Assessing Individual Talent in Second Language Production and Perception

Matthias Jilka[1], Volha Anufryk[2], Henrike Baumotte[2], Natalie Lewandowski[2], Giuseppina Rota[2,3], Susanne Reiterer[4,5]

[1]Institute of English Linguistics, University of Stuttgart
[2]Institute of Natural Language Processing, University of Stuttgart
[3]Institute for Medical Psychology and Behavioural Neurobiology, University of Tübingen
[4]Section of Experimental MR, University Clinic Tübingen
[5]Hertie Institute for Clinical Brain Research, Tübingen
Germany
jilka@ifla.uni-stuttgart.de, natalie.lewandowski@ims.uni-stuttgart.de

1. Introduction

In this paper we introduce a large-scale project that aims to provide a comprehensive examination of talent in pronunciation, especially of a second language. The project investigates this ability with respect to its multiple phonetic/linguistic manifestations, but also considering psychological and behavioral influences on pronunciation performance and, most importantly, with the objective of finding neural correlates of pronunciation talent, i.e., differences in brain activity between talented and untalented speakers.

The present study focuses mainly on the first part of this undertaking, describing the extensive tests necessary to measure phonetic talent in its various dimensions such as production and perception, the segmental and suprasegmental levels of speech or different utterance forms such as spontaneous speech, reading and imitation.

The tests are intended to provide general insights into the nature of pronunciation talent as well as the interactions between the examined talent-related parameters. They are also meant to serve the purpose of allowing a reliable classification of talent level to be used in the selection of subjects for the neuroimaging studies in the second part of the project.

1.1 Individual differences in second language acquisition ability

It is a common observation that individuals differ greatly from each other in their ability and velocity in acquiring a non-native, or "second" language (L2). This is especially true for the acquisition of the L2 sound system (including all segmental and prosodic manifestations on both the phonetic and the phonological level). The possible factors responsible for these differences are numerous, and a large number of them has been shown to significantly influence the learners' performances.

The mainly phonetics-oriented studies of second language acquisition (SLA) concentrate on external factors concerning circumstances of language acquisition and use, such as age of learning (e.g., Johnson & Newport, 1989), length of residence / age of arrival (e.g., Flege et al., 1995) or amount of L1 / L2 use (e.g., Piske et al., 2001). Characteristics inherent to the learner are the focus in cognitive/psychological investigations, which examine, e.g., working memory (e.g., Papagno & Vallar, 1995), motivation (e.g., Moyer, 1999) or personality factors such as empathy and extraversion (e.g., Edmondson & House, 1993). This is also the case for neurolinguistic approaches that assume physiological factors like special genetic equipment, e.g., regarding brain anatomy or a particularly evolved neural organization, i.e., more or less effective neural processing of language-related tasks. It is especially this latter aspect - with respect to phonetic/phonological abilities - that constitutes a major final object of study for the project presented here.

1.2 Talent

The stated objective of this project naturally requires that the notions of "talent" in general as well as language and pronunciation talent in particular be addressed.

With respect to a definition of what constitutes the core of talent it is generally agreed that a distinction has to be drawn between proficiency, i.e., the overtly observable performance of a particular skill, and talent per se. Factors such as motivation, practice and experience contribute to the degree of the proficiency but are not part of the talent. The consequences of this view correspond to popular concepts that consider talent to be an innate, somewhat mysterious ability that a person either has or does not have.

The idea that this ability is somehow reflected in the brain seems to be a logical extension of this line of reasoning, but is not as widespread. While there are some neurologically-oriented studies of talent, or rather genius, as typified by such efforts as the examination of sections of Einstein's brain (Diamond et al., 1985), the study of this field is not broadly established. Investigations of the neurological substrates of language talent have for example been undertaken by Geschwind and Galaburda (1985) in their description of pathological language talent as being related to the increased growth of particular brain areas. Influential neuropsychological models of the source and structure of talent are based on a model of distinct faculties where special abilities are adjacent to each other and linguistic talent is comparable to musical, logical, spatial talent etc. (Gardner, 1983).

In the majority of cases the measurement of language talent simply consists of tests of general ability. However, in individual cases of exceptional, sometimes pathological, language talents (i.e., extremely fast and successful L2 learners) extensive tests have been applied in the attempt to diagnose the exact nature of the skill. Novoa et al. (1988), for example, used a test battery that examined a speaker's abilities with respect to IQ, vocabulary skill, general language aptitude, verbal fluency, verbal memory, apprehension of abstract patterns, understanding (digit symbol test), and learning of code systems.

Just as language talent is generally considered one of many discrete talents within the model of distinct faculties, it is also generally assumed that language talent consists of different independent linguistic skills. A widely acknowledged and fundamental distinction has been drawn between two substrates of linguistic ability, namely talent for grammar vs. talent for accent (Schneiderman & Desmarais, 1988). The separate position of pronunciation skills is widely accepted. In fact, this generally assumed special difficulty of pronunciation acquisition as opposed to that of other aspects of grammar is virtually proverbial, as is made clear by the commonly used term "Joseph Conrad Phenomenon" (e.g., Bongaerts et al., 1995; Bongaerts, 1999; Guiora, 1990), which refers to the Polish-born novelist's native-like abilities in English grammar, vocabulary and style being opposed to his strongly accented pronunciation.

1.3 Method

1.3.1 Main objectives

The primary purpose of the testing procedures is the assessment of the speakers' phonetic abilities in different areas (e.g., production vs. perception, segmental vs. suprasegmental characteristics). As these assessments will be correlated with the results of the psychological tests and general information about the speakers' habits and history concerning language use, it should be possible to distinguish between the speakers' superficial proficiency and their inherent abilities. The phonetic tasks contribute to this objective by restricting the influence of such factors as practice and experience. The results of the assessment lead to the classification

of distinct groups of subjects (talented, normal, untalented) to be selected for neurolinguistic examination, i.e., repetition of modified versions of the phonetic tasks during an fMRI scan.

1.3.2 Speakers and languages

One hundred and five native speakers of German served as test subjects. Among them a core control group of 50 university students of English shared an additional number of variables such as age (range: 20-23 years), age of onset of L2 English learning (10 years) and experience/training in this L2 (same type of English instruction in the formal setting of German schools, comparable amounts of experience in English-speaking environments). The remaining test subjects were of all ages and backgrounds, found via ads or personal contacts. They were chosen due to their self-professed pronunciation talent or lack thereof. 15 native speakers of English also participated in order to provide comparable data. The languages used are English, German (for prosody tasks) and Hindi (as an unfamiliar language). English was chosen as the large number of learners increases the likelihood of finding talented individuals with native-like pronunciation skills. Furthermore, comparative linguistic descriptions, both for the segmental and prosodic characteristics are easily available. The testing of prosodic phenomena in particular is facilitated, because detailed intonation descriptions and resynthesis tools for both German (e.g., Mayer, 1997) and English (e.g., Jilka et al., 1999) exist.

2. Test procedures

2.1 Personal data

An extensive questionnaire deals with the relevant biographical data and, following examples from Moyer (1999) or Markham (1997) investigates language-related attitudes and experiences that are assumed to influence language ability. Specific questions thus address issues such as general motivation, self-assessment of one's own abilities, preferred learning styles, the extent of instruction, the extent of L2 experience, use and input, the extent of L1 use and input and the attitude towards foreign languages, especially English.

2.2 Psychological testing

2.2.1 Personality questionnaires

According to Trait Theory, personality reflects a human being's individual consistency in behavior across time and situations (Allport, 1961). We applied several questionnaires that aim to determine whether certain personality characteristics correlate with linguistic abilities.
The Neuroticism Extraversion Openness Personality Inventory assesses personality traits according to five major factors, namely neuroticism, openness, extraversion, agreeableness, conscientiousness (Costa & McGrae, 1992). The BIS/BAS questionnaire (Carver & White, 1994) on the other hand investigates whether a person's actions are driven more by the positive motivation toward something desired (Behavioral Approach System – BAS) or the desire to avoid a negative experience (Behavioral Inhibition Sytem – BIS). A questionnaire assessing empathy, E-Skala (Leibetseder et al., 2001), is administered to examine if this characteristic is associated with a greater readiness to adapt to unfamiliar phonetic features.

2.2.2 Cognitive tests

While certain cognitive abilities such as working memory or musicality are often assumed to correlate with linguistic performance, this is not the case for others, e.g., intelligence. We therefore conducted tests assessing some of these abilities.

The Raven Advanced Progressive Matrices Test (Raven et al., 1998), for example, addresses the capacity of non-verbal logical reasoning and recognition. A different type of intelligence assessment, namely verbal intelligence, is performed by the Mehrfachwahl-Wortschatz-Intelligenztest (MWT), which focuses on extended vocabulary knowledge in the native language (Lehrl, 2005). The latter test would be expected to show a stronger connection to linguistic, though not necessarily phonetic performance.

We are also administering standard tests of phonological working memory (Gathercole et al., 1994), as it has repeatedly been argued that "phonological short term memory" can predict success in foreign as well as native language learning. Many studies have reported such correlations, but their exact nature is still unclear. Research with children has demonstrated that the ability to learn new words is greatly affected by working memory span, a correlation that appears to continue into adolescence (Gathercole et al., 1994).

In addition, a task evaluating musicality by asking the listener to recognize tonal and rhythmic changes in pairs of sound files (Gordon, 1989) is also included, as is a test measuring mental flexibility (Simon, 1990), defined as the ability to handle different situations in different ways and thus easily adapt to changes, learn from mistakes and be creative in solving problems.

2.3 Language aptitude and proficiency

There are of course established tests for assessing general linguistic aptitude and a large number of tests for simply evaluating proficiency in a specific language. Results from such test batteries serve as useful comparisons to tasks focusing on phonetic abilities exclusively.

2.3.1 General language aptitude

The Modern Language Aptitude Test, short MLAT (Carroll & Sapon, 1959), was conceived for native speakers of English in order to predict success/talent at learning a second language.

We have applied a shortened version of the MLAT to our native speakers of German. The results can therefore not be analyzed according to the given interpretation scores, but nevertheless allow a comparison of the speakers amongst each other.

Of the three tasks we used, "Spelling Cues" has a phonological component in that the use of an alternative spelling system is meant to test the ability to associate sounds and symbols (and vocabulary knowledge in non-native speakers of English). The example shown below demonstrates that an alternatively spelled word ("nme") must be associated with one of four possible synonyms or otherwise related terms:

nme *A. sea-animal* *B. decoration* *C. foe* *D. druid* *E. numbness*

We also used a task that tests comprehension of grammatical structure by requiring a comparison of the syntactic functions of elements in different sentences ("Words in Sentences"), as well as a test of verbal memory ("Paired Associates"), in which test participants must memorize 24 Kurdish vocabulary items within 2 minutes and then pick the correct item out of five choices offered.

2.3.2 English proficiency

We also applied a small portion of the Test of English as a Foreign Language (TOEFL) in order to get a clearer impression of our German tests subjects' overall English proficiency.

The tasks are limited to a listening comprehension test (participants listen to a lecture of approximately two minutes in duration and then answer questions about the content of what they heard in a multiple choice scenario) and a so-called "structure" test in which both grammaticality judgments and vocabulary knowledge are required.

2.4 Testing phonetic ability

Contrary to the questionnaires and tests described in Sections 2.1. to 2.3., which were conducted via the internet, the tasks intended to assess the test participants' phonetic abilities were performed in an approximately 90-minute session in the anechoic chamber of the Institute of Natural Language Processing at the University of Stuttgart (45 minutes of production tasks + 45 minutes of perception tasks).

2.4.1 Speech production

Following the example of studies like Bongaerts et al. (1995), Markham (1997) or Flege and Hillenbrand (1987) we have used several different elicitation techniques to see whether degree of accent varies as a function of type of elicitation. We applied the full range of techniques in order to both get a complete picture of speakers' abilities before brain imaging and to elicit a variety of types of intonational configurations, differing speaking rates and degrees of fluency.

Tasks for **segmental production** can be included in all controlled elicitation techniques, while for the produced spontaneous speech it is of course only possible to analyze what the speakers happen to say. In both the reading and imitation tasks, all phonemes, important allophonic variations and phonotactic constellations of the L2, i.e., English are covered. The known problematic areas for German speakers reach from cases of heavy (e.g., production of uvular /ʀ/, lack of dental fricatives, clear /l/ in all positions) to medium (final devoicing of voiced obstruents, insertion of glottal stops before morphemes that start in a vowel, raising of /æ/ to /ɛ/) to light foreign accent (e.g., subtle differences in vowel quality).

The tasks for **prosody production** use especially imitation and reading tasks to elicit tunes (i.e., combinations of pitch accents and boundary constellations) associated with particular discourse situations (e.g., declaratives, Yes/No-questions, continuation rises). Insights gained in previous work (Jilka, 2000) facilitate the identification of areas of possible deviations due to foreign accent (e.g., category transfer in the realization of continuation rises, subtle variation in the phonetic realization of equivalent categories).

Quasi-spontaneous speech

Spontaneous speech is by definition the most natural form of speech and certainly reflects overall abilities the best, allowing especially representative impressions of fluency, speaking rate, choice of words, choice of prosodic patterns and segmental realizations. On the other hand it is very difficult to elicit specific prosodic or segmental realizations of phenomena assumed to be particularly interesting. We have attempted to control spontaneous output to a certain degree by having the subjects narrate a short Gary Larson cartoon (Larson, 1992), see also Figure 1, which to a small degree suggests vocabulary choices, and thus segmental realizations and also triggers certain prosodic constellations like continuation rises.

At the start of each recording session, speakers were also asked to introduce themselves and talk about their experiences with the English language/culture. This monologue is then extended into a short conversation with the instructor, resulting in about 4 minutes of quasi-spontaneous speech altogether. Due to the laboratory situation, the speech production can of course not be called completely spontaneous, therefore we term these tasks quasi-spontaneous.

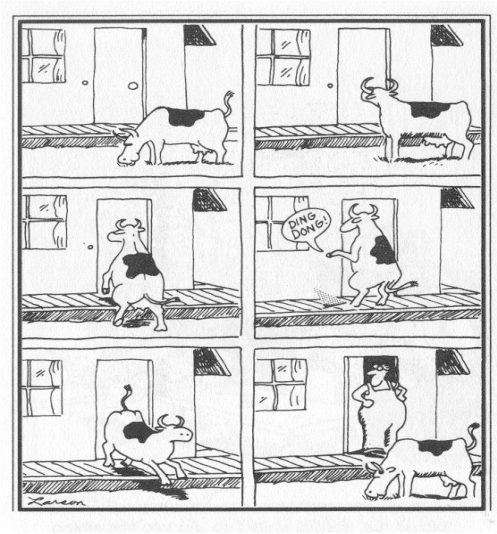

Figure 1. Cartoon to be narrated in English by the native German test subjects (words like "cow", "doorbell", "farmer", "grass" etc. are likely to occur)

Read speech

Reading tasks have the advantage of allowing a controlled coverage of the phoneme inventory and specific phonotactic constellations as far as segments are concerned, as well as a reasonably controlled elicitation of pitch accent distribution and tunes associated with specific discourse situations. They also make use of avoidance strategies with respect to problematic sounds, words or sentence structures much more difficult, thus reducing the potential influence of syntactic/morphological errors on evaluation. It has to be taken into account, however, that the produced speech is restricted to the prosodic patterns specific to this style of speech and may also elicit a more careful segmental articulation than "regular" speech.

For the English reading tasks we used the standard IPA text for segmental coverage "The Northwind and the Sun" and a page-long extract from the Mark Twain short story "Mrs. McWilliams and the Lightning" (Twain, 1968) which contains more difficult words such as, for example, "unlocatable" or "deliberately" as well as various dialog situations eliciting questions, exclamations etc. We also included individual sentences in order to enforce the production of specific segmental and/or prosodic constellations (e.g., reading the short phrase "She's teething" in such a way that it reflects different emotional states like surprise, happiness, pity or anger – the voiced dental fricative also proved to be challenging).

Equivalent sentence-reading tasks were also applied in German, focusing on prosodic choices and realizations, e.g., reading the sentence "Alle Politiker sind nicht korrupt" in such a way that it means either that all politicians without exception are not corrupt or that many politicians are indeed corrupt, but not all of them (the second interpretation requires the use of a hat pattern). Additionally, the speakers were asked to read "Der Nordwind und die Sonne" in German, imitating an English accent. This task follows a procedure introduced in Flege and

Hammond (1982) and constitutes a form of delayed mimicry that tests the speakers' awareness of the phonetic features of other languages and their ability to realize them.

2.4.2 Speech perception

Phonetic ability may also manifest itself in how semantic and pragmatic information in speech is perceived (and possibly replicated). Our speech perception tasks concentrated mainly on the detection, comprehension and interpretation of suprasegmental features that control overall phrase intonation, i.e., particularly pitch, but also duration and loudness.

We describe intonation using the ToBI system (Silverman et al., 1992). Intonational differences can thus be categorical (different tone category, e.g., pitch accent) or realizational (same category, different phonetic realization). The existence of detailed intonation "grammars" for English (Jilka et al., 1999) and German (Mayer, 1997) as well as a resynthesis tool (Möhler, 1998) allows for the deliberate production of phonological and phonetic modifications of the original models without significant distortion effects.

Pair comparisons

Our pair comparison tasks test the listeners' ability to recognize prosodic differences between otherwise identical utterances. The differences were created artificially by using resynthesis to modify one of the recordings in such a way that an intonational category (i.e., pitch accent or boundary tone) is either affected in its completeness, e.g., added, deleted or replaced by another one (see Figure 2) or manifests a different phonetic realization, e.g., different timing or pitch height (see Figure 3). It is expected that categorical differences are perceived more easily than realizational ones such that talented listeners would be more sensitive to the latter than less talented listeners.

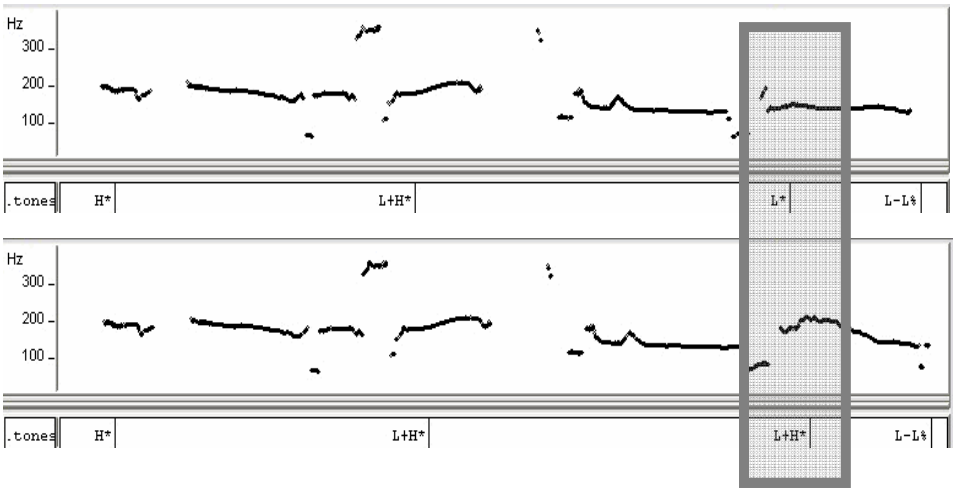

Figure 2. Example for a pair comparison with a categorical difference between the two versions of the utterance "People will get used to the idea". The top version has a low (L*) pitch accent on "idEa", giving it emphasis but no particular F_0 movement. The bottom version shows an additional rise-fall movement associated with a L+H* pitch accent

The pair comparisons were carried out in German and English to determine whether listeners would perform better in their native language (native speakers of English also participated as control subjects). Furthermore, they were also conducted using low-pass filtered speech, which is devoid of segmental information, thus leaving the listener only with

prosodic information from the F_0 contour that may or may not be recognized as language-specific (additional language identification tests in low-pass filtered speech were carried out as well).

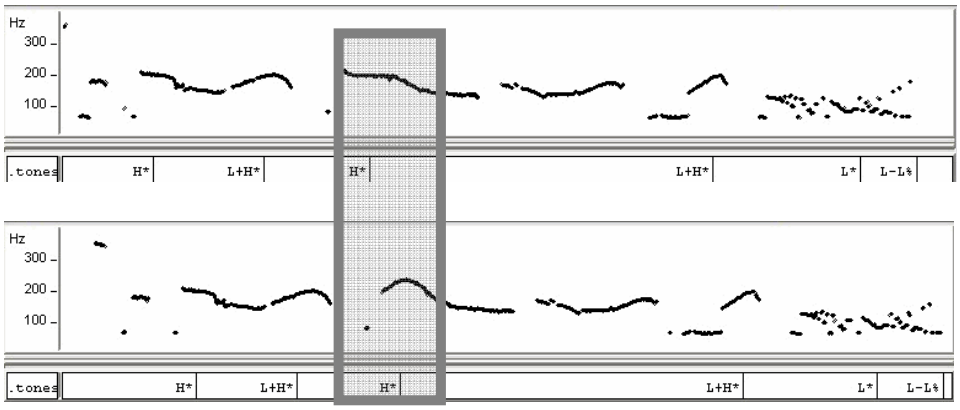

Figure 3. Example for a pair comparison with a realizational difference between the two versions of the utterance "The State is not planning on putting more police on the road". Both versions have a high (H*) pitch on "PLAnning", but in the bottom version it occurs somewhat later and higher, causing a slightly more pronounced up-and down movement of the contour

Prosody interpretation

Exercises on prosody interpretation are intended to test experience and awareness in the ability to draw linguistic information from the prosodic structure of an utterance without help from the general context. The listeners hear an utterance and are then asked to decide which one of four suggested interpretations matches it best. In an alternative scenario two prosodically different versions of an utterance are presented, and the listener has to answer a question about their interpretation, e.g., "In which version does the speaker sound angrier?". As in the pair comparison, results in the German version of this task might be expected to be better than those in the non-native language English, with differences among speakers in German indicating differences in general language awareness.

Accent identification

This is the only speech perception task not dealing exclusively with prosodic features. The test participants listen to various versions of "Der Nordwind und die Sonne" as read by non-native speakers of German and are asked to identify the speakers' respective native languages. Four possible answers are offered in a multiple choice scenario. The task assesses general awareness of and experience with characteristic segmental and prosodic features of languages.

2.4.3 Speech perception and production combined

Imitation tasks combine the abilities addressed in the two preceding subsections: the correct perception of a model and the ability to reproduce a correct representation of it. They allow the maximum amount of direct control over what a speaker will say and thus facilitate the direct testing of complex tonal constellations and difficult segments as well as of subtle phonetic variation on both the segmental and suprasegmental level. Besides German and English we also added Hindi as a language that the test participants were not familiar with.

The Hindi imitations of words and short phrases focus exclusively on the segmental aspects of speech, especially the perception of those sounds and phonotactic structures not present in the native language and the ability to reproduce the perceived acoustic patterns.

The German and English imitations emphasize prosody to a much larger extent. They require the speakers to perceive, understand and mimic appropriate prosodic patterns, both in categorical and realizational dimensions. Prosodic constellations of interest (e.g., focus, particular F_0 movements, certain types of discourse situations) were created directly by the model speaker or again using resynthesis. Prosody imitation is the main challenge in the German tasks, even though many of our native German speakers (those who are actually speakers of the Swabian dialect) also did not find it easy to reproduce the standard High German pronunciation of the model. The English imitation tasks are also designed to test the reproduction of prosodic phenomena, but here too the imitation of either the standard British English (Received Pronunciation) or General American model proved to be difficult. Additionally, general English comprehension problems may influence the performance.

Direct imitation

The approach of directly imitating a given model is often criticized as not being a reflection of linguistic skill at all since it produces behavior temporarily exceeding actual competence and thus has no "carry-over into the post-imitative tasks" (Barry, 1989:167). Others, however, are more supportive of the approach. Markham (1997), for example, argues for its use because acquisition itself is a strongly imitative phenomenon and Kuhl and Meltzoff (1995) show that imitation can result in nonidentical, but functionally equivalent reproductions of the modelled behavior, a fact that would speak against the view that direct imitation bypasses all levels of linguistic processing. Thus, a number of studies (e.g., Markham, 1997; Neufeld, 1987) use direct imitation in their language tests.

Stimulus presentation in this part of the test was thus quite straightforward in that the speakers simply imitated the utterance they were presented with. In the Hindi test, however, the stimuli were presented three times in a row before imitation, such that the speakers had the chance to familiarize themselves a little more with the unknown sounds they encountered.

Delayed imitation

Delayed imitation tasks strive to preserve the advantage of all imitations, namely the high degree of control and precision, while avoiding the apparent disadvantage of allowing direct imitations from sensory memory that bypass actual linguistic abilities. For this reason some consider delayed imitation to be the best possible elicitation technique (e.g., Piske et al., 2001; Flege, 1995; Flege et al., 1999). Such tasks typically take the form of a presentation of a question, an answer and then the question repeated, after which the listener/speaker is required to reproduce, i.e., imitate, the given answer. The procedure thus has somewhat greater demands on memory capacity than direct imitation.

Figures 4 and 5 show examples of utterances elicited by delayed imitation.

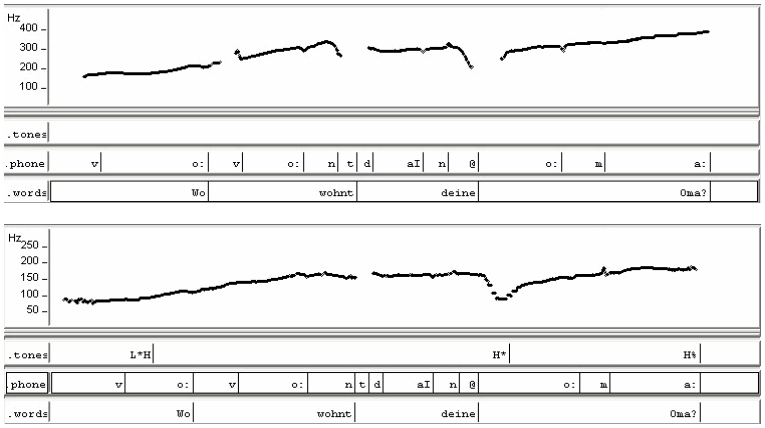

Figure 4. Successful delayed imitation of the incredulous question "Wo wohnt deine Oma?" (Where does your grandma live?) – top contour: original, bottom contour: imitation

Figure 4 depicts an imitation where the speaker succeeded (both acoustically and visually) in reproducing the incredulous question "Wo wohnt deine Oma?" (Where does your grandma live?), which manifests the unusual contour shape of a steady rise from the start to the end. The intonation shape is triggered by the preceding context "Meine Oma muss sich immer vor den Gorillas in Acht nehmen" (My grandma must always beware of the gorillas).

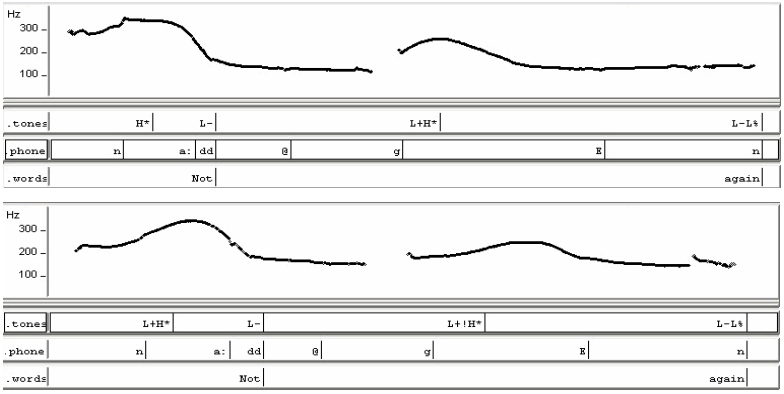

Figure 5. Failed delayed imitation of the exclamation "Not again!" top contour: original, bottom contour: imitation

Figure 5 on the other hand shows a failed delayed imitation of the short exclamation "Not again" with a General American accent both in terms of intonation and segmental pronunciation (low back vowel /*a*/ in "not"). The speaker manages to reproduce the vowel but has problems with the intonation contour. Especially the second F_0 peak (on "aGAIN") occurs too late and is not high enough, but also the anticipation of the first peak (on "NOT") is not reflected in the German speaker's version. For this reason she produces a much steeper rise toward it (accordingly labeled with L+H* instead of just H* as in the original).

3. Evaluation methods

3.1 Overview

The abundance of data created by such an extensive experiment requires a large effort in evaluating and analyzing the many different types of results and relating them to each other.

Several established psychological tests (intelligence, empathy, mental flexibility) provide automatic scoring and interpretation. This is also the case for MLAT and TOEFL, however only parts of those tests were used, such that the individual scores of the respective test parts have to be analyzed. The perception tests were also designed in such a way that the listeners provided clearly defined results, either deciding whether a difference was perceived or not (pair comparison) or selecting an answer in a multiple choice scenario (interpretation, accent identification). Performance in the perception tasks can therefore be assessed in a relatively straightforward fashion, even though the significance of specifically targeted phonetic phenomena (i.e., subtle realization differences, particular constellations like hat patterns, recognition of discourse elements like incredulous questions, sarcasm etc.) must be investigated in greater detail. This aspect is even more important in the analysis of the production tasks, where the realization of the targeted segmental and prosodic phenomena must undergo thorough expert analysis. The test subjects' overall English proficiency will also be assessed by native speakers of English in a large web-based evaluation (see also Section 3.3.). A smaller evaluation of the performance in the Hindi imitation tasks will also be carried out.

3.2 Talent vs. proficiency

Distinguishing superficial proficiency and inherent talent is of course far from being a trivial challenge, as talent is just one of several factors contributing to the speaker's actually perceivable performance.

It is possible to use the information gathered in the questionnaires addressing biographical data and personality factors to better assess the influence of experience and practice as well as the degree of motivation. Drawing connections between this information and actual performance would not be a straightforward process, but would at least allow the consideration of differences, for example between students of English and test subjects with little or no academic experience, or speakers with the long-term experience of living in an English-speaking country and others who only spent a two-week vacation in, say, Scotland.

More importantly however, some tasks were designed in such a way that the contribution of experience and practice is neutralized. In the tasks involving German (interpretative reading, pair comparison, prosodic interpretation, imitation) the test participants, as native speakers of German, should have comparable amounts of experience, whereas in the Hindi task, none of them had any experience at all. As none of the participants were professional linguists, major differences in theoretical knowledge especially about phonetics are also extremely improbable. In the English tasks, an effect of practice and experience can of course not be excluded, even though it may be smaller in tasks involving prosody only, as learners are less likely to have acquired any conscious knowledge about this aspect of speech.

3.3 Perceptual evaluation

A perceptual evaluation is necessary in order to obtain a reliable score of the test subjects' abilities in relation to each other, but also in absolute terms.

A web-based evaluation scheme has been devised in order to reach a large number of native English-speaking raters (of several accents of English). It focuses on the assessment of two production tasks, the narration of the Gary Larson cartoon and the reading of "The North Wind and the Sun". 120 speakers (105 German test subjects + 15 native speakers of English with various accents) must be rated. We intend to use 200 raters overall, who are divided into four groups rating 60 speakers each with an overlap of 30 speakers per group such that each speaker will be rated by 100 raters. The presentation order will be random, the two tasks separated. We follow the recommendation of Piske et al. (2001) in using a representative sample of raters, though there may not be a significant difference between expert and naïve raters (e.g., Bongaerts, 1999). Southwood and Flege (1999) recommend a 9- or 11-point scale to best exploit listeners' full range of sensitivity. The generally accepted criterion for determining near-nativeness is to be within two standard deviations of the mean rating for native speakers (e.g., Flege, 1995; Bongaerts et al., 1995; Piske et al., 2001).

The stimuli are rated on a visual analogue scale (Wewers & Lowe, 1990), which consists in a straight line with only the two extreme boundaries labeled. The scale used in our study, as shown in Figure 6, is unipolar, denoting the absence of a phenomenon, i.e., foreign accent at one end and its maximum intensity at the other end. Any click on the scale in between these extremes is assigned to the corresponding numerical value. There are no further marks on the scale, which may allow the rater less possibilities of orientation but is claimed to provide a more uniform distribution of scores along the scale's entire length (Scott & Huskisson, 1976). Visual analogue scales are easily understood, quickly scored and can reflect fine discriminations.

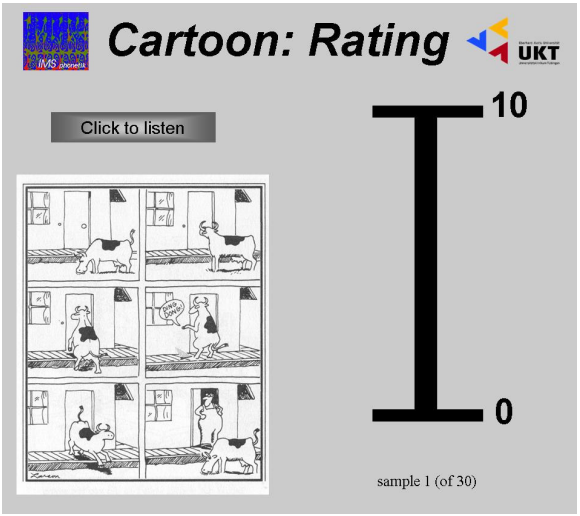

Figure 6. Example of the web-based evaluation (here: rating of performance in narration of cartoon)

4. Preliminary results

Due to its size, this project has necessitated a prolonged period of preparation, starting with the design of the various tasks, the recruitment of speakers, and the actual recordings. While an extensive quantity of data exists, it has not yet been analyzed in great detail. Similarly, as stated in Section 3.3., the perceptual evaluation has not yet been carried out. For working purposes such as a first overall examination of the general relations between the various tasks as well as the conduction of neurolinguistic pilot studies (see also Section 4.2.), we used preliminary expert scores based on sample surveys of about 15 minutes each of the

perfomance in all tasks. Initial impressions of several tasks seem to be quite distinct: e.g., the reading task that requires test participants to "fake" an English accent seems to be a good indicator of untalented performers (though not necessarily of talented ones), on the other hand, good imitations skills in the Hindi imitation and some of the more difficult prosody imitation tasks in German and English seem to correlate with general talent and high performance.

4.1 Correlations of automatic scores

Results for the tasks involving automatic scoring (i.e., psychological and perceptual testing) are readily available for statistical analysis, albeit without allowing in-depth looks at the actual characteristics and objects of the individual tasks.

It can be briefly summarized that the preliminary expert talent/performance scores in the production tasks correlate best, i.e., significantly, with the results from the pair comparisons in English, the accent identification task, the English interpretation task, the TOEFL listening comprehension and, maybe not surprisingly, the MLAT scores. As far as the psychological tests are concerned we have found correlations with verbal intelligence, musicality, and the personality factor openness.

Considering correlations among the automatically scored tasks, first results show relationships between better performances in the interpretation tasks and the personality factor empathy. Performance in the German interpretation tasks correlates especially strongly with the accent identification, possibly indicating a heightened awareness of more subtle characteristics of the native language. Performances in all the pair comparisons (German, English, low-pass filtered) correlate with each other, which may mean that linguistic knowledge/awareness actually does not play a role as opposed to purely perceptual abilities. This view may be supported by the fact that better performance in the low-pass filter-tasks correlates with the personality factors of less inhibition and more conscientiousness, which could be interpreted as expressions of greater effort and motivation.

It has to be emphasized, however, that these are also only preliminary results that will be subject to further scrutiny.

4.2 Pilot study of possible neuronal correlates of talent

In order to learn more about the technical feasibility of the planned neuroimaging study and to get a more concrete idea of what shape eventual differences between talented and untalented speakers might take with respect to processing in the brain, a pilot study was conducted. Three native speakers of German matched for age, level of education, and onset of L2 acquisition participated. One speaker had been classified as "talented", the other two as "untalented". Due to the variability of human brain anatomy it is of course obvious that no stable conclusions can be drawn from an examination of only three different individuals.

The test participants were presented with 12 pairs of German, English and Hindi words matched for number of syllables (N = 3) and length (2s) which they had to imitate while being scanned by means of functional magnetic resonance imaging (fMRI). Such a task had not been part of the regular phonetic test. For the actual scan considerably longer sessions based on the known task types (perception, reading, imitation) are used.

Echo Planar Images (EPIs) (sparse sampling design; 72 volumes, TE = 48 ms, TR = 10000 ms, TA = 3s, voxel size 3 x 3 x 3 mm, FoV = 192 x 192 mm) and a T1-weighted anatomical image (176 volumes, TE = 3.19 ms, TR = 1300, thickness of 1 mm, FoV = 256 x 256 mm) were acquired on a 1.5 Tesla whole body scanner (Siemens Vision, Erlangen, Germany). Pre-processing and statistical analyses were carried out by means of SPM5.

Results showed that the pronunciation of German, English, and Hindi materials elicited activation of a bilateral temporal network mainly restricted to the superior temporal

gyri and the motor areas. First level statistical analyses indicated that the extent of activation necessary to sustain the task was significantly greater for untalented speakers (see also Figure 7). These preliminary results thus suggest that high proficiency correlates with reduced effort in speech production, and enhanced cortical efficiency.

It remains to be seen whether this result will be confirmed and what other more specific insights will be gained during the further course of the experiment.

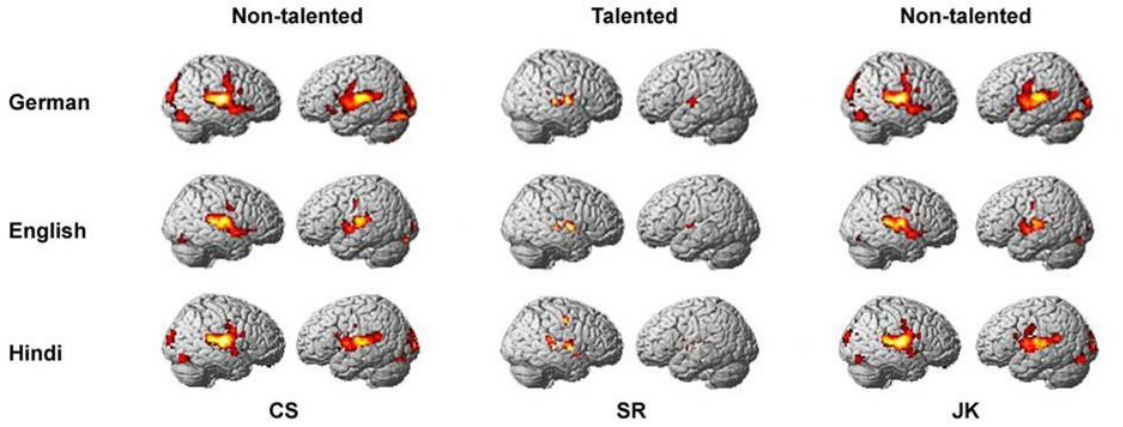

Figure 7. Brain activation in the three speakers CS, SR and JK during word imitation tasks of German, English and Hindi. The talented speaker SR generally shows a much smaller extent of activation (first level fixed effects analysis, $p < 0.001$, FWE corrected).

Acknowledgement

The research project "Language Talent and Brain Activity – The Neural Basis of Pronunciation Talent" is supported by the German Research Council (DFG): Grants DO 536/6-1, AC 55/7-1

References

Allport, G. W. (1961). *Pattern and growth in personality.* New York: Holt, Rinehart, and Winston, Inc.

Barry W. (1989). Perception and production of English vowels by German learners: Instrumental-phonetic support. *Phonetica, 46*, 155-168.

Bongaerts T., Planken B., & Schils E. (1995). Can late starters attain a native accent in foreign language? In D. Singleton, & Z. Lengyel (Eds.), *The age factor in second language acquisition* (pp. 30-50). Clevedon: Multilingual Matters.

Bongaerts T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 133-160). Mahwah: Lawrence Erbaum Assoc.

Carroll J., & Sapon S. (1959). *Modern Language Aptitude Test (MLAT): Manual*. New York: The Psychological Corporation.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology, 67*, 319-333.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO personality inventory (NEO PI-R) and NEO five factor inventory. Professional manual*. Odessa, Florida: Psychological Assessment Resources.

Diamond M. C., Scheibel A. B., Murphy G. M., & Harvey T. (1985). On the brain of a scientist: Albert Einstein. *Experimental Neurology, 88*(1), 198-204.
Edmondson, W. J., & House J. (1993). *Einführung in die Sprachlehrforschung*. Tübingen/Basel: Francke.
Flege J. (1995). A critical period for learning to pronounce foreign languages. *Applied Linguistics, 8*, 162-177.
Flege J., & Hammond R. (1982). Mimicry of non-distinctive phonetic differences between language varieties. *Studies in Second Language Acquisition, 5*(1), 1-17.
Flege J., & Hillenbrand J. (1987). Limits on phonetic accuracy in foreign language speech production. In G. Ioup, & S. Weinberger (Eds.), *Interlanguage Phonology* (pp. 176-203). Cambridge, MA: Newbury House.
Flege J., Munro M. J., & MacKay I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America , 97*(5), 3125-3134.
Flege J., Yeni-Komshian G., & Liu H. (1999). Age constraints on second language acquisition. *Journal of Memory and Language, 41*, 78-104.
Gardner H. (1983). *Frames of mind: The theory of multiple intelligences*. New York. Basic Books.
Gathercole S. E., Willis C. S., Baddeley A. D., & Emslie H. (1994). The children's test of nonword repetition: A test of phonological working memory. *Memory 2*(2), 103-127.
Geschwind N., & Galaburda A. M. (1985). Cerebral lateralization: Biological mechanisms, associations, and pathology: I. A hypothesis and a program for research. *Archives of Neurology, 42*, 428-459.
Gordon, E. (1989). *Advanced measures of music audiation*. Chicago: GIA Publications.
Guiora A. (1990). A psychological theory of second language pronunciation. *Toegepaste Taalwetenschap in Artikelen, 37*, 15-23.
Jilka M. (2000). *The contribution of intonation to the perception of foreign accent*. PhD Dissertation. AIMS 6(3), University of Stuttgart.
Jilka M., Möhler G., & Dogil G. (1999). Rules for the generation of ToBI-based American English intonation. *Speech Communication, 28*, 83-108.
Johnson J., & Newport E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*, 60-99.
Kuhl, P., & Meltzoff, A. (1995). Vocal learning in infants: Development of perceptual-motor links for speech. *Proceedings of the International Conference of Phonetic Sciences Stockholm*, 146-149
Larson, G. (1984). *The far side gallery 1*. London: Warner Books
Lehrl S. (2005). *Mehrfachwahl-Wortschatz-Intelligenztest. MWT-B*. Spitta Verlag.
Leibetseder, M., Laireiter, A. R., Riepler, A., & Köller, T. (2001). E-Skala: Fragebogen zur Erfassung von Empathie - Beschreibung und psychometrische Eigenschaften. *Journal für Differentielle und Diagnostische Psychologie 22*, 70-85.
Markham D. (1997). *Phonetic imitation, accent, and the learner*. PhD Dissertation. Travaux de l'Institut de Linguistique de Lund 33. Lund University Press
Mayer J. (1997). Intonation und Bedeutung: Aspekte der Prosodie-Semantik-Schnittstelle im Deutschen. PhD Dissertation. AIMS 3(4), University of Stuttgart.
Möhler G. (1998). *Theoriebasierte Modellierung der Deutschen Intonation für die Sprach-synthese*. PhD Dissertation. AIMS 4(2), University of Stuttgart.
Moyer A. (1999). Ultimate attainment in L2 phonology: The critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition, 21*, 81-108.
Neufeld G. (1987). On the acquisition of prosodic and articulatory features in adult language learning. In G. Ioup, & S. Weinberger (Eds.), *Interlanguage phonology* (pp. 321-332). Cambridge, MA: Newbury House.
Novoa L., Fein D., & Obler L. (1988). Talent in foreign languages: A case study. In L. Obler, & D. Fein (Eds.), *The exceptional brain: The neuropsychology of talent and special abilities* (pp. 294-302). New York. Guildford.
Pagagno, C., & Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *Quarterly Journal of Experimental Psychology, 48A*(1), 98-107.
Piske T., MacKay I., & Flege J. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics, 29*, 191-215.

Raven J., Raven J. C., & Court J. H. (1998). *Manual for Raven's advanced progressive matrices (1998 Edition)*. Oxford, England: Oxford Psychologists Press.

Schneiderman E., & Desmarais C. (1988). A neuropsychological substrate for talent in second-language acquisition. In L. Obler, & D. Fein (Eds.), *The exceptional brain: The neuropsychology of talent and special abilities* (pp. 103-126). New York: Guildford.

Scott, J., & Huskisson, E. C. (1976). Graphic representation of pain. *Pain, 2*, 175-184.

Silverman, K., Beckman, M., Pitrelli, J. , Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992). ToBI: A standard for labelling prosody. Proceedings of ICSLP 1992, 867-870.

Simon J. R. (1990). The effects of an irrelevant directional cue on human information processing. In R. W. Proctor, & T. G. Reeve (Eds.), *Stimulus-response compatibility* (pp. 31-86). Amsterdam: Elsevier.

Southwood M., & Flege J. E. (1999). Scaling foreign accent: Direct magnitude estimation versus interval scaling. *Clinical Linguistics and Phonetics, 13*, 335-349.

Twain, M. (1968). *Short stories*. New York: Airmont Books.

Wewers, M. E., & Lowe, N. K. (1990). A critical review of visual analogue scales in the measurement of clinical phenomena. *Research in Nursing and Health, 13*, 227-236.