



OXFORD SOUND DAY

POSTER SESSION

5 March 2011

Organizing Committee:

Anastassia Loukina

Anna Stirr

Sergio Grau Puerto

Lauren Hall-Lew

The workshop is sponsored by the Career Development Grant
Humanities Division, University of Oxford

Contents

May Chan, “Analysing consonantal speech sounds: a sociophonetic approach”	3
Steven Chance, “Chimpanzee-human neuro-anatomical differences in auditory association cortex”	3
P. García-Bellido, A. Benítez-Burraco, K. Park, B. Molineaux, “Timing the integration of utterance duration and task shift in a case of genetic anomaly implicating 7q31 with language disorders.”	4
Rob Godman, David Mapp and Lee Richardson. “The making and performance of Alyona”	4
Grace de la Flor and Marina Jirotko, “Understanding observable order for software design: using the tools and techniques of conversation analysis”	5
Sergio Grau, “Automatic Speech-to-Phoneme alignment of the Spoken British National Corpus”	6
Marinda Hagen, “The effect of singing on the Perception of Dutch-accentedness in English”	6
Lauren Hall-Lew, “Sound and Sociophonetics”	7
Zoe Handley, “Second Language Speech Learning”	7
Greg Kochanski, “Quasi-duration: Quantifying how rapidly sounds change.”	8
Alexander Krasovitsky, “Sound change across generations: evidence from North Russia”	8
Noel Lobley, “Recording and remembering the sounds of Africa: ethnomusicology, sound archiving and sound elicitation”	9
Anastassia Loukina and Greg Kochanski, “Measuring linguistic rhythm”	9
Kevin R. Page. “Linked Data, e-Research, and Music Information Retrieval”	10
Zeynab Raeesy, “Acoustic-to-Articulatory Inversion based on Dynamic MRI”	10

Jonathan Roberts, “Sounding Authentic: Helping English People Sing Javanese”	11
Duncan Williams, “Evaluating timbre in location recordings”	11

May Chan, “Analysing consonantal speech sounds: a sociophonetic approach”

Spectrography has been employed for the analysis of speech sounds in the last decades, and many significant discoveries about speech has been made using this technique. This poster will present techniques which are employed in spectrographic analysis of consonantal speech sounds in sociophonetic studies. The formant profiles of various consonants as well as the question of how one particular ‘consonant’ can be realised in different ways will be explored. The aim is to uncover the acoustic properties of consonantal sounds which are significant in explaining how speakers present their own identities and perceive those of others.

Steven Chance, “Chimpanzee-human neuro-anatomical differences in auditory association cortex”

Background: Asymmetry of minicolumnar stacks of cells in auditory cortex may underlie functional hemisphere contrasts in sound processing (temporal vs. spectral) in humans. Minicolumns in human Planum Temporale (PT), in Wernicke’s language area, are wider in the language-dominant, left hemisphere, unlike other primates. However, previous studies in chimpanzees did not analyze age effects.

Methods: Minicolumns, neurons, and glial cells were measured in post-mortem PT cortex (BA22) from both hemispheres of 15 chimpanzees and 15 humans. (Age ranges: chimpanzee: 1-57 years, human: 2-57 years).

Results: Minicolumns were wider* and neurons were larger in human PT compared to chimpanzees. Absence of minicolumn asymmetry in chimpanzees was confirmed*. Human minicolumn width peaked in adulthood, then declined with age, whereas the chimpanzee trajectory declined from an early peak coincident with or prior to peak brain size and sexual maturity. (* $p < 0.05$)

Discussion: Our findings indicate that developmental synchronicity of brain regions differs between hominoid species: the PT matures late compared to the rest of the brain in humans, whereas, in chimpanzees, the PT matures at about the same time as other areas. Asynchronous development of brain regions may have increased in human evolution, with a particularly late maturational peak in asymmetric language areas.

P. García-Bellido, A. Benítez-Burraco, K. Park, B. Molineaux, “Timing the integration of utterance duration and task shift in a case of genetic anomaly implicating 7q31 with language disorders.”

The most basic biological endowment needed for communication is that which allows the brain to copy sounds. This requires having an auditory pathway that responds to changes in the environment and a motor system that times muscular changes within a limited time scale. In this experiment we measure the significant differences observed between the performance of a subject (A), affected by a genetic anomaly (a t[7;11] [p13;p13] along with a pericentromeric inversion with rupture points in 7p13 and 7q31), and a control subject (C) matched for sex, age, schooling, and linguistic and socioeconomic background. For this experiment the subjects repeat 253 different nonsensical utterances. We measure the deviation of each subject from each given utterance with respect to two clocks: the total utterance time and the number of task shifts. We look at 9 possible integrative types these two clocks produce and determine how A and C differ in each of them. Since 7q31 is the *locus* for *FOXP2* (which could be consequently mutated in A), this experiment could be of interest for understanding how genes affect pathways involved in regulating the timing of muscular activity as a response to auditory stimulation.

Rob Godman, David Mapp and Lee Richardson. “The making and performance of Alyona”

Алёна is a 1-hour continuous set, presented as a live performance of unlikely juxtapositions of poetry, hip-hop and immersive sound. Our aim is to engage audiences through performance and documentation by offering unique and unrepeatably live events where the use of technology is transparent and visual.

David says: *Jung describes the anima and animus as elements of his theory of the collective unconscious, a domain of the unconscious that transcends the personal psyche. The central concept of the piece is to present acousmatic immersive sound as the anima of hip-hop. The composition of the*

script then, is a matter of exaggerating those aspects of rap, which the (unbalanced!) narrator perceives as masculine, and counterbalancing them (to use Jung's language) with a prose-style suitable for reciting over the immersive sound. So, the hip-hop poetry becomes aggressive, angry, crass, vulgar and, in short, explicit, whilst the prose becomes contemplative, ambiguous, mysterious and, in short, implicit. Whilst in the rap they revolve around hatred and vengeance, in the prose-poetry they revolve around love and reconciliation.

This poster discusses the interdisciplinary relationships found within [Alyona] and methods of delivery for linear narrative including aesthetic considerations of 'liveness'.

<http://www.robgodman.com/RobGodman/Alyona.html>

Grace de la Flor and Marina Jirotko, “Understanding observable order for software design: using the tools and techniques of conversation analysis”

When designing software for a community of skilled practitioners it is important to understand their requirements for new technological systems. This understanding must inevitably be based upon the ways in which the members of a community themselves organise work activities, communication and collaboration. In the domain of Requirements Engineering we uncover the organisation of work through sociologically-informed methods, particularly ethnomethodology, conversation analysis and interaction analysis.

For this poster, we will present the qualitative methods used to understand conversational data within its context of production. Conversation analysis is used to analyse the sequential organisation of talk in order to reveal the ways in which mutual intelligibility is achieved amongst the participants of a setting. Through an analysis of reflexively emerging interaction we not only analyse the sounds produced by co-conversationalists as publicly observable/hearable phenomena but also the language games that link the rules and meaning of language to its situated use and the mechanics of embodied interaction that convey non-verbal or paralinguistic information.

Informed by these qualitative approaches software requirements are identified and proof-of-concept prototypes examined for how well technologies may be aligned with the endogenous organisation of work within a community of practice. Examples of detailed naturalistic studies of work and interaction will be presented in the poster from domains as diverse as clinical

medicine, stock market trading and palaeography.

Sergio Grau, “Automatic Speech-to-Phoneme alignment of the Spoken British National Corpus”

This poster describes the creation of an automatic word and phoneme alignment between the audio recordings of the Spoken British National Corpus (BNC) and their corresponding word-level transcriptions. The work presented here is part of the “Mining a Year of Speech” project which aim is to produce automatic speech-to-phoneme alignments of an approximately one year of audio recordings.

The Spoken BNC recordings consist of unscripted, spontaneous speech conversations in different recording conditions, accents and background noises. The range of topics covers from radio programs to family conversations, council meetings or chemistry courses. The Spoken BNC was originally recorded on analogue cassette tapes between 1991 and 1994. These tapes have been recently-digitised by the British Library. The resulting dataset is composed of approximately 2,000 digital audio files with an average duration of 45 minutes and their associated word-level transcriptions.

This poster describes the dataset, the automatic alignment process, the results obtained and the difficulties encountered.

Marinda Hagen, “The effect of singing on the Perception of Dutch-accentedness in English”

To test the hypothesis that Dutch speakers of English have a less strong foreign accent when they sing than when they speak, twenty native English listeners were asked to judge matched spoken and sung texts by eleven Dutch speakers of English and six native speakers of English, who served as a control group. The scores for the spoken and sung passages by the native English speakers differed little, while the scores for the sung versions by the Dutch speakers were substantially higher than those for the spoken versions.

To test the further hypothesis that the results are explained by the unavailability of duration and F0 for identifying foreign accent in singing, we asked the same twenty native speakers to evaluate speech in which original segment durations and F0 contours have been removed. Our findings reveal that speech in which original segment durations and F0 contours have been

made unavailable sounds less accented. They support our further hypothesis that the low degree of foreign-accentedness in sung texts is due to the unavailability of duration and F0 for identifying foreign accent in singing. A practical conclusion is that a considerable component in Dutch speakers' foreign-accentedness resides in non-segmental properties of their speech.

Lauren Hall-Lew, “Sound and Sociophonetics”

Sociophonetics is the study of language and society with respect to the phonetic qualities of speech. Sociophoneticians look at how speakers convey and construct aspects of their identity and other social information through variability in their pronunciation. The analysis of phonetic variables considers the acoustic properties of vowels, consonants, and suprasegmental features such as voice quality, pitch, and rhythm. Although traditionally restricted to the acoustic domain, researchers are increasingly interested in testing how listeners interpret social information about a speaker in the process of speech perception. This poster will give an overview of how acoustic data has been analysed in sociolinguistic research and how it has recently been used to create more rigorous test stimuli in perception experiments.

Zoe Handley, “Second Language Speech Learning”

My research focuses on perceptual approaches to second language speech learning, or pronunciation training. Two main approaches to perceptual training have been investigated in the literature: (1) perceptual fading in which learners are presented stimuli from opposite ends of a synthetic continuum in which the differences between the members of the non-native phonemic contrast have been exaggerated (Jamieson and Morosan, 1986), and (2) High-Variability Pronunciation Training (HVPT) in which learners are presented minimal pairs containing the non-native phonemic contrasts in forced-choice identification format (Logan et al., 1991). While these techniques, in particular HVPT, have been demonstrated to be effective in lab-based studies, they typically involve just one simple task, e.g. forced-choice identification, which the learner is required to repeat *ad nauseum*. To ensure that learners persist with training when a wider variety of phonemic contrasts are to be trained, I believe that a greater variety of tasks is required. In this poster, I present an experiment which compared two versions of HVPT, one employing an identification task and one employing a discrimination task.

Future research will explore a wider variety of tasks in combination with the use of synthetic continua.

Greg Kochanski, “Quasi-duration: Quantifying how rapidly sounds change.”

A certain amount of prosody and intonation can be correctly perceived even in unfamiliar languages or when speech is too distorted to be understood. Since durations of phones or syllables are often correlated with basic prosody properties like emphasis, this suggests that humans may have ways of perceiving something akin to syllable duration without a detailed segment-by-segment understanding of the sound. As it turns out, one can construct a simple algorithm with those properties: it measures how far (along the time axis) one can go before the acoustic spectrum changes substantially. This poster presents the quasi-duration algorithm and how it relates to durations of speech segments (phonemes).

Alexander Krasovitsky, “Sound change across generations: evidence from North Russia”

The focus of the study is ongoing sound change observed through variation across several generations of speakers. Until recently there has been a consensus that a large number of North Russian dialects systematically discriminate low and mid vowels in unstressed syllables.

Recent fieldwork in one of the North Russian rural communities has shown that the discrimination model may hold only for some of the older speakers whereas in the majority of idiolects, particularly within middle and younger generation, we find numerous instances of vowel neutralization. One example is the neutralization of low and mid vowels, /a/ and /o/, in unstressed syllables.

We find three types of the idiolects which present three successive stages of the loss phonological distinction between /o/ and /a/ in unstressed syllables. Type A consistently discriminates between low and mid phonemes in unstressed syllables (some of the older speakers). The opposite of this is Type C with /o/ and /a/ systematically neutralized in [ɐ] or [ə] (middle and younger generation). Multiple choices in type B, typical for the majority of the older speakers of the dialect, reveal a competition of the discrimination and neutralization pattern found in A and C accordingly.

Noel Loble, “Recording and remembering the sounds of Africa: ethnomusicology, sound archiving and sound elicitation”

Enchanted by recordings of African pygmy song, in 1986 Louis Sarno, an American carpenter, put down his tools and bought a one-way ticket to the Central African Republic. He still lives there today recording songs from the forest and has become an advocate for indigenous land and political rights. During the last three decades, he has built an unprecedented archive of recordings of music, which also includes images of performance and everyday life of a little known and endangered nomadic community. Until recently, this archive remained virtually unknown, lying wrapped in an old jumper inside a battered suitcase in a storeroom in Oxford. Using examples from Sarno’s archive and from my own doctoral fieldwork in the Eastern Cape of South Africa, during which I examined the potential contemporary relevance of the musicologist Hugh Tracey’s Sound of Africa series, a recording map of the musical memory of sub-Saharan Africa, I illustrate the value of ‘sound elicitation’. This method, the circulation of archival recordings through local social mechanisms, is an attempt to build ongoing relationships between sound recordings and indigenous communities, to enhance collections of field recordings, and to try and ensure sensible and reciprocal sharing of musical and cultural knowledge.

Anastassia Loukina and Greg Kochanski, “Measuring linguistic rhythm”

People often speak about ‘rhythm’ of a language and that different languages or dialects have different rhythm. This has led to substantial interest in quantitative measures that would capture this distinction. Traditionally, most of the rhythm metrics relied on duration. However, experimental evidence suggests that the perception of rhythm rests on a combination of different acoustic properties that are not limited to duration. These include changes in F0, rate of spectral change and auditory prominence. In this poster I will present an overview of different approaches to measuring linguistic rhythm and report experimental results that evaluate reliability of some of these methods.

Kevin R. Page. “Linked Data, e-Research, and Music Information Retrieval”

The Linked Data movement encourages a Semantic Web built upon HTTP URIs that are published, linked, and retrieved using RDF and SPARQL, while the Linked Data cloud already contains several music related datasets that hold great potential for enhancing the process of research in the field of Music Information Retrieval (MIR) and which, in turn, can be enriched by MIR results.

We present a system that demonstrates the utility of Linked Data for enhancing the application of MIR workflows through several related aims: to enable MIR researchers to utilise these datasets by incorporation in their research systems and workflows; to publish MIR research output on the Semantic Web linked to existing datasets; and to present MIR research output, with cross-referencing to other linked data sources, for manipulation and evaluation by researchers and re-use within the wider Semantic Web.

By way of example we gather and publish metadata describing signal collections derived from the country of an artist; genre analysis over these collections and integration of collection and result metadata enables us to ask: “how country is my country?”. While the demonstrator embodies a specific analysis (genre) and collection selection (by nation) to a simple use case, the approach and technologies are more generic and widely applicable – the common data model of RDF extends a myriad of possibilities for linking with data models and categorisations within and without the MIR community.

Zeynab Raeesy, “Acoustic-to-Articulatory Inversion based on Dynamic MRI”

A speech wave is the result of pushing the air through human vocal tract, where distinct vocal tract shapes lead to production of distinct sounds. The aim of acoustic-to-articulatory inversion research is to estimate the human articulatory configuration when producing different sound waves. In other words, acoustic-to-articulatory inversion is the problem of finding a function that maps the acoustic to the articulatory domain.

In this research, I focus on developing machine-based approaches for modelling the articulation according to speech. I use data-oriented learning techniques for finding the correlation between different sound units and vocal tract shape. The model is provided with the acoustic and the corresponding articulatory data, and is trained to find the mapping between them. The

trained model can then be used to derive the articulatory parameters for unobserved acoustic data. This research is developed using a dynamic Magnetic Resonance Imaging database, which consists of midsagittal MRI images of human vocal tract (head) while speaking, and the corresponding acoustic data recorded simultaneously during articulation.

In this poster, I present an overview of the current status of this project and a brief insight to the further stages.

Jonathan Roberts, “Sounding Authentic: Helping English People Sing Javanese”

Sounding convincing when performing a vocal art from another culture requires attention to sound on many levels. Pronunciation and voice quality are as important as intonation, phrasing and idiomatic rhythm. The mechanics of how we use our voices can be so ingrained and habitual that changing and maintaining control of all these variables at one time can be difficult.

Sung poetry is a core element in Javanese musical expression and essential to the performance of gamelan. Vocal training in Java, however, focuses entirely on content rather than technique, leaving the foreign student attempting to gradually adjust their sound by trial and error. In order to recreate a full gamelan sound with groups in the UK, one is faced with the challenges of helping non-singers attracted by the idea of a percussion orchestra to sing in a foreign language and singers with existing technique to change how they use their voice. This poster will address some of the issues that arise in learning and teaching Javanese vocal music cross-culturally and some of the techniques that can be used to help English people make a Javanese sound.

Duncan Williams, “Evaluating timbre in location recordings”

Recording engineers are often charged with making recordings ‘in-field’ that require artistic and creative judgements to be made. Timbral fidelity, unlike perceived loudness or frequency response, is the least well understood and difficult to control of perceptual characteristics. The long-term goal of this work is to provide a statistically valid ‘rule of thumb’ for the recording engineer wishing to achieve specific timbral characteristics in their location recordings, by evaluating a range of recordings with acoustical and perceptual measures.

This poster presents the current methodology - recording with multiple microphones, followed by acoustical analysis, listener evaluation and statistical analysis. The accompanying presentation will demonstrate example sound files from the stimulus set for the audience to experience.