The Common European Framework of Reference for Languages: a European framework for foreign language speech development?

Bettina Beinhoff
bettina.beinhoff@anglia.ac.uk
Anglia Ruskin University Cambridge, United Kingdom

ABSTRACT

This paper explores the representation of speech development and particularly perceptive skills in the CEFR level descriptions. The speech-related CEFR sections and related level descriptions are vague and none of the assumptions made therein (such as the supposed linear progression between levels) have been sufficiently tested yet. This paper presents an exploratory study on speech perception in language learners at different levels of proficiency and from different first language (L1) backgrounds (Spanish and German). The study is based on transcriptions in response to short narratives, and investigates what kind of influence listeners' levels of proficiency in the second language (in this case English) and their L1 backgrounds have on how intelligibility is perceived. The results suggest that proficiency levels and L1 background do indeed influence intelligibility (though not always as anticipated) and partially confirm the idea of a linear progression as proposed in the CEFR.

Keywords: *Language assessment, second language acquisition, speech perception, intelligibility, pronunciation, non-native speech*

I. INTRODUCTION

Ever since its publication in 2001, the Common European Framework of Reference for Language (CEFR, Council of Europe 2001) has received much attention from a variety of professional communities, such as language assessors, language teachers and researchers in the area of second and foreign language acquisition. Each of these communities has made significant contributions to the discussion and it is probably fair to say that in this way, the CEFR has stimulated a truly interdisciplinary debate.

The CEFR was introduced to serve as a 'common basis for the elaboration of […] curriculum guidelines […] across Europe' (Council of Europe 2001: 1). Its central part is a description of proficiency levels of learners' skills with the aim of facilitating

comparisons in language teaching and assessment across Europe. After presenting a brief overview of the main areas of research dealing with the CEFR, this paper will focus on phonological competence which is one of six communicative language competences (together with lexical, grammatical, semantic, orthographic and orthoepic competence) that define learners' levels of proficiency within the CEFR. In the following I will be mainly concerned with the representation of speech perception in the CEFR and I will argue that although much work still needs to be done in order to develop the CEFR to a stage where it can be fully used in language teaching and assessment, the CEFR has the potential to be an incredibly useful tool for bringing more structure into the description of 'learner Englishes' and it can provide an alternative to the increasingly criticised use of native speaker norms in second language (L2) and foreign language education. This discussion will be followed by an explorative study which takes 'intelligibility' as a means of testing whether differences between learners at different CEFR proficiency levels actually exist.

Throughout the debate surrounding the CEFR, many perceived shortcomings of the framework were raised across a wide range of areas, such as lack of detail in its descriptor scales which restricts the CEFR's applicability in language testing (Weir 2005) and specifically in test development (Alderson et al. 2006), the political and social agenda underlying the CEFR and the way this affects language testing (Fulcher 2004) and possible 'abuses' of the CEFR in testing migrant's language abilities (Krumm 2007). Nevertheless, the CEFR has been noted to make valuable contributions, including providing a general framework for language testing which builds on familiar concepts (the notion of 'beginner', 'intermediate' and 'advanced' skills) to develop new approaches (e.g. Little 2007), its potential for improving curriculum design (Westhoff 2007) and to spark discussion about the applicability and comparability of language tests across platforms (e.g. Tannenbaum and Wylie 2008), across national borders (e.g. Phakiti and Roever 2011) and across languages (e.g. Breton et al. 2010).

While most of these debates were mainly centred on language testing and assessment, further research looked at specific areas of language use and development, such as discourse structure, where for example Evison (2013) investigated turn-taking in academic spoken discourse and McCarthy (2010) studied fluency and its connection to turn-taking mechanisms such as turn-openings and turn-closings. Most attention,

however, has been devoted to issues concerning lexical and grammatical development within the CEFR with studies based on several of the 'bigger' European languages; e.g. English (Milton 2010), French (Kusseling and Lonsdale 2013) and German (Hancke and Meurers 2013). Moreover, some studies attempted to link criterial features of the CEFR to second language acquisition theory (see Hawkins and Filipović 2012 and Hawkins and Buttery 2010 for a discussion of the development of grammatical features within the CEFR with regard to Universal Grammar and cognitive theories), however, none of them dealt with phonological development.

II. SPEECH DEVELOPMENT IN THE CEFR

Compared to discourse structures and lexical and grammatical development, phonological development has received little attention in research and discussions on the CEFR so far. This is problematic because more general findings from studies in grammatical and lexical development or discourse structure may not be applicable to phonological development because it is known to diverge from other competences and does not follow the same development rates (Flege and Bohn 1989). In very extreme cases the differences between a learner's level of phonological competence and the level of competence in other areas of development can lead otherwise highly proficient second or foreign language learners to speak with an accent that is hardly intelligible (cf. the 'Joseph Conrad Phenomenon' in Scovel 1981). This discrepancy is not addressed in the CEFR proficiency level descriptions or in CEFR-related research publications.

In addition, CEFR level descriptions provide no detail on speech production and perception. While all CEFR level descriptions are intentionally vague, mostly as a matter of practicality and for historical reasons (Trim 2010), the descriptions for speech production and perception are particularly limited. There is some information on 'phonological control' which refers to speech production and mainly focuses on accentedness, intelligibility and overall fluency (see table 1).

Table 1: Requirements for phonological control (Council of Europe 2001: 117)

Note: The labels A1-C2 refer to the proficiency level of the learner, where A1 stands for beginner, A2 for waystage or elementary, B1 for threshold or intermediate, B2 for upper intermediate, C1 for advanced and C2 for mastery level.

	PHONOLOGICAL CONTROL
C2	As C1
C1	Can vary intonation and place sentence stress correctly in order to express finer shades of meaning.
B2	Has acquired a clear, natural, pronunciation and intonation.
B1	Pronunciation is clearly intelligible even if a foreign accent is sometimes evident and occasional mispronunciations occur.
A2	Pronunciation is generally clear enough to be understood despite a noticeable foreign accent, but conversational partners will need to ask for repetition from time to time.
A1	Pronunciation of a very limited repertoire of learnt words and phrases can be understood with some effort by native speakers used to dealing with speakers of his/her language group.

Perceptive skills, however, are not described as part of phonological control, nor are they described in any detail anywhere else in the CEFR. Rather, speech perception is mentioned as part of other skills; for example under "General phonetic awareness and skills" which combines production and perception:

> "Many learners, particularly mature students, will find their ability to pronounce new languages facilitated by: an ability to distinguish and produce unfamiliar sounds and prosodic patterns; an ability to perceive and catenate unfamiliar sound sequences; an ability, as a listener, to resolve (i.e. divide into distinct and significant parts) a continuous stream of sound into a meaningful structured string of phonological elements; an understanding/mastery of the processes of sound perception and production applicable to new language learning. These general phonetic skills are distinct from the ability to pronounce a particular language." (Council of Europe 2001: 107)

Teasing out the features which relate specifically to perception, the above quote differentiates between the ability to

(a) distinguish unfamiliar sounds and prosodic patterns,

(b) perceive unfamiliar sound sequences,

(c) turn a continuous stream of speech into meaningful input and

(d) understand or master the processes required in learning how to perceive (and presumably understand) a new language.

These perceptive skills are not linked to any specific proficiency level whereas more general issues of speech perception are part of "Listening comprehension". However, much of the listening comprehension level descriptions refer to the discourse level and lexical features, including content and register of speech, and are otherwise rather vague.

II.1. Linear progression

The information given in the CEFR at each proficiency level suggests a linear progression from very basic skills and knowledge at level A1 and A2 through the intermediate stages B1 and B2 to the advanced stages C1 and C2. As Hulstijn (2007) quite rightly pointed out, it is by no means clear whether this progression from one level to the next corresponds to the learners' actual development. Widely discussed theories in L2 speech acquisition, such as the Speech Learning Model (Flege 1995) and the Native Language Magnet Model (Kuhl 1993) propose specific ways in which learners progress in acquiring the sound system of a new language. For example, in the Speech Learning Model, Flege (1995) suggests that it is necessary to create a category for a new speech sound in order to be able to successfully identify it and distinguish it from similar first language (L1) and L2 sounds. Similarly, Kuhl's Native Language Magnet Model (1993) suggests that new speech sounds are perceived with reference to already known speech sounds (typically from the learner's L1). Only with repeated exposure to these new sounds does the learner (implicitly or explicitly) perceive these new sounds as meaningful new speech sounds. This indicates that speech perception and speech production are closely linked, and, in fact, research suggests that these two levels interact (e.g. Lacabex et al. 2008). This suggests that research into L2 speech production

and perception could potentially add much-needed detail to the CEFR level descriptions.

However, so far, research conducted with reference to L2 speech acquisition models has not looked at the specific development of L2 learners or include L2 learners at different levels of proficiency in line with the CEFR. Rather, the focus was mostly on learners from different first language backgrounds (e.g. Anderson-Hsieh et al. 1992, Munro et al. 2006). A notable exception is Hulstijn et al.'s (2011) study which looked at pronunciation skills to determine the connection between speaking proficiency and linguistic competences within the CEFR for learners of Dutch as an L2. This study found that pronunciation skills are an important factor in determining a learner's proficiency level, together with knowledge of vocabulary and grammar in combination with processing speed. This particular study looked at language production, but given the importance of perceptive skills in the L2 acquisition process, it is crucial to establish whether L2 speech perception is a distinguishing factor between CEFR proficiency levels, and if it is, to investigate what the differences between the proficiency levels are.

II.2. Native speaker norms and non-native speaker performance

As a brief additional note, I would like to add a few comments on a potential area of use for the CEFR which has not received much attention in academic debate as yet. Variability in learners' (non-native) speech perception is usually studied in terms of how it varies from native speaker (NS) abilities and to what extent it follows NS norms. This, however, fails to address the fact that for many, if not for most, non-native speakers, NS proficiency in the perception and production of their second/foreign language is unattainable. In addition, there has been much debate about the suitability of NS norms for non-native speakers, especially with regard to the ever increasing international and intercultural use of the English language (cf. e.g. Canagarajah and Wurr 2011, Cook 1999, He and Zhang 2010, Timmis 2002).

This debate reveals a dichotomy between those proposing 'nativeness' (which in itself is difficult to define) as a central pronunciation target and those who regard intelligibility to a wide native and non-native speaker audience as more important (Levis 2005). This need for more suitable norms and aims for language learners has

been considered by Fitzpatrick and Racine (2013) in their recent study on using L1 performance profiles as an alternative tool for L2 assessment, albeit without explicit reference to the CEFR.

This lack of alternative norms and factors is particularly an issue in speech development where, so far, most research included well-established factors such as 'Age of Learning' and 'Length of Stay in Target Language Country' (cf. Piske et al. 2001), which are not necessarily relevant for the majority of language learners. Especially with regard to the worldwide use of English, many learners – in Europe and beyond – tend to learn English in their first language environments and may not necessarily travel to a target language country; and if they do, they may not stay in the country for extended periods of time. Thus, norms and factors which were developed in contexts where language learners lived in the L2 environment for extended periods of time may not be suitable for most learners. In fact, factors such as 'age' have been shown to have a very different impact for learners in L1 environments as compared to those in L2 contexts (see Lecumberri and Gallardo 2003).

Thus, a more appropriate way of studying and assessing non-native speech would consider the learners' progress in relation to their own development and that of language learners from comparable backgrounds. In this way, the reference framework which the CEFR provides could potentially be used to track and assess learner development based on individual progress, rather than native speaker norms. Furthermore, specifying CEFR proficiency levels for non-native speech development would provide more detailed and appropriate descriptions of learner language than what is currently available (see e.g. the descriptions of different learner Englishes in Swan and Smith 2001). Descriptions of language learners' typical speech patterns (if these really exist) at any given proficiency level could therefore also contribute greatly to fairer assessment and realistic learning aims and targets.

III. INTELLIGIBILITY AND CEFR PROFICIENCY LEVELS: AIMS OF THE STUDY

The above discussion reveals that it is by no means clear whether the learner progression as suggested in the CEFR corresponds to the learners' actual development. In addition, the information given on speech perception in the CEFR is almost non-existent and does not give any information on what learners 'can do' at any particular level. The following study is a very first attempt at shedding some light on

(a)   whether there are any differences in speech perception between learners at different CEFR proficiency levels, and

(b)   if there are differences, to reveal any general trends of what might constitute these differences.

Intelligibility is one of the main concerns in the CEFR level descriptions for 'Phonological control' and it also features strongly under 'General phonetic awareness and skills'. In addition, further research into issues of intelligibility is crucial as it is still not fully determined what pronunciation features contribute to intelligibility (cf. Harding 2013). Therefore, this study will use 'intelligibility' to address the main research aim by investigating if learners of English from different proficiency levels (B1 and C2) can understand given speech samples equally well. It is possible that 'intelligibility' may follow a linear progression as it is necessary to understand before it is possible to engage in communication. Therefore, it seems likely that there may be differences between learners at B1 and C2 level. Another aim is to find out what it is that inhibits intelligibility and whether it is different for learners at B1 and C2 level of proficiency.

Given that the L1 is considered to strongly influence speech perception, this study will also investigate if learners of English from two different L1s (German and Spanish) show differences in how they perceive L2 speech. This is quite likely and is implicitly included in the CEFR section on perception where distinguishing unfamiliar sounds and perceiving unfamiliar sound sequences are central concerns (German and Spanish are phonologically very different and may therefore lead to different problems with intelligibility).

Finally, to account for the global use of English and the growing number of learners and non-native speakers (NNS) of English which – at least theoretically – makes NNS to NNS communication much more likely than NNS to NS communication, this study

aims to investigate intelligibility of NNS English. This study uses Spanish NNS accents in English as the speech stimuli because it is one of the most widely distributed NNS accents in English. Previous research is unclear about whether sharing the same L1 as the speaker is beneficial in understanding L2 communication, such as Spanish learners of English understanding Spanish-accented English (a detailed discussion is beyond the scope of this paper, but see e.g. Major et al. 2002 and Munro et al. 2006), however, it is possible that familiar sounds and patterns from L1 Spanish may be easier to understand for Spanish L1 speakers compared to, for example, German L1 speakers, because Spanish speakers are more familiar with such structures.

IV. METHODOLOGY AND PROCEDURE

In this study, 20 German and Spanish learners of English (from here on called "listeners") responded to recorded speech samples spoken by four Spanish NNS of English (from here on called "speakers"). The listeners were German and Spanish learners of English at different CEFR proficiency levels: ten German learners of English; five at proficiency level B1 (i.e. threshold level) and five at proficiency level C2 (i.e. proficient user) and ten Spanish learners of English; five at proficiency level B1 and five at proficiency level C2. In addition, five English NS took part in the study as a control group. The proficiency levels B1 and C2 were chose because the likelihood to obtain distinctive results was deemed relatively high.

All participants were students in Cambridge, UK, where the English NS and the C2 listeners studied a variety of subjects at the local universities and the B1 listeners were students at residential language schools. All listeners grew up monolingual with German, Spanish or English as their L1; the German and Spanish listeners had learned English at schools in their L1 environments before coming to Cambridge. Their proficiency levels were established through proficiency tests they had taken up to two months previously (for the B1 listeners) and up to one year previously (for the C2 listeners, who had been living in the UK since then). The Spanish and German listeners had stayed in English-speaking countries between three weeks and six months (for level B1) and between five months and seven years (for level C2). The age at which they started learning English was similar across all proficiency levels. The Spanish listeners

started learning English when they were between 6 and 9 years old and the German listeners when they were between 8 and 13 years old.

The speech samples were recordings of four Spanish NNS of English with varying degrees of influence from their L1 Spanish in their English pronunciation. All of these speakers were female. The speech samples were recorded in a sound-treated room with a digital recorder. The speakers were asked to describe three different picture stories and to talk about themselves without revealing any personal information. From these recordings, short utterances were isolated using the speech processing software Audacity (version 2.0.3). These utterances varied in length between five and 15 words. For the experiment, eight speech files per speaker were selected from these utterances (i.e. 32 speech stimuli in total).

The speech samples contain a great range of variation that is generally found in Spanish NNS accents of English (cf. Coe 2001), such as

- Variation in vowel length and vowel quality

- Variation in specific consonants (e.g. pronunciation of English /z/ towards /s/ or /□/; /b/ towards /v/ or /β/; /r/ towards [□] and [r]; /h/ towards /x/)

- Regular sentence rhythm, which often leads to vowels being produced as full vowels (instead of the weak forms) in unstressed syllables

- Strong devoicing of final consonants

- Narrower pitch range, leading to what is often perceived as a 'flat' intonation

The four speakers in this study show all of the above types of variation but differ in the extent to which this variation surfaces in their accents. In an auditory analysis, the accent of Speaker 1 showed a greater amount of the above features than any of the other three accents. The accent of Speaker 4 showed relatively few of these features while the accents of Speaker 2 and Speaker 3 ranked in-between the two.

Speech samples were played in silent conditions over headphones using a Praat (Boersma and Weenink, 2013, version 5.3.42) script on a laptop. The samples were arranged in two different randomisation patterns to address possible influences of

adjacent speech samples on the responses and also to avoid samples of the same speaker appearing in direct succession.

The listeners were asked to transcribe exactly what they heard. They could listen to each sample only once and had to write down what they heard immediately after each file was played. The task was self-paced to give sufficient time for the transcriptions. Prior to the actual task the listeners did a short familiarisation session with sentences which were not used in the study. In addition, the listeners filled out a questionnaire which asked for additional information on the participants' background and familiarity with languages other than their L1 and familiarity with NS and NNS accents of English. As expected, all Spanish listeners were more familiar with Spanish-accented English than the German and English listeners.

V. RESULTS

The 32 speech stimuli were transcribed by all 25 listeners resulting in 800 transcribed sentences. These were coded against a transcription of the speech stimuli (which were carefully transcribed by the researcher and checked by a colleague). For each transcription the number of correctly transcribed words was counted; spelling mistakes were not considered to be incorrect transcriptions in cases where the listener had clearly identified the intended word. Intelligibility scores were obtained by calculating the percentage of correctly transcribed words in each stimulus (following Derwing and Munro 1997); mean scores for each participant group were calculated for each sentence and across all stimuli of each speaker. The mean scores for stimulus transcriptions ranged from 16% (lowest) to 100%, where listeners at B1 level usually achieved the lowest value (in stimulus 1 from Speaker 1, Spanish B1 listeners scored 24% as the lowest value and German B1 listeners 16%) while the lowest value for listeners at C2 level was 24% for the German listeners and 43% for the Spanish listeners. Each group managed to transcribe a number of stimuli 100% correct. Across speakers, the lowest value for correct transcriptions was 47% (for the Spanish B1 group's transcriptions of Speaker 1's utterances) and the highest was 94% (Spanish C2 group's transcriptions of Speaker 3's utterances).
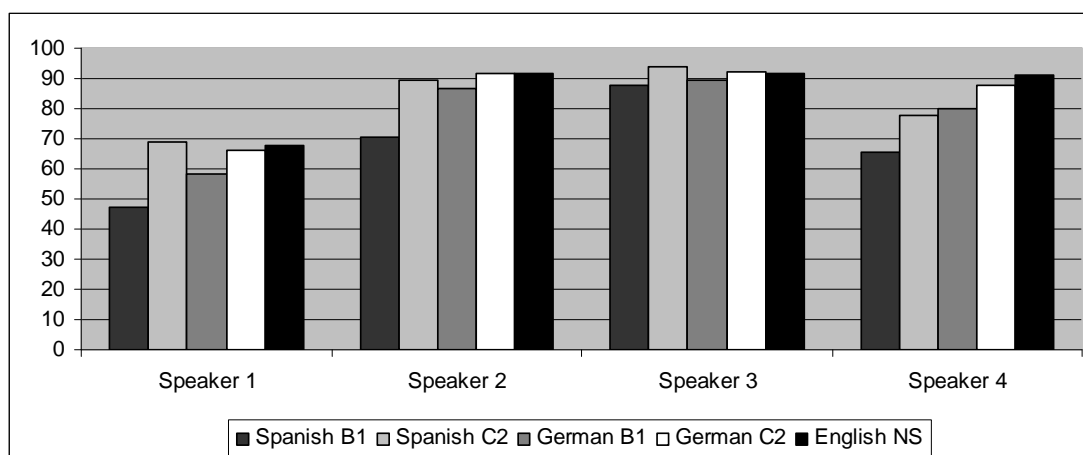
Figure 1: Mean percentages of correct transcriptions across Speakers.

The percentage of correctly transcribed words is lowest for Speaker 1 (47% correct transcriptions for the Spanish B1 group) and highest for Speaker 3 (94% for the Spanish C2 group). All participant groups seemed to have understood Speaker 3 very well, the above values are almost at ceiling level for this speaker across all groups (between 88% and 94%; see figure 1). The main differences between listener groups are evident in the results for Speaker 1, 2 and 4, where in general listeners at B1 level achieved fewer correct transcriptions than the corresponding listeners at C2 level, and for Speaker 4 the results of the German B1 listeners are level with those of the Spanish C2 listeners. According to the results, German listeners made more correct transcriptions of the utterances of Speaker 2 and 4 than the corresponding Spanish listeners; for Speaker 1, the Spanish C2 level listeners made more correct transcriptions than the German C2 listeners, but the Spanish B1 listeners transcribed fewer words correctly than the German B1 listeners. The number of correct transcriptions of the English NS is always very similar to that of the German C2 group. However, this general trend is only partially confirmed by a one-way ANOVA (with the factor 'listener-group' and the dependent factor 'speaker') and a post-hoc Tukey test, which reveal significant differences between the following pairs (see table 2).

Table 2: Significant differences in correct transcriptions (non-significant results are not reported).

	Speaker 1	Speaker 2	Speaker 3	Speaker 4
Spanish B1	Spanish C2 (p = 0.017) German C2 (p = 0.044) English (p = 0.024)	Spanish C2 (p = 0.01) German B1 (p = 0.03) German C2 (p = 0.00) English (p = 0.00)	Spanish C2 (p = 0.09)	German B1 (p = 0.05) German C2 (p = 0.002) English (p = 0.001)
Spanish C2	Spanish B1 (p = 0.017)	Spanish B1 (p = 0.01)	Spanish B1 (p = 0.09)	None
German B1	None	Spanish B1 (p = 0.03)	None	Spanish B1 (p = 0.05)
German C2	Spanish B1 (p = 0.044)	Spanish B1 (p = 0.00)	None	Spanish B1 (p = 0.002)
English	Spanish B1 (p = 0.024)	Spanish B1 (p = 0.00)	None	Spanish B1 (p = 0.001)

The results listed in table 2 indicate that the only statistically significant differences occur between the Spanish B1 group and other groups, which suggests that the other groups transcribed the stimuli equally correctly. A closer look at the transcriptions reveals that the lower number of correct transcriptions for the Spanish B1 listeners is influenced by a comparatively high number of blanket statements such as 'I did not understand a word'. In quite a few cases listeners from this group seemed to have not understood the utterance at all or they may have given up because the utterance appeared to be too difficult to understand. This did not happen as much with the German B1 listeners who by and large at least managed to understand a few words, although in this group there are also some cases where no word has been transcribed or where all transcribed words were incorrect.

It was not always possible to determine what lead to misunderstandings of certain words or whole utterances. However, there are a number of items which give some idea of what may have been the main issues involved. For example the word "catch" in the utterance "cat trying to <u>catch</u> a fish" (Speaker 1, stimulus 3) was misunderstood to mean "cut" by eight listeners (two Spanish B1, 2 Spanish C2, one German B1 and three German C2 listeners) or "cat" by three listeners (one Spanish B1, one Spanish C2 and one German B1 listener). This indicates that the final voiceless affricate was not perceived as such by these listeners, and indeed, the speaker placed more emphasis on the initial part of this affricate, though the final fricative is still audible. It also appears that in this case the vowel was perceived as a more open vowel by those who

understood "cut". In fact, in this example many listeners opted to transcribe a word which would actually make sense in this context; i.e. "trying to *cut* a fish" makes more semantic sense than "trying to *cat* a fish". Interestingly, only one German listener (from the C2 group) was able to identify the word correctly and all English NS transcribed "catch" correctly. All other listeners either transcribed "cat" or "cut" or did not attempt to transcribe the word.

Another word which stands out as difficult to perceive and transcribe is the word "beaver" in the utterance "the <u>beaver</u> got a fish in his hands" (Speaker 2, stimulus 1). All listeners were able to correctly transcribe this sentence except for the word "beaver" which was transcribed as "river" by four listeners (one Spanish B1, two German B1 and one English NS listener) and as "weaver" by one German C2 listener. Other transcriptions were nonsense entries (such as "beaber" or "viver", usually with an added question mark); five listeners correctly identified the word "beaver" (two Spanish C2, two German C2 and one English NS listener). All the other listeners did not attempt to transcribe this word. Misunderstandings were apparently caused by the speaker's pronunciation of the /b/ sounds which varied towards a bilabial fricative, which explains why some listeners perceived it as a /v/ or /w/. Again, some listeners opted for the closest known word (such as "river" and "weaver") in their transcriptions. An almost identical pattern emerged in sentence 1 of speaker 3 "the white cat is looking at the <u>beaver</u>", where all listeners were able to transcribe the sentence correctly but for the final word, which was transcribed as "river", "viver", "weaver" or similar.

Further combinations of factors caused misunderstandings, most notably final devoicing in combination with variation in vowel quality. In stimulus 2, speaker 2 "he suddenly <u>hugged</u> the hedgehog", the word "hugged" is transcribed as "hack" by eight listeners (three Spanish C2, one German B1, four English NS listeners) and as "hacked" by three listeners (one Spanish C2 and two German C2 listeners); one German C2 listener was undecided between "hack" and "hug", six listeners transcribed the word correctly (two Spanish B1, one German B1, two German C2 and one English NS listener). In this utterance, the speaker devoiced the final consonant cluster /gd/ and produced the vowel slightly more fronted.

Another noteworthy case is stimulus 8 by speaker 2 "and then I was <u>living</u> for three years", where the word "living" was transcribed as "leaving" by three German B1 and three German C2 listeners. All other listeners transcribed this word correctly, indicating that the variation in vowel length in this item was only problematic for the German listeners. However, in another instance of variation in vowel length (in stimulus 4 of speaker 3 "the cat is <u>hitting</u> the beehive") eight listeners transcribed "hitting" as "heating" (two Spanish B1, two Spanish C2, two German B1, one German C2 and one English NS listener). In addition, six listeners transcribed this word as "hid in" (one Spanish C2, one German B1, one German C2 and three English NS) which is due to the speaker's voicing of the intervocalic consonant. Another instance of vowel variation which lead to misunderstandings was in stimulus 8, speaker 4 "the box was full of <u>bees</u> and they start chasing the hedgehog". Here the word "bees" was transcribed as "beers" by seven listeners (two Spanish B1, one Spanish C2, two German C2 and two English NS listeners) and as "bears" by four listeners (two Spanish C2, one German B1 and one German C2 listener). One listener (German C2) was undecided between these two options and wrote down both, four listeners (Spanish C2, German B1, German C2 and English NS) transcribed the correct word, all other listeners did not attempt to transcribe this word.

VI. DISCUSSION

The main aim of this study was to investigate whether there are any differences between learners at different CEFR proficiency levels in order to determine whether the proficiency levels as set out in the CEFR correspond to actual differences in the performance of language learners. There are uncertainties about the validity of assumed learner progression from one level to the next (cf. Hulstijn 2007) especially with regard to speech perception and production. The results of the present study, however, reveal that learners of English at B1 level had in general more difficulties understanding the speech stimuli than learners of English at C2 level, which was especially visible in the results for speaker 1, 2 and 4. Interestingly, these differences were only statistically significant for the Spanish listener groups and not for the German groups (a finding which is in itself interesting and should be investigated further). Nevertheless, these

results confirm that there is indeed a difference between these two proficiency levels with regard to speech perception at least in some L1 groups. Considering that the development of second/foreign language speech production and perception are considered to be closely interconnected (cf. Lacabex, García Lecumberri and Cooke 2008), this result confirms that some of Hulstijn et al.'s (2011) findings on pronunciation skills may also be applicable for speech perception in some contexts.

In the transcriptions, some items stood out since many participants had difficulties understanding them. These items were analysed more closely to find general trends of what might cause intelligibility issues between learners at different proficiency levels. The main finding here is that many Spanish B1 listeners did not transcribe a particular stimulus because they did not understand a single word of the stimulus (as many of these listeners later reported or noted on their transcription sheets). It is also likely that they may have perceived the utterance as too difficult to understand and thus did not attempt to transcribe anything, or 'gave up'. It is quite possible that these listeners may have been able to transcribe at least some part of these stimuli had they been given the opportunity to listen to them again. After all, 'repetition' of speech is included in level A2 of 'phonological control' (i.e. speech production; Council of Europe 2001) and though it does not appear in the CEFR section on "General phonetic awareness and skills" there is no reason why it should not be part of level descriptions for speech perception. There were no further findings regarding a possible influence of specific patterns on the intelligibility between proficiency levels, however, this does not necessarily mean that there is none. Bearing in mind that the speech stimuli for the present study were based on natural speech in order to account for the kind of accents these learners are likely to encounter in everyday life, it is possible that speech stimuli which are controlled for specific features may reveal perceptual differences. However, based on the present study, it is not possible to show any general trends of what might constitute differences in speech perception and intelligibility between learners at B1 and C2 proficiency level other than that Spanish B1 listeners were more likely not to understand and transcribe entire stimuli.

The transcriptions and especially those of the difficult items were further analysed to reveal any differences between the Spanish and German learners of English. Based on the CEFR's account of "General phonetic awareness and skills" (Council of Europe

2001) it is tempting to predict that the Spanish listeners – who are more familiar with Spanish accents than the German listeners, as revealed by the background questionnaire – would generally be better able to understand the Spanish accents because these accents may include more familiar sounds and prosodic patterns. And indeed, in one stimulus the word "living" was misunderstood to mean "leaving" by most German speakers but not by the Spanish listeners. This may indicate that Spanish listeners are more likely to be able to process variation in vowel length as compared to German listeners (who would not normally have this feature in their accent). However, a different stimulus containing the same feature only in a different word ("hitting") was misunderstood to mean "heating" by both Spanish and German listeners. Clearly this aspect would be worth further investigation; based on the present results, however, it appears that sharing the same L1 was not an advantage in understanding the speakers. In fact, for the stimuli spoken by speakers 2 and 4, the German listeners made fewer incorrect transcriptions than the Spanish listeners. However, for speaker 1, the Spanish C2 listeners transcribed more items correctly as compared to the German C2 group, while the Spanish B1 group made fewer incorrect transcriptions than the German B1 group.

VII. CONCLUSION

This paper was made up of two parts; in the first part it presented a brief overview of previous research into the Common European Framework of Reference for Languages (CEFR; Council of Europe 2001) and gave an account of the representation of speech development in the CEFR. In the CEFR, speech production is covered by the level descriptions of "phonological control" which focuses mainly on intelligibility and accentedness of the speaker as perceived by native speakers of the language (cf. Council of Europe 2001: 117) and is otherwise rather vague. Speech perception, however, is not included in any detail and only some possible features concerning speech perception are included in the section on "General phonetic awareness and skills" (ibid.: 107). This was followed by a review of previous research into speech perception which pointed out the main areas that would benefit from further investigation in relation to the CEFR; i.e. the supposed linear progression between proficiency levels as described in the CEFR,

but not confirmed by research findings, and the close connection between speech perception and production, which is included to a limited extent in the CEFR section on "General phonetic awareness and skills". In addition, this paper briefly explored the potential of the CEFR to act as an alternative way of describing and evaluating learner speech without necessarily having to resort to native speaker norms and targets. Given that there is a rising interest in such an alternative (cf. Fitzpatrick and Racine 2013) and considering that native speaker norms are not always suitable in language learning contexts (cf. e.g. Cook 1999, He and Zhang 2010) it is well worth exploring this possibility.

The second part of this paper was devoted to an exploratory study of speech perception by learners of English at two different CEFR proficiency levels, with ten German learners of English (five at proficiency level B1 and five at proficiency level C2) and ten Spanish learners of English (five at proficiency level B1 and five at proficiency level C2). In addition, five English native speakers took part as a control group. These participants listened to speech samples based on recordings of four Spanish NNS of English with varying degrees of influence from their L1 Spanish in their English pronunciation. The participants were asked to transcribe each stimulus exactly as they heard it. The transcriptions were then analysed for the number of correctly transcribed words and for any specific sounds and patterns which may have been difficult to understand.

The study shows that some of the assumptions made in the CEFR are valid and correspond to learners' differences at the two levels of proficiency; for example, listeners at C2 level consistently transcribed more sentences correctly than the corresponding B1 level group (though this was only statistically significant for the Spanish listeners). While this study only looked at these two levels, which are quite different in terms of the learners' proficiency in the language, the results show that it would be worth looking at differences between adjacent levels, e.g. C1 and C2, which would also confirm whether the assumption that C1 and C2 are the same (as depicted in "Phonological control") holds in speech perception.

While this study did not find any particular patterns regarding what was difficult to understand for B1 vs. C2 learners and for German vs. Spanish learners, it must be born

in mind that the stimuli were naturalistic and geared towards assessing overall intelligibility. However, there were some items which proved difficult for many listeners and these may be a suitable starting point for further investigation. In general, what this study shows is that intelligibility is a suitable measure with which to investigate differences in the perception of accents between speakers at different proficiency levels.

In this paper, I hoped to show that it is crucial to conduct research into the representation of speech development in the CEFR, for the benefit of learners, language practitioners, researchers and the further development of the CEFR. The results from the study, though limited, have revealed that there are indeed several aspects which are worthy of further investigation in the area of speech perception. It is hoped that the beginning wave of research into the CEFR and especially its proficiency level descriptions will encourage further research within this framework in all areas of language acquisition research.

Notes

REFERENCES

Alderson, J.C., Figueras, N., Kuijper, H., Nold, G., Takala, S., and Tardieu, C. 2006. "Analysing tests of reading and listening in relation to the common European framework of reference: The experience of the Dutch CEFR construct project". *Language Assessment Quarterly: An International Journal*, 3 (1), 3-30.

Anderson Hsieh, J., Johnson, R., and Koehler, K. 1992. "The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure". *Language Learning*, 42 (4), 529-555.

Audacity software, Version 2.0.3 (August 2013). Open Source Software http://audacity.sourceforge.net/

Breton, G., Bolli, G.G., and Perlmann-Balme, M. 2010. "All different all equal? Towards cross-language benchmarking using samples of oral production in French, German and Italian". *ForumSprache*, 2010 (4), 5-19.

Canagarajah, A., and Wurr, A. 2011. "Multilingual communication and language acquisition: New research directions". *The Reading Matrix*, 11 (1), 1-15.

Coe, N. 2001. "Speakers or Spanish and Catalan". In Swan, M. and M. Smith (Eds.) *Learner English*. Cambridge: Cambridge University Press, 90-112.

Cook, V. 1999. "Going beyond the native speaker in language teaching". *TESOL Quarterly*, 33 (2), 185-209.

Council of Europe 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment.* 09 August 2011 <http://www.coe.int/t/dg4/linguistic/cadre_en.asp>

Derwing, T.M. and Munro, M.J. 1997. "Accent, intelligibility, and comprehensibility". *Studies in Second Language Acquisition*, 19 (1), 1-16.

Evison, J. 2013. "A corpus linguistic analysis of turn-openings in spoken academic discourse: Understanding discursive specialisation". *English Profile Journal,* 3, e4.

Fitzpatrick, T. and Racine, J.P. 2013. "Using L1 performance profiles as an alternative to native speaker norms". Unpublished poster presented at BAAL 2013, September 2013.

Flege, J.E. and Bohn, O.S. 1989. "An instrumental study of vowel reduction and stress placement in Spanish-accented English". *Studies in Second Language Acquisition* 11, 35-62.

Flege, J.E. 1995. "Second language speech learning. Theory, findings, and problems". In Strange, W. (Ed.) *Speech Perception and Linguistic Experience*. Baltimore: York Press, 233-277.

Fulcher, G. 2004. "Deluded by artifices? The common European framework and harmonization". *Language Assessment Quarterly: An International Journal*, 1 (4), 253-266.

Hancke, J., and Meurers, D. 2013. "Exploring CEFR classification for German based on rich linguistic modelling". *Learner Corpus Research Conference*, September 2013, Book of Abstracts, 54-56.

Harding, L. 2013. "Pronunciation assessment". In Chapelle, C.A. (Ed.) *The Encyclopedia of Applied Linguistics*. Oxford: Blackwell, 4708-4713.

Hawkins, J.A. and Buttery, P. 2010. "Criterial features in learner corpora: Theory and illustrations". *English Profile Journal*, 1, e5.

Hawkins, J.A. and Filipović, L. 2012. *Criterial Features in L2 English: Specifying the Reference Levels of the Common European Framework* (Vol. 1). Cambridge University Press.

He, D. and Zhang, Q. 2010. "Native speaker norms and China English: From the perspective of learners and teachers in China". *TESOL Quarterly*, 44 (4), 769-789.

Hulstijn, J.H. 2007. "The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency". *The Modern Language Journal*, 91 (4), 663-667.

Hulstijn, J.H., Schoonen, R., de Jong, N.H., Steinel, M.P. and Florijn, A. 2012. "Linguistic competences of learners of Dutch as a second language at the B1 and B2 levels of speaking proficiency of the Common European Framework of Reference for Languages (CEFR)". *Language Testing*, 29 (2), 203-221.

Krumm, H.J. 2007. "Profiles instead of levels: The CEFR and its (ab) uses in the context of migration". *The Modern Language Journal*, 91 (4), 667-669.

Kuhl, P.K. 1993. "Innate predispositions and the effects of experience in speech perception: the native language magnet theory". In de Boysson-Bardies, B., S. De Schonen, P. Jusczyk, P. McNeilage and J. Morton (Eds.) *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*. Dordrecht: Kluwer, 259-272.

Kusseling, F. and Lonsdale, D. 2013. "A corpus-based assessment of French CEFR lexical content". *Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 69 (4), 436-461.

Lacabex, E.G., García Lecumberri, M.L. and Cooke, M. 2008. "Identification of the contrast full vowel-schwa: training effects and generalization to a new perceptual context". *Ilha do Desterro* 55, 173-196.

Lecumberri, M.G. and Gallardo, F. 2003. "English FL sounds in school learners of different ages". *Age and the Acquisition of English as a Foreign Language*, 115-35.

Levis, J.M. 2005. "Changing contexts and shifting paradigms in pronunciation teaching". *TESOL Quarterly* 39 (3), 369-377.

Little, D. 2007. "The Common European Framework of Reference for Languages: Perspectives on the making of supranational language education policy". *The Modern Language Journal*, 91 (4), 645-655.

Major, R.C., Fitzmaurice, S.F., Bunta, F. and Balasubramanian, C. 2002. "The effects of non-native accents on listening comprehension: Implications for ESL assessment". *TESOL Quarterly*, 36 (2), 173-190.

McCarthy, M. 2010. "Spoken fluency revisited". *English Profile Journal*, 1, e4.

Milton, J. 2010. "The development of vocabulary breadth across the CEFR levels". In Bartning, I., M. Martin and I. Vedder (Eds.) *Communicative Proficiency and Linguistic Development: Intersections between SLA and Language Testing Research*, 211-232. 28 November 2013 < http://eurosla.org/monographs/EM01/EM01tot.pdf >

Munro, M.J., Derwing, T.M. and Morton, S.L. 2006. "The mutual intelligibility of L2 speech". *Studies in Second Language Acquisition* 28, 111-131.

Phakiti, A. and Roever, C. 2011. "Current issues and trends in language assessment in Australia and New Zealand". *Language Assessment Quarterly*, 8 (2), 103-107.

Piske, T., MacKay, I.R.A. and Flege, J.E. 2001. "Factors affecting degree of foreign accent in an L2: a review". *Journal of Phonetics* 29, 191-215.

Praat – Doing Phonetics by Computer. Version 5.3.42 (March 2013). Paul Boersma and David Weenink. University of Amsterdam <http://www.fon.hum.uva.nl/praat/>

Scovel, T. 1981. "The effects of neurological age on nonprimary language acquisition". In Andersen, W. (Ed.) *New Dimensions in Second Language Acquisition Research*. Rowley, MA.: Newbury House, 33-42.

Swan, M. and Smith, B. 2001. *Learner English*. Cambridge: Cambridge University Press.

Tannenbaum, R.J. and Wylie, E.C. 2008. Linking English-language test scores onto the Common European Framework of Reference: An application of standard-setting methodology. Princeton, NJ: Educational Testing Service. 28 November 2013 <http://origin-www.ets.org/Media/Research/pdf/RR-08-34.pdf>

Timmis, I. 2002. "Native-speaker norms and International English: a classroom view". *ELT Journal*, 56 (3), 240-249.

Trim, J.L.M. 2010. "The Modern Languages Programme of the Council of Europe as a background to the English Profile Programme". *English Profile Journal*, 1, e2.

Weir, C.J. 2005. "Limitations of the Common European Framework for developing comparable examinations and tests". *Language Testing*, 22 (3), 281-300.

Westhoff, G. 2007. "Challenges and opportunities of the CEFR for reimagining foreign language pedagogy". *The Modern Language Journal*, 91 (4), 676-679.