentation. For example, English stress makes its largest contribution in adverse listening conditions, but Czech stress, less based on prominence, might contribute relatively less to segmentation in noisy conditions. S2S will test trade-offs between reliability and intelligibility cross-linguistically. Dutch and English share similar distributional and acoustic stress properties, so should show comparable results. French, however, is characterised by fixed word-final lengthening. Thus French shows reliable prosodic patterns, but, like Czech, is unlikely to survive noisy conditions. The above predictions are for L1 listeners. Segmentation patterns will be studied for L2 listeners as well, e.g. English learning French or Czech learning English. Effectiveness of L2 word segmentation should depend on the similarity between L1 and L2 stress patterns and FPD (notion of cue transferability), interacting with auditory salience when in noise (Theme II).

**Methods.** S2S has expertise in standard phonetic and psycholinguistic paradigms (e.g. identification, discrimination, word-monitoring, word-spotting, cross-modal priming, intelligibility in noise), and pause-detection, a paradigm developed by Bristol, validated by York, which reflects degree of lexical processing when the pause is heard. It provides fast reaction times, thought to aid understanding of the time course of processing, and is easily transferable across languages, since the event to detect (silence) is defined similarly for all speakers of all languages.

**Materials** will vary from cross-spliced spontaneous speech, to copy-synthesized speech using PROCYSY (allowing fast spectrotemporal manipulation of natural-sounding synthetic speech). Work on English and Dutch will build on that in the literature. Casual non-spontaneous speech may also be used for full control as appropriate.

**Indicative resources.** 1 ESR based at Bristol. Visits to other hosts dictated by needs: expect Prague, Geneva, Nijmegen and possibly Aix; and to Cambridge or York if PROCYSY is used.

**Links.** To other Theme 1 projects, projects 5 and 8, and will inform Theme IV. Entails collaboration between phoneticians, other linguists (phonologists/morphologists), and psychologists, possibly pairing computational psychologists and experimental psychologists too.

## **Project 4: Perceptual salience of FPD 2 - Cross-linguistic comparison of contextual sensitivity in speech perception**

*Coordinators: Gaskell (UK), Nguyen (FR)*

*Team: Ernestus, Frauenfelder, Meunier, Post, Volin*

Identification of segments and words often requires FPD to be evaluated and potentially re-evaluated in the light of following speech. E.g. *bad girl* may sound more like *bag girl* due to regressive place assimilation. The listener must evaluate the /g/ in *bag girl* relative to the place of articulation of the following consonant. Previous research has shown an exquisite sensitivity to the FPD of assimilation and has highlighted a context-sensitive recognition process for assimilated speech. However, current debate focuses on whether the perceptual system becomes tuned to the assimilation processes specific to a listener's native language. We will address this question by manipulating both the language of the stimuli and the listener's native language.

This project will combine phonetic and psycholinguistic expertise across a range of European languages. We will use pairs and triplets of languages varying in the presence of key assimilatory processes. E.g.: place assimilation (English [present], French [absent]); voice assimilation (Czech [present], Dutch [present], English [absent]). Language triplets will be particularly important because we can then use stimuli for which two different sets of listeners are matched in terms of their experience of the language (i.e., it is a foreign language to both), but experience of the particular assimilation type varies. We can then examine the extent to which perception of the foreign language is shaped by language-specific experience. The influence of FPD on contextual sensitivity will be examined by developing continua with varying degrees of assimilation and examining how native language experience shapes perception across the assimilation spectrum. Computational analyses, developed alongside the behavioural work, predict that perception of mild assimilation will be dominated by language-universal perceptual mechanisms, whereas perceiving stronger assimilations will rely on language-specific compensation.

**Materials.** Interdisciplinary collaboration in the generation of suitable materials representing the assimilation continua in the respective languages will be crucial. This will require a good phonetician to do detailed phonetic analyses to identify dimensions of variation within each assimilation process and for each language. This will allow identification, for each type of assimilation studied, of both the acoustic cues that are shared between languages and those that are particular to each language. It will also allow generation of two types of continua: those that are plausible examples of assimilation in more than one language, and those that are idiosyncratic.

**Methods.** Psycholinguistic methods will range from 'low-level' tasks (e.g. phoneme monitoring, mismatch negativity) to more lexical procedures (e.g. cross-modal priming, word monitoring). The language-dependency of the recognition system may vary with the processing level, with low-level processes being language-general and higher-level processing being more language specific. The range of tasks will ensure that variation of this type is captured.

**Indicative resources.** 14 months of ER at York, then 20 months ER time at Aix and Geneva. Visits to Nijmegen and Prague. This project involves phoneticians and psychologists.

**Links.** Informed by Project 2. Related to Project 6. May inform Theme IV projects.

### 3. Theme II: Imperfect knowledge/imperfect signal

**Rationale.** While much work in FPD and ASR has been applied to clean speech material, “everyday” speech communication can be characterised as involving “imperfect signals”. In the former case, the target speech reaching our ears is likely to be partially masked both by other sources and by reverberations from the target speech source itself. Possible mechanisms for recovering information about the target from the background have been proposed based on the perceptual coherence of components which arise from the same source. Project 5, coordinated by a computer scientist and a linguist, will investigate the interrelation of auditory coherence and FPD. Speech acquisition may well be shaped by the structures delivered by early (and possibly not speech-specific) processes in hearing, which are likely to arrive at higher processing centres having undergone perceptual grouping. Consequently, FPD may be closely related to auditory coherence. Project 5 will investigate this idea using different types of signal degradation (competing speech from familiar and unfamiliar talkers, stationary noise), tasks, and listener populations (L1 vs L2), all of which the partners already use. TTS will build into the synthetic signal attributes of FPD identified as contributing to signal robustness, and assess the effect on intelligibility, varying background noise, semantic coherence, and type of context. Particular effects of FPD can thus be identified, and inform the auditory modelling activities. Some of these methods have been used by partners.

A less-recognised form of distortion results from “imperfect knowledge” of the language being used. In modern Europe, much speech communication takes place amongst non-native speakers and listeners. Project 6, again led by a linguist and a computer scientist, aims to gain a better understanding of the role of interference from the native language when attempting to understand non-native speech by carrying out behavioural tests of L2 perception in a number of conditions, including situations where the signal is degraded by noise. Surprisingly, there are few computational models of non-native perception. The second part of project 6 will build on the behavioural studies of the first part to construct a detailed computational model of L2 perception. A deeper understanding of non-native perception will also highlight the processes which contribute to the robustness of native perception and lead to the development of effective teaching regimes.

**Training outcomes.** (1) Sophisticated technical skills: (i) acoustic phonetics, (ii) formant synthesis, (iii) experimental design, (iv) statistics, (v) programming (vi) ASR (HMMs, ANNs). (2) Strongly interdisciplinary: phoneticians learn computational modelling and auditory psychophysics; engineers/computer scientists learn acoustic phonetics, experimental design, statistics and prosodic phonology.

**Intellectual outcomes.** (1) Better understanding of speech processing in real-world conditions; (2) insights into the role of exposure to noise in both L1 and L2 acquisition; (3) integration of an auditory grouping perspective into polysystemic theories; (4) clearer understanding of universals and language-specific processes in speech perception; (5) development of better regimes for language learning in a wide range of languages used in the European Community.

## Project 5: Perceptual coherence and the role of FPD in noise

*Coordinators: Cooke (UK), Hawkins (UK)*

*Team: Barker, Brown, Giurgiu, Lecumberri, Local, Mattys, Van hamme, Norris, Wells*

Perceptual experiments have shown that synthetic speech is more intelligible in noise when it contains FPD that mimics natural speech patterns. Knowledge of the FPD for one's native language may also result in speech perception advantages over non-native listeners in noise. Intriguingly, FPD that affects intelligibility in noise is often not easily noticeable in good listening conditions. One hypothesis is that the FPD enables better grouping of the auditory signal, and thereby more efficient lexical access and understanding. Perceptually coherent signals should result in structures that are robust in everyday noise conditions (including competing speakers).

This project aims to use computational techniques to build linguistic structures such as prosodic trees on the basis of 'perceptual coherence', that is, the grouping of sound components such as harmonics and formants into larger units. The project will examine the relationship between FPD and auditory perceptual coherence and test the hypothesis that FPD-derived coherence contributes to speech processing in noise. There will be three linguistic-phonetic foci: (1) Short-term spectrotemporal changes, e.g. near segment boundaries, that may contribute low-level auditory coherence. (2) Systematic variation that spreads over several syllables and provides information either about a single phoneme (e.g. /r/ resonance in English, vowel harmony in French) and/or about prosodic structures such as accent groups and intonational phrases (e.g. f<sub>0</sub> contour; strengthening at phrase boundaries). (3) Phonetic variation due to morphological structure as described above. Pilot data analysed in Cambridge and Sheffield suggests that morphological differences may involve elements of (1) and interact with prosodic strengthening (3), so this third focus links low-level auditory processes with higher-order prosodic structure as well as grammar.

Three additional project options are available. 1. To study speech produced in the presence of a N-talker babble for various N, to examine how FPD is affected by speech production changes brought about by noise. 2. To include audio-visual experiments, thus broadening to an ecologically-valid, multi-sensory approach that addresses general perceptual, rather than just auditory, coherence. (3) To use other languages (particularly Romanian) as a testbed for a predictive account of coherence-inducing FPD.

**Methods.** Paired stimuli will be constructed that are identical except for the presence/absence of FPD, and their word intelligibility in noise assessed. Stimuli whose intelligibility is enhanced in noise will be used for the computational modelling. Intelligibility tests will follow standard procedures. Computational auditory scene analysis algorithms and simulations of auditory 'glimpsing' opportunities will be used to determine which parts of the speech signal are most salient in noise. Explanations for improvements in intelligibility will be sought with reference to auditorily-salient information.

**Indicative resources.** 1 ESR with a computing background hosted at Sheffield; 1 ESR with linguistic background based in Cluj. Visits: Cambridge and/or York for FPD and to Leuven for noisy episodic representations. Integrates computer science with phonetics & phonology.

**Links.** Auditory coherence for prosodic structuring links with theme III. Listening in noise will feed into theme IV by addressing the issue of what kinds of episodic representations are useful in noise. Auditory grouping of simultaneous speech is linked to turn-taking in project 1.

## Project 6: Low-level processing in L2 speech perception

*Coordinator: Lecumberri (ES), Giurgiu (RO)*

*Team: Bowers, Cenoz, Cooke, Van Compernelle, van Dommelen, Duběda, Johnsen, Koreman, Mattys, Meunier, Post, Svendsen*

Speech recognition by L2 learners suffers in adverse listening conditions. Several factors are likely to be involved in this performance disparity: (i) interference from the native language; (ii) incomplete acquisition (lack of FPD) of L2 categories; and (iii) presumed universal confusions due to inherent maskability of certain sounds. It is important to tease apart these factors if effective L2 training regimes are to be devised. In addition, knowledge of L2 perception will lead to insights into processes and representations used by native listeners. The purpose of this study is to clarify the role of the above factors via a tightly-linked series of behavioural and computer modelling studies.

**Perception.** To assess the role of L1 interference in L2 perception, common speech perception tasks will be carried out in a range of language communities. In one set of experiments, VCV tokens will be used. Data will be collected from native language groups as well as L2 learners. Amongst other contrasts, we will study the different status and FPD of the interdental fricatives (i.e. the initial sounds in the English words “this” and “think”) in Spanish, Czech and English, and their acquisition by learners in each language group. Results will feed into a computational modelling study described below. A parallel thread will investigate L2 learning of subsegmental lexical information affected by assimilatory processes (e.g. when *bad girl* sounds like *bag girl*; and voicing assimilation in French). Studies of disambiguation of such assimilated sequences by native speakers suggest that FPD must be part of the lexical representation of the word, mediating between acoustic input and phonological representation. We will assess (1) whether L1 assimilation patterns are transferred to L2, whether appropriate or not, (2) whether L2 learners acquire assimilatory contingencies as L1 learners do, i.e. implicitly.

**Computational modelling.** Since non-native listener populations vary widely in the degree and type of exposure to the L2, it is difficult to control for each of the factors which influence L2 speech perception. One solution is to construct a computer model in which L2 ‘exposure’ can be varied continuously. A model can also shed light on the relative roles of native language interference and incomplete acquisition. A computational model will be constructed using standard ASR techniques. To model a L2 learner’s capabilities, different model sets will be trained for the L1 and L2 using speech material for each of the languages used in the perception tests. By varying the amount of training material used to train the L2 model, the process of L2 acquisition will be simulated. For speech in noise data, a missing data ASR model developed at Sheffield will be used to model energetic masking. Consequently, the effect of adverse conditions on L2 speech perception will be tested to analyse the robustness of acquired representations. This modelling study will not only lead to a better understanding of early processes in L2 perception, but may cast light on the reasons for robustness of native perception.

**Indicative resources.** 1 ESR with background in linguistics hosted by Basque Country. 1 year of engineering ER based in Prague. Between these two Fellows, visits to up to 5 other countries for perception testing with L2 learners. Visits to Sheffield, Leuven and/or Trondheim for computer modelling. Links linguistics, phonetics, psychology and computer science.

**Links.** L2 issues permeate all themes. In particular, projects 1, 3 and 4 involve cross-linguistic comparisons and data, with project 4 also addressing assimilation.

#### 4. Theme III: Beyond short units of speech

**Rationale.** Historically, much work in linguistics has attempted to construct a description of the speech signal based on the concatenation of short segmental units. While some currents in linguistics have moved away from the segmental perspective, it remains the dominant paradigm for nearly all work in ASR. In fact, recent efforts have been made to make use of suprasegmental information in ASR, with little success to date. The S2S network provides a timely opportunity for experts in intonation and prosody to contribute insights into the role and perceptual salience of long-term information and variation. The aim of theme III, then, is to initiate the process of moving from a frame-based to a tree-based view of speech in both linguistics and engineering.

To segment speech into long-term units (e.g. words, phrases) adult L1 listeners use linguistic knowledge of the units themselves and of typical patterns of phonetic detail. FPD, context (syntactic, semantic), and language background (L1, L2) can be manipulated to find which aspects of segmentation are knowledge- vs. signal-driven, general vs. language-specific. Project 7 aims to integrate stochastic and cognitive model approaches using segmental FPD, prosodic and syllabic factors, with longer time scales than has been done before, by, for instance, tracking mutual dependencies between the pitch contour and syllabic attributes, like segmental composition of the onset or phonological category of the nucleus, which may facilitate lexical access.

Not only are segmental and prosodic research traditionally separated, but intonologists often focus solely on the  $f_0$  contour, with little regard for co-occurring properties of the speech signal that may crucially influence the way the  $f_0$  contour is interpreted, or even  $f_0$  properties that appear to interact with lexical identification. Segmentalists, on the other hand, often avoid intonational and other prosodic issues except duration. S2S researchers are well placed to overcome this traditional fragmentation, for all phonetics/linguistics groups have contributed to the understanding of both long- and short-domain effects in their respective fields. The challenges of integration are significant, so though Project 8 may appear narrow, it is in fact amongst the hardest in S2S.

**Training outcomes.** Experimental design, semi-automatic acoustic analysis of large sets of speech units, statistics, psycholinguistics. Design, development and expertise on new generation of ASR. Computer programming and signal processing skills in the field of features extraction, corpora evaluation for ASR assessment. ER and ESRs will receive training in the following areas (where necessary; training will be tailored to the Fellow's needs): (i) experimental design; (ii) perceptual testing; (iii) intonational analysis; (iv) phonetic and phonological analysis; (v) data treatment; (vi) statistical analysis; (vii) speech synthesis; (viii) Conversation Analysis and functional approaches to intonation.

**Intellectual outcomes.** A better understanding of (1) the interface between syntax/morphology and phonetics, (2) the integration of lexical-based and prosody-based segmentation, (3) the role of FPD in sound-meaning relations in intonation, (4) the interaction between auditory and visual signals in speech intelligibility. These outcomes will directly benefit research and applications in ASR, speech synthesis and second language learning, and ultimately, they will lead to the first integrated theory of the interaction between suprasegmental, segmental and subsegmental features in signalling prosodic information.

## Project 7: Integration of multiple units in computational models

*Coordinators: Cutugno (IT), ten Bosch (NL), Barker (UK)*

*Team: Bowers, Van Compernelle, Hawkins, Mattys, Moore, Norris, Local, d'Imperio, Strik, Svendsen, Van hamme, Wells*

One of the challenges of making use of FPD in ASR is how to incorporate long-term structures which represent speech dynamics and suprasegmental processes into existing recognizers which employ short-term representations. The frame-based nature of most current work in ASR is at odds with the richer, multiple tree-based representations implied by approaches such as Polysp. Indeed, in ASR, suprasegmental information is typically seen as the source of distracting variation rather than as valuable information. The purpose of this project is to attempt to develop an effective, statistical framework for ASR which is capable of exploiting the information available at multiple time scales. The study has two components. In the first, researchers will build on existing work at Naples and Nijmegen into novel speech feature representations and ASR architectures. The second study is equally adventurous and will examine multimodal FPD.

**1. Multiple units in ASR.** Naples studies multilevel stochastic architectures allowing parallel analysis of speech under different time scales, each making use of different feature sets. Parallel, hierarchical and factorial HMMs are being developed specifically to model speech dynamics. Nijmegen uses a different set of features, but has similar aims. Project researchers will spend time in each other's labs to better understand the differences between the two approaches and will construct hybrid architectures which will be evaluated using a number of different features based on the modulation spectrogram and the temporal evolution of energy and fundamental frequency.

**2. Audiovisual FPD.** In noisy conditions, it is known that listeners become observers, with eyes tracking to the interlocutor's lips. Indeed, the speech reading benefit has been estimated as equivalent to a reduction in the noise level of 15 dB. However, the way that auditory and visual signals interact with respect to fine phonetic detail has not been studied to date. The purpose of this part of the project is twofold: (i) to examine if the generation and use of FPD differs when the receiver has access to visual information, and (ii) to integrate visual information, primarily from the lips and jaw, into the recognition process. This subproject will draw on existing expertise in audiovisual speech recognition at Sheffield in conjunction with the work on multilevel statistical architectures detailed above.

**Indicative resources.** 1 ESR with computational background hosted by Naples. Extended visits to Nijmegen for (1) above and to Sheffield for (2). Other visits to Aix, Trondheim or Cambridge, for phonetics, and/or to Bristol for hierarchical processing, as interests and progress dictate.

**Links.** This project links to project 5 since suprasegmental and multimodal cues are known to be especially important in adverse conditions: for instance,  $f_0$  variation helps to ensure that the target speech stands out against other speech sources in the background, while energy modulations at the rate of syllables and above help speech to resist the masking effects of both stationary and non-stationary noise; visual cues are unaffected by acoustic noise.

This project has the potential to integrate 'prosodic' and 'segmental' phoneticians, phonetic Conversation Analysts, intonational phonologists, computer scientists/engineers and computational and experimental psychologists.

## Project 8: Prosodic structure and FPD: segmental-suprasegmental interaction

*Coordinators: Palková (CZ), Post (UK), Nguyen (FR)*

*Team: Barker, Duběda, D'Imperio, Giurgiu, Gussenhoven, Hawkins, Howard, Local, Ogden, Strik, Wells, Svendsen*

The aim is to elucidate how disparate acoustic parameters covary to cause language-specific interpretations of prosodic properties and conversational functions. S2S partners differ widely in their methods. Intonologists (D'Imperio, Gussenhoven, Palková, Post, Volín) use laboratory phonology to produce *formal phonological* descriptions of f0. Conversation Analysts (Howard, Local, Ogden, Wells) study *functional* roles of correlated phonetic parameters in conversations.

**Methods.** The S2S groups have expertise in virtually all paradigms used to explore prosodic structure and its interaction with segments, and will have easy access to pairs of languages which are more (Dutch-English, French-Italian) or less closely related (initially Czech, extended if possible to other S2S languages). Intonation systems differ in each to a greater or lesser extent. Cross-linguistic comparisons will show to what extent the dependencies are language-specific.

**Production studies** will identify correlations between acoustic parameters reflecting tempo, timing, articulatory setting, voice quality and intonation, and compare functionally- and formally-oriented analyses. There are 3 foci. (1) How systematic variation in FPD influences perceptual parsing of speech into longer linguistic units e.g. intonational phrases. (2) Dependencies in FPD in various types of questions. (3) Covarying parameters that modify categorical interpretations of intonation contours when alignments between segmental and suprasegmental properties change.

**Perception studies** will (1) test hypotheses formulated from the combined results of the production analyses. Various linguistic and psycholinguistic paradigms will be used to test the naturalness of resynthesized stimuli and changes in the interpretation of utterances as a function of FPD parameters. E.g. parameters will be systematically manipulated in brief utterances and placed in their original conversational contexts to assess whether listeners' interpretation changes. (2) Interactions between acoustic and visual cues to prosodic information will be assessed in noise and in quiet; visual cues are expected to aid comprehension in noise.

**Materials.** Groups will use each others' data, which includes much spontaneous speech (conversational and other e.g. dictations, interviews), as well as controlled read sentences. At least three corpora include many different styles (Czech: 400 speakers; English: IViE: 126 speakers; Norwegian: 50 hrs). Most perception experiments will manipulate natural speech via resynthesis (PROCSY, Residual-LPC-based concatenative synthesis, PRAAT PSOLA resynthesis).

**Indicative Resources.** 2 ESRs, hosted by Aix and Cambridge. With strong backgrounds in at least two of the following, they will learn about the others: CA, intonational phonology, acoustic phonetics, (audio-visual) experimental design/analysis. Visits appropriate for the ESR's profile to: Sheffield/York for CA & audio-visual; Nijmegen, Prague or elsewhere for intonation, data collection and perception experiments. If feasible, Prague's ER will help and mentor both ESRs.

**Links.** Project 8 addresses severe fragmentation within four areas of linguistics: CA, intonational phonology, prosodic phonetics, segmental phonetics. Thus it has direct links to Theme I, projects 5 (perceptual coherence) and 7.

## 5. Theme IV: Exemplars and abstraction

**Rationale.** Not only is ASR performance much worse than HSP, but improvements in state-of-the-art ASR systems are asymptoting to a level of performance far below that required for many practical applications. Yet research in HSP has seen the introduction of computational models based on search algorithms that strongly overlap with the techniques used in contemporary ASR. Unfortunately, such models typically use abstract symbols as input rather than real speech signals. In consequence, and also because the emphasis is typically on word recognition, potentially informative FPD is neglected: HMMs' probability density functions discard FPD because they lack the linguistic-phonetic structures to map it to, while standard HSP models can give no role to *phonetic* detail because they use abstract input symbols that neglect it.

The potential relevance to ASR of research on episodic memory and hence on perception of FPD has already been realised by a few researchers, and some early work shows promise. Likewise, some researchers, including those in S2S, are trying to 'bridge the gap' between ASR and HSP. This Theme expands these lines of investigation by linking the key research centres with each other and with phoneticians, and will provide a major contribution to a unified view of speech recognition.

The three Theme IV projects share the above aims but take different approaches. [Project 9](#) seeks to further improve its recent combined episodic and abstractionist SR model. The focus for the first half of the project will be on weighting acoustic-phonetic information relative to more abstract knowledge. Later this project may seek to incorporate systematic linguistic knowledge fed to it by Project 10, using the bottom-up information in ways that humans are thought to. [Project 10](#) takes a similar conceptual approach, but for HSP. It will adapt an existing computational model of spoken word recognition, SHORTLIST, to use more complex, flexible data structures that represent the linguistic knowledge signalled by FPD. These will be Prosynth prosodic structures, developed for TTS and applied to speech perception: The Prosynth/Polysp framework has abstract structure with FPD properties attachable at all nodes in the structure. Which node a feature and its value is placed on in a tree varies according to the type of structure which is described (e.g. a content word vs. a function word; a morphological boundary that coincides with a syllable boundary vs. one that does not). [Project 11](#) will design and build a computational model for HSP using subphonemic features as input. Lexical access and competition will be studied using search techniques (exemplars) from ASR. This model will extend the SpeM model of HSP that has been developed in Nijmegen. An aim is to simulate key effects observed in HSP while avoiding probabilistic symbolic input representation.

Towards the end of the RTN the three projects will combine the insights gained to propose a single, elegant solution to the representation of linguistic-phonetic knowledge for ASR and HSP. If the Polysp approach is validated, then we will be able to offer a formal solution to the polarization of the episodic/abstractionist debate.

**Training outcomes:** (i) ASR development and evaluation (ii) Models of HSP with an emphasis on the differences/complementarities of episodic/abstractionist modelling; (iii) Experimental design used in HSP; (iv) Acoustic phonetics; (v) PROCSY; (vi) corpus analysis.

**Intellectual outcomes:** (i) an ASR concept that combines abstractionist and exemplar modelling; (ii) understanding of the respective roles of early abstraction and attention to FPD.

**Project 9: Hybrid episodic-abstract computational modelling: ASR**

*Coordinators: Van Compernelle (B), Moore (UK)*

*Team: ten Bosch, Cutugno, Demuyneck, Hawkins, Norris, Strik, Svendsen, Van hamme*

Two fundamentals of traditional HMM ASR systems, generalization of phoneme-like units and a top-down search procedure, are in conflict with current understanding of HSP, viz. the preservation of FPD and the importance of bottom-up processing. Properties of recognition of novel words and L2 learning require early abstraction of (sub)-phoneme-like units; while other phenomena hint at detailed segmental matching. Two recent projects at Leuven have tried to embed these concepts in new ASR architectures. Within the TEMPLATE project the recognition units are of arbitrary length and the matching paradigm is exemplar based. The FlaVoR project explores a new search paradigm for an HMM based ASR in which bottom-up (data-driven) and top-down (knowledge-driven) processes are combined. Recently, very preliminary, though impressive, results have been obtained by a combination of both systems. A bottom-up phonemic recognizer (abstractionist model) defines the search space. In the final recognition (ASR search process) scores from multiple knowledge sources (acoustic and linguistic) are combined. The acoustic contributors are the score from the phonemic recognizer and the score from an exemplar based recognizer that incorporates fine phonetic detail and longer speech units into the recognition process.

**Methods.** The bottom-up phoneme recognizer can rely on gross properties (traditional HMM) or it can examine FPD (e.g. segmental recognition). The former is likely to have the best generalization behaviour, but the latter will work best in the case of sufficiently similar examples in the database. Advantages and shortcomings of both approaches will be evaluated. The final recognition needs a weighting of the HMM and exemplar scores. We will investigate which factors influence the weighting function. The L1/L2 situation forms a specific test case in which some or all of the components may be fully derived from L1, while only a limited number of the components are adapted to L2.

An ER is needed because this work is highly innovative, especially the aim of comparing L1 and L2 recognition. In addition, most Fellows engaged in computational projects are likely to visit Leuven, so the ‘resident’ Leuven Fellow will play a significant mentoring and helping role throughout Theme IV, as well as for other projects, e.g. project 5.

**Indicative resources.** 1 ER with an engineering or computational background, hosted by Leuven. Visits: substantial periods at Cambridge to learn acoustic phonetics including elements of linguistic structuring of FPD with Hawkins, and HSP with Norris and Hawkins; also to Sheffield to enrich experience of ASR approaches to modelling episodic representations (Moore).

**Links:** Projects 7, 10, 11. This project links engineers/computer science with phonetics.

## Project 10: Hybrid episodic-abstract computational modelling: HSP

*Coordinators: Norris (UK), Ogden (UK)*

*Team: Bowers, Cooke, Ford, Hawkins, Mattys, Moore, Nguyen, Van Compernelle*

Although standard HSP models have had considerable success in simulating some of the basic phenomena of spoken word recognition, any further progress will be critically dependent on constructing models that pay proper attention to the FPD available in the speech signal. This project aims to achieve this by constructing a new computational model that uses FPD to guide the construction of prosodic and grammatical structures that can drive the recognition process.

Computationally, Prosynth/Polysp prosodic trees are XML structures with nodes in the prosodic tree linked to nodes in the grammatical tree. The aim is to establish whether these trees and their relationships are all that are needed to model speech perception. If so, an experienced listener can be seen as placing feature values on the nodes of a general tree or linked set of trees. The main difference between this and a standard perception model is that the FPD maps onto every type of node, at any height in the tree, not just the bottom level of the tree that in a standard model specifies phonemes/phonological features. This conceptually simple difference is challenging to model.

**Methods and materials.** We will progress from a ‘proof of principle’ evaluation of Prosynth/Polysp, to the development of a complete computational model of speech recognition.

In phase 1 we will construct a modelling framework with flexible, complex data structures that can represent prosodic, phonological, morphological and lexical information, their inter-relationships, and the FPD that accompanies these structural distinctions. Flexibility is important as the representations specified by Prosynth/Polysp will evolve during the course of the project. Because each utterance type has a unique structure, we will work first with a small set of structures whose FPD is perceptually salient, so that clear predictions can be made based on behavioural data.

In phase 2, the model will be tested on annotated handcrafted input representing the FPD that we know is extracted from the signal. This will be used to incrementally construct Polysp structures consistent with the current input. These structures, the values of whose nodes will be assigned probabilistic weightings according to their consistency with the input, will then be used to constrain all hypotheses about the signal’s linguistic structure: featural, segmental, syllabic, prosodic, grammatical and lexical. The structured representations will be revised and updated as new evidence arrives in the same kind of updating and re-evaluation that occurs in existing ‘abstractionist’ models like SHORTLIST, but with the advantage of being further constrained by attention to FPD and knowledge about what it signifies. Standard search procedures will be used to generate a continuously evolving measure of the probability of alternative structural hypotheses (cf the methods used in SpeM for deriving ‘word activation’ scores). Weightings for specific nodes and properties will be derived from perceptual data and judgments of the auditory salience of specific utterances’ acoustic properties in the ambient noise conditions.

In phase 3 we will try to incorporate automatic extraction of a limited set of FPD features from real speech, using easily-segmentable utterances with differing grammatical status, so that the focus is not on word recognition per se, but on the pattern and temporal order in which systematically varying FPD is mapped onto the nodes of the data structures.

**Indicative resources.** 1 ESR with a background in computer science, computational linguistics, or engineering, hosted at Cambridge. Visits: (i) York, to learn prosodic phonology (ii) Leuven to learn hybrid ASR computational modelling techniques.

**Links:** This project has close but complementary links with projects 7, 9 & 11 and can use FPD insights from theme I and, in time, project 8. Primary link between psychological modelling and experimental phonetics; anticipated extensions to engineering models of ASR.

## Project 11: Lexical decoding of speech using sub-phonemic features

*Coordinators: ten Bosch (NL), Moore (UK)*

*Team: Baayen, Ernestus, Gussenhoven, Hawkins, Ford, Van Compernelle, Strik*

This project addresses the lexical coding and decoding of FPD by means of probabilistic subphonemic feature representations that are automatically derived from the speech signal. Subphonemic feature vectors for each 10 millisecond interval in the unfolding speech signal provide an excellent window on the fine phonetic detail across many acoustic and articulatory dimensions. The main aim is to improve computational modelling of HSP by using key techniques from ASR. The starting point will be existing models of ASR and HSP, such as the conventional HMM-based ASR models (Nijmegen), exemplar-based models (Leuven & Sheffield), and the ASR-based model of HSP called SpeM (Nijmegen). For three related issues (see below) we will first carry out a corpus-based survey, followed by psycholinguistic experiments on production and comprehension. This work is interdisciplinary and also addresses fragmentation within linguistic phonetics, and within computational modelling. It will be carried out in collaboration with Cambridge, Leuven and Sheffield.

**Methods.** We will use subphonemic feature vectors to address the extent to which FPD mediates lexical competition in speech comprehension and language production. Recent studies show that listeners are highly sensitive to FPD present in laboratory speech, and phonetic analyses of speech corpora have shown that the fine phonetic detail of acoustic realisations in speech corpora bear the traces of lexical competition in the speakers' mental lexicons. Almost all studies focused on durational differences. The sequences of probabilistic feature vectors representing speech tokens provide a unique opportunity to investigate a much broader palette of potential differences in FPD, ranging from place and manner to nasality and intensity of frication. This will be applied e.g. to (1) the FPD found for the Dutch singular-plural system (morphology) and (2) temporal alignment of features of prosodic information, in particular the intonation contour, with subphonemic detail at the segmental level, with special emphasis on lexical disambiguation during comprehension, which is a completely uncharted domain of inquiry.

Year 1 will involve corpus-based research on acoustic encoding, focused on FPD in lexical competition in Dutch and English, in close collaboration with Cambridge and the Nijmegen ESR assigned to Project 2. Beginning also in Year 1, we will carry out comprehension and production experiments to assess the role of FPD (including prosodic alignment) in mediating lexical competition, and in the singular/plural system, in Dutch and English. In Year 2, we will focus on computational modelling for ASR and HSP, especially comparing lazy and greedy learning algorithms to evaluate performance that uses more ('episodic') or less ('abstract') signal information.

**Indicative resources.** One ER, hosted by Nijmegen, with a mathematical/computational background (or with a linguistics background and strong scientific or mathematical and statistical skills, and experience of or aptitude for programming). Visits: (i) Cambridge, to study FPD in morphology; (ii) Cambridge or Sheffield, to study prelexical representation; (iii) Leuven, to learn about alternative modelling techniques.

**Links.** Theme I (especially Projects 2 & 3), Projects 8, 9, 10. This work links computational modelling with phonetics and morphology; and addresses fragmentation within linguistic phonetics, and within computational modelling.

## **6. Summary of project rationale and structure**

Each Fellow can select an individually-tailored path. S/he might spend a year working on a particular phonetic issue in a particular language, to learn about how FPD works, and two years contributing to the design of a hybrid episodic-abstract model. This fellow would therefore select from two Projects (1 and 10). S/he might spend months 1-6 on Project 1, months 7-22 on Project 10, months 23-26 on Project 1 again to investigate questions arising from the modelling (possibly in a different phonetics department), and the remaining time mainly in the host institution finishing the modelling and writing up a PhD thesis. S/he would normally be expected to have a background in computer science/engineering, and, in addition to working in at least one phonetics department, would normally travel to more than one institution to learn specific computing techniques. There would be some flexibility in choice of host institution. If the project were conducted mainly in just two institutions, the host institution would be the one housing the computer modelling department. But if, say, the Fellow's main phonetics department and secondary computer modelling department were in the same institution, then this institution could be equally appropriate as the host, depending on the exact balance of supervisory expertise required.

The second alternative above would be expected to be more appropriate for ERs than ESRs. An ER choosing this alternative would be expected to foster links for ESRs between phonetics and computer modelling, and/or between fragmented branches of the same discipline e.g. intonational and segmental phonetics. Such an ER might be based in NTNU or Sheffield, which have both phonetics and computer modelling, but work on long-domain problems with Naples, for example.

**Table 2. S2S senior project scientists**

For each site, the individual whose name is underlined is the scientist in charge at that site. Each scientist's main areas of expertise is in parentheses after each name.

Cambridge (UK)	<b><u>Prof. Sarah Hawkins</u></b> (S2S Coordinator) (acoustic phonetics, FPD, perception, adverse conditions, formant synthesis) <b>Dr. Mike Ford</b> (psycholinguistics, spoken word recognition, morphology) <b>Dr. Michele Miozzo</b> (neurolinguistics, psycholinguistics, morphology) <b>Dr. Dennis Norris</b> (spoken word recog, computational perception models, memory) <b>Dr. Brechtje Post</b> (phonetics, phonology, morphology, intonation, French, L2)
K.U.Leuven (B)	<b><u>Prof. Dirk Van Compernelle</u></b> (exemplar-based ASR, speech recognition architectures) <b>Prof. Hugo Van hamme</b> (robust recognition, missing data techniques) <b>Dr. Kris Demuyne</b> (ASR search)
Charles Univ. Prague (CZ)	<b><u>Prof. Dr. Zdena Palková</u></b> (syntax, intonation, prosody) <b>Dr. Jan Volín</b> (intonation, L1-L2 interaction) <b>Dr. Tomáš Duběda</b> (rhythm)
Aix-en-Provence (FR)	<b><u>Prof. Noël Nguyen</u></b> (artic. & acoustic phonetics, especially French, FPD, perception) <b>Dr. Mariapaola D'Imperio</b> (prosody, intonation, tonal alignment, lab phonology) <b>Dr. Christine Meunier</b> (acoustic phonetics, cross-linguistic production and perception)
Naples (IT)	<b><u>Dr. Francesco Cutugno</u></b> (synthesis & recog., HSP & ASR models, multilevel annotation) <b>Dr. Anna Corazza</b> (nat. lang. proc., machine learning, statistical POS, syntactic parsing)
Radboud & MPI (NL)	<b><u>Dr. Louis ten Bosch</u></b> (ASR, phonetics, articulatory features, human-machine interaction) <b>Prof. dr. Carlos Gussenhoven</b> (general and experimental phonology, prosody, intonation) <b>Dr. Helmer Strik</b> (ASR, pronunciation variation & assessment, episodic speech recognition) <b>Prof. dr. Harald Baayen</b> (psycholinguistics, lexical statistics, morphology) <b>Dr. Mirjam Ernestus</b> (psycholinguistics, phonetics, casual speech)
Trondheim (NO)	<b><u>Prof. dr. Wim van Dommelen</u></b> (phonetics, Foreign Language prod & perc, hearing in noise) <b>Prof. Torbjørn Svendsen</b> (speech processing, ASR, TTS, pronunciation modelling ) <b>Dr. Magne H. Johnsen</b> (machine speech and pattern recognition) <b>Dr. Jacques Koreman</b> (phonetics, voice quality, speech & speaker recognition)
Cluj-Napoca (RO)	<b><u>Prof. Dr. Mircea Giurgiu</u></b> (engineering methods for ASR, TTS, web-based applic' tns) <b>Dr. Luciana Leev</b> (automatic phonetic transcription, morphology, syntax, intonation)
Basque Country (ES)	<b><u>Dr. María Luisa García Lecumberri</u></b> (phonetics, L2 acquisition, adverse conditions) <b>Prof. Jasone Cenoz</b> (multilingualism, psycholings, language planning, L2 acquisition)
Geneva (CH)	<b><u>Dr. Ulrich Frauenfelder</u></b> (psycholinguistics, logopedics, cognitive psychology)
Bristol (UK)	<b><u>Dr. Sven Mattys</u></b> (cognitive psychology, psycholinguistics, HSP, language development) <b>Dr. Jeff Bowers</b> (cog psych, reading, computational modelling, speech perception)
Sheffield (UK)	<b><u>Prof. Martin Cooke</u></b> (comp. models of HSP and scene analysis, robust ASR) <b>Prof. Bill Wells</b> (speech development and difficulties; phonetics; interactional analysis) <b>Dr. Sara Howard</b> (children's speech disorders; auditory phonetics; electropalatography) <b>Prof. Roger Moore</b> (sp. tech. algorithms, applic' tn, assessm't; bridging HSP & ASR) <b>Dr. Jon Barker</b> (robust ASR, audiovisual speech perception) <b>Dr. Guy Brown</b> (computational auditory scene analysis, auditory modelling)
York (UK)	<b><u>Dr. Richard Ogden</u></b> (auditory phonetics; interactional analysis; Finnish) <b>Prof. John Local</b> (auditory phonetics; interactional analysis; attitude) <b>Dr. Gareth Gaskell</b> (experimental psycholinguistics, spoken word rec, comp. modelling)

**Other contributors:** Many partner institutions include other experts, not named above, who will help and advise Fellows as appropriate.

## 7. Relevant publications from S2S partners

### Cambridge

- Hawkins, S. (2003) Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, 31, 373-405.
- Scharenborg, O., Norris, D., ten Bosch, L., & McQueen, J. M. (2005) How should a speech recognizer work? *Cognitive Science*, 29, 867-918.
- Grabe, Esther, Carlos Gussenhoven, Judith Haan, Erwin Marsi and Brechtje Post (1998) Preaccentual pitch and speaker attitude in Dutch. *Language and Speech* 41, 63-85.

### K. U. Leuven

- M. De Wachter, M. Matton, K. Demuynck, P. Wambacq, R. Cools, D. Van Compernelle (2005). Template Based Large Vocabulary Recognition. conditionally accepted for publication by *IEEE Trans. Speech Audio Proc.*
- K. Demuynck, T. Laureys, D. Van Compernelle and H. Van hamme (2003). FLVoR: a Flexible Architecture for LVCSR. In *Proc. Eurospeech*, 1973-1976.
- H. Van hamme. (2004) PROSPECT Features and their Application to Missing Data Techniques for Robust Speech Recognition. In *Proc. Interspeech*, 101-104,

### Charles University, Prague

- Palková, Z. (2004) The set of phonetic rules as a basis for the prosodic component of an automatic TTS synthesis in Czech. In: Z. Palková & J. Janíková (Eds.), *AUC - Philologica 1/ 2004: Phonetica Pragensia X*, pp. 33-46.
- Volín, J. (2005) Rhythmic properties of polysyllabic words in British and Czech English. In: J.Čermák, A.Klégr, M.Malá & P.Šaldová (Eds.) *Patterns*, pp. 279-292. Praha: KMF.
- Duběda, T. & Keller, E. (2005) Microprosodic aspects of vowel dynamics—an acoustic study of French, English and Czech. *Journal of Phonetics* 33/4, pp. 447-464.

### Aix-en-Provence

- D'Imperio, M., Espesser, R., Loevenbruck, H., Menezes, C., Nguyen, N., & Welby, P. (2005) Are tones aligned with articulatory events? Evidence from Italian and French, in Cole, J., & Hualde, J. (eds.) *Papers in Laboratory Phonology IX: Change in Phonology* (Mouton de Gruyter, The Hague).
- Nguyen, N., & Fagyal, Z. (2006) Acoustic aspects of vowel harmony in French, *Journal of Phonetics* (cond. accepted).
- Meunier, C. (2005) Invariants et variabilité en phonétique. In Nguyen, N., Wauquier-Gravelines, S., & Durand, J. (eds.) *Phonologie et phonétique: Forme et substance* (Hermès, Paris) pp. 349-374.

### Naples

- Corazza A., Ten Bosch L. (2003) Learning rule ranking by dynamic construction of context-free grammars using AND/OR graphs". *Proceedings of Eurospeech* 03, Geneva.
- Cutugno F., Coro G., Petrillo M. (2005) Multigranular scale speech recognizers: technological and cognitive view". *Lecture notes in Computer Sciences*, Springer-Verlag, 3673:327-331.
- Cutugno F., Fougeron C. A computer-based tutorial on Models of Speech Perception. In *Proceedings of Method and Tool Innovations for Speech Science Education (MATISSE)*, April 16-17 1999, London, 85-88, 1999.

### Nijmegen

- Gussenhoven, C. (2004) *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.
- M. Pluymaekers, M. Ernestus & R.H. Baayen (2005) Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America*, 118(4), 2561-2569.
- O. Scharenborg, D. Norris, L. ten Bosch & J.M. McQueen (2005) How should a speech recognizer work? *Cognitive Science: A Multidisciplinary Journal*, 29(6), 867-918.

### Trondheim

- van Dommelen, W.A. & Werner, S. (2006) The effect of speaking rate on perceived quantity in Finnish and Norwegian reiterant speech. *To appear* In Gösta Bruce and Merle Horne (eds.), *Nordic Prosody, Proceedings of the IXth Conference*, Lund 2004. Peter Lang, Europäischer Verlag der Wissenschaften, Frankfurt am Main, 67-75.
- Koreman, J. (2006) Perceived speech rate: The effects of articulation rate and speaking style in spontaneous speech. *Journal of the Acoustical Society of America* 119, 582-596.
- T.Holter, T. Svendsen (1999) Maximum likelihood modelling of pronunciation variation, *Speech Comm.*

### Cluj-Napoca

- Giurgiu, M. (2003) Information Retrieval System based on Romanian Continuous Speech Recognition, *Proc. of 2003 IEEE Int. Conf. on Systems, Man & Cybernetics (SMCC'03)*, Washington D.C. Vol.2, pp.1104-1109.
- Giurgiu, M. (2004) Wavelet Analysis and Teager Energy Operator Applied for Speech Recognition in Noisy Environments, *Proc. of Int. IEEE Conference Communications 2004 (ISBN: 973-640-036-0)*, Bucharest, 181-184.
- Vogliano, G. and Giurgiu, M. (2005) An Approach for Non-linear Voice Activity Detection: data from Romanian and Italian, *Proceedings of IPSI 2005 Conference*, Spain. Abstracts pp.19-20, CD ISBN: 86-7466-117-3

**Basque Country**

- García Lecumberri, M.L. & Cooke, M. (in press) Effect of masker type on native and non-native consonant perception in noise. *Journal of the Acoustical Society of America*.
- García Lecumberri, M.L. & Gallardo, F., (2003) English FL sounds in school learners of different ages. In *Age and the Acquisition of English as a Foreign Language*, MP García Mayo, ML García Lecumberri, (eds.). Clevedon, UK: Multilingual Matters: 115-135.
- Cenoz, J. (2003) The additive effect of bilingualism on third language acquisition: A review. *The International Journal of Bilingualism*, 7:1, 71-88.

**Geneva**

- Dufour, S. & Frauenfelder, U. H. (2007) L'activation et la sélection lexicales lors de la reconnaissance des mots parlés: modèles théoriques et données expérimentales. *L'Année Psychologique*, 107, 87-112.
- Dufour, S., Nguyen, N. & Frauenfelder, U.H. (in press) The perception of phonemic contrasts in a non-native dialect. *Journal of the Acoustical Society of America*.
- Franck, J., Lassi, G., Frauenfelder, U., & Rizzi, L. (2006) Agreement and movement: a syntactic analysis of attraction. *Cognition*, 101, 173-216.

**Bristol**

- Mattys, S.L., White, L., & Melhorn, J.F (2005) Integration of multiple segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, 134, 477-500.
- Mattys, S.L., Jusczyk, P.W., Luce, P.A., & Morgan, J.L. (1999) Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, 38, 465-494.
- Bowers, J. S., & Schacter, D. L. (1990) Implicit memory and test awareness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 404-416.

**Sheffield**

- Newton, C., Wells, B. (2002) Between word junctures in early multiword speech. *J. Child Language* 29, 275-299
- Maier V and Moore R K. 'An investigation into a simulation of episodic memory for automatic speech recognition', *Proc. INTERSPEECH 2005* Lisbon, 5-9 September (2005)
- J. Barker, M.P. Cooke and D.P.W.Ellis (2005) Decoding speech in the presence of other sources. *Speech Communication* 2005, 45:5-25
- Cooke, M.P. (2006) A glimpsing model of speech perception in noise. *Journal of the Acoustical Society of America* 119, 1562-1573.
- Barker, J. and Cooke, M.P. (2007) Modelling speaker intelligibility in noise. *Speech Communication* (in press)

**York**

- Gaskell, G. (2003) Modelling regressive and progressive effects of assimilation in speech perception. *J. Phonetics* 31: 447-463.
- Local, J. (2003) Variable domains and variable relevance: interpreting phonetic exponents. *J. Phonetics* 31: 321-339.
- Ogden, R. (2001) Turn transition, creak and glottal stop in Finnish talk-in-interaction. *Journal of the International Phonetic Association*, 31(1):139-152.