UvA-DARE (Digital Academic Repository)

Multilingual sound perception and word recognition

Escudero, P.

Published in:
Stem-, Spraak-, en Taalpathologie

Link to publication

Citation for published version (APA):
Escudero, P. (2007). Multilingual sound perception and word recognition. Stem-, Spraak-, en Taalpathologie, 15(2), 93-103.

Multilingual sound perception and word recognition

Paola Escudero, University of Amsterdam

Address: Institute of Phonetic Sciences

                  University of Amsterdam

                  Spuistraat 210, office 348

                  1012 VT Amsterdam

Email address: escudero@uva.nl

1

ABSTRACT

In this paper, I review two recent studies conducted to examine the sound perception and word recognition abilties of adult multilingual and bilingual speakers. The abilities to perceive sounds and identify words in a new language are essential for the global understanding of such language. However, these abilities have only recently received attention within the broad areas of bilingualism and multilingualism research. The first study reviewed here shows that speakers of a third language, in this case Dutch, perceive the contrast between words such as "tak" ("brunch") and "taak" ("task") differently from native Dutch listeners. It is also shown that this difference may decrease with proficiency in the Dutch language. The second study shows that problems with the perception of sound contrasts, in this case vowel contrasts, also lead to problems with recognizing new words containing such contrasts. This second study uses the same eye-tracking paradigm used in previous similar studies and finds that bilinguals who learn similar sounding words with their spelling differences can differentiate between the words in a word identification task. In contrast, bilinguals who learn the same words only through listening to their auditory forms match them to their two pictures indifferently. Thus, these two studies show the problems that bilinguals and multilinguals have with perceiving the sounds and identifying the words of a new language, as well as how they learn to master these abilities and what sources of information, e.g. spelling, can help them in achieving their goal.

INTRODUCTION

About 6000 languages are spoken in the world (Grimes 1992, as cited in Tucker 2002). However, according to sociolinguistic studies (e.g. Cheshire 1991, as cited in Tucker 2002), only a few languages are considered means of wider communication because they create links between many different conmunities, e.g. Arabic, English, French, Mandarin, Spanish. In addition to their strength as means of global communication, these languages tend to have a large number of speakers as well as economical, social and educational importance. Thus, many speakers of other less represented languages are forced to use and master one or more of these powerful languages. Therefore, languages such as English, French, Spanish and Chinese are also spoken as second, third and fourth languages by many more speakers than those who have them as their native language. This situation gives rise to a large number of speakers who have two or more languages as their primary means of communication. In fact, surveys suggest that there are many more bilingual or multilingual individuals in the world than there are monolingual. Given this large number of bilinguals and multilinguals, it seems of great importance to compare their language proficiency to that of monolingual speakers in order to improve our understanding of the human language faculty and to enlarge the scope of language assessment and policy making tools. Fortunately, researchers have begun to investigate the language proficiency of adult bilinguals (see for example work by Dijkstra and colleagues in the Netherlands, and Grosjean and colleagues in Switzerland). In addition, current efforts have already produced extensive literature in the domain of language development in bilingual children, as shown by the work of Genesee and colleagues in Canada. It seems of importance to continue this line of research with multilingual children and adults, and especially in language abilities which have not been throroughly studied such as language comprehension and phonological development.

The present article reviews two recent studies which aim at furthering our understanding of the development of language comprehension in bilingual and multilingual children and adults. The perception of sounds and the recognition of words containing such sounds are crucial abilities which enable rapid communication. The first study,

namely, Brasileiro & Escudero (in preparation), examines the Dutch vowel perception of adult multilingual speakers (Portuguese-English-Dutch) as compared to monolingual adults and to monolingual and bilingual children. The authors show how these diverse populations perform in the same task and how we can explain the divergences in their results. In the second study, i.e. Escudero, Hayes-Harb & Mitterer (submitted), it is shown how problems with the perception of vowel sounds may affect the recognition of words and their meanings. These authors examine the comprehension of words in highly proficient Dutch-English bilingual adults who have problems differentating between certain English vowel sounds. In the discussion session, a summary of the findings and discussion of their implications is presented.

STUDY I: Sound perception in multilingual speakers

Brasileiro & Escudero (in preparation) investigated how the sound perception of multilingual speakers compared to that of monolingual and simultaneous bilingual children and monolingual adults. Specifically, they examined the perception of similar sounding Dutch vowels which are not present in the multilinguals' first language, or bilinguals' other language, i.e. Portuguese. The authors used one Dutch vowel contrast, namely that between the Dutch words 'taak' ('task') and 'tak' ('brunch'). The vowels in these two words are represented with the phonetic symbols /a/ and /ɑ/, respectively. The difference between them lies in their quality, as measured by their first and second spectral formants, as well as in their quantity, as measured by their duration. Figure 1 shows the mean values of these Dutch vowels taken from Adank, van Hout & Smits (2004) together with the closest Portuguese (Escudero, Boersma, Bion, Rauber, Nobre & Wempe submitted).

<Figure 1 about here>

As we can see, there are indeed acoustic differences between the Portuguese and Dutch vowels, and therefore the authors expected perceptual difficulties in bilingual and multilingual listeners.

Participants. With respect to the multilingual group, 22 adult trilinguals whose native language was Brazilian Portuguese were selected. These participants had learned English in Brazil and had moved to the Netherlands after the age of 18. They had had different length of stays in the Netherlands, 5 years in average, and were divided in two proficiency groups, low and high proficiency, based on a a general proficiency test developed by the Council of the European Union (i.e. Dialang, www.dialang.org). Brasileiro & Escudero's adult monolingual and children data were reported before in Brasileiro (2006). These groups consisted of 14 adult Dutch speakers, 14 monolingual Dutch children (3-5 years), 14 simultaneous bilingual children (3-5 years), 14 monolingual Brazilian children (3-6 years) and 14 monolingual Brazilian adults.

Methodology. Brasileiro (2006) and Brasileiro & Escudero (in preparation) set out to investigate the perceptual use of the main auditory dimensions which differentiate the vowels in 'taak' and 'tak', i.e. the vowels /a:/ and /ɑ/. As shown in Figure 1, these vowels are differentiated in F1 and duration values, but F2 also plays a differentiating role, i.e. a typical /ɑ/ token has lower F1, F2 and duration values than a typical /a:/ token. In order to measure the relative use of spectral (a combination of F1 and F2) versus durational differences in vowel perception, the authors presented

all listeners with synthetic isolated vowels representing possible tokens of the two Dutch vowels. To create the synthetic stimuli, three steps between the typical values of the Dutch vowels were considered for both spectral and duration dimensions. These three different steps for the two dimensions yielded a 4x4 matrix with 16 different spectral-duration combinations. The three steps between the 4 tokens were equally distant along the auditory scale of mels for spectrum and along a logarithmic scale for duration.

Given that the participants included young children, the authors mentioned that they could only present a limited amount of tokens. Therefore, to reduce the number of stimuli, they used Boersma & Escudero's (2005) proposed technique to collect the minimum possible data in order to reliably measure relative cue weighting. Based on a perceptual experiment and data analysis reported in Escudero & Boersma (2004), Brasileiro & Escudero argue that the relative use of two dimensions in perception can be derived from the responses given to the stimuli shown in Figure 2, stimuli which fall on the edges of the 4 x 4 matrix. Thus, the stimuli set was reduced to 12 tokens which we presented 4 times in 4 blocks of 12.

<Figure 2 about here>

As for the task that the listeners performed, again, the fact that the authors had young children in their pool of participants made the authors device a task in the form of a game. This game was presented on a computer screen and it featured colourful drawings of a 'cavemaster' and two 'caveboys'. The task was to identify the caveboy who had best imitated what the master said. The master produced the 12 target stimuli shown in Figure 2 and the caveboys produced one of the two typical instances of the vowels /ɑ/ and /a:/ shown in Figure 1. This type of perception task is know as XAB categorization because the listener is presented with three tokens and has to classify the first token (X) as being similar to either the second (A) or the third (B).

Results. Following Escudero & Boersma (2004), the relative use of the two dimensions used to differentiate between the vowels, i.e. the *relative cue reliance*, was computed as the ratio between the reliance on each individual dimension, namely the duration and spectral dimensions. The *duration reliance* was computed as the percentage of /ɑ/ responses along the vertical edges of Figure 2, i.e. the total /ɑ/ responses given to the stimuli on the left edge minus the total /ɑ/ responses for the stimuli on the right edge divided by the possible number of responses along each edge, which is 12. Similarly, the *spectral reliance* was computed as the percentage of /ɑ/ responses along the horizontal edges of Figure 2 divided between 12, i.e. top edge responses - bottom edge responses / 12. In order to easily compare individual listeners, all participants were grouped in the 6 different classes proposed by Escudero & Boersma (2004), depending on their cue reliance ratio. Thus, in general, if a participant's ratio is larger than 4, the subject is judged to rely "exclusively on duration", if it is between 2 and 4, she is judged to rely "mainly on duration", and if it is between 1 and 2, she relies on "duration & spectrum". Similarly, if a participant's ratio is between 0.5 and 1, she relies on "spectrum & duration", if it is between 0.25 and 0.5, she relies "mainly on spectrum", and if her reliance ratio is less than 0.25, she relies "exclusively on spectrum". Table 1 shows the number of native Dutch adults and multilingual participants who were considered to belong to each of the six classes according to their relative cue reliance ratios.

<Table 1 about here>

As we can see, there is a large difference between groups because 79% (11 participants) of the Dutch participants relied mainly or exclusively on spectral differences to classify the vowels in 'taak' and 'tak', while only 14% (3 participants) of the multilingual participants relied mainly on this cue. Moreover, the majority of the multilingual participants, i.e. 60% (13 participants) relied mainly or exclusively on the durational differences, while none of the native Dutch adults showed this pattern of perceptual reliance. Brasileiro (2006) and Brasileiro & Escudero (in prep.) report that the large majority of the monolingual Dutch, monolingual Brazilian, and simultaneous Dutch-Portuguese bilingual children, i.e. 79%, 72%, and 67% respectively, showed the same pattern as the native Dutch adults. Also, 54% of the monolingual Brazilian adults relied on both spectral and durational cues, and 38% had the same pattern as the monolingual Dutch adults, namely reliance on spectral cues mainly or exclusively. Only one participant of this last group showed the same pattern as the majority of the multilingual listeners, namely reliance on duration cues mainly or exclusively.

With respect to the children, the results show that simultaneous bilinguals perform similar to Dutch adults and dutch monolingual children, at least at 3-5 years of age. Thus, it can be concluded that a possible difference between monolingual and bilingual children cannot be found at this age. Further research should show whether younger bilingual children (1-3 years of age) behave differently that their monolingual counterparts. On the other hand, the results of the monolingual Brazilian children are quite interesting because they look almost identical to the children exposed to Dutch. This seems to suggest that their first language, i.e. Portuguese, allows them to differentiate between the Dutch vowels, perhaps associating them to the two Portuguese vowels shown in Figure 1, namely /a/ and /ɔ/. It is worth mentioning that Brasileiro (2006) states that many children could not perform the task, and therefore were removed from the data analysis reported above, because they could not hear the differences between the vowels, i.e. they heard both vowels as Portuguese /a/.

As for the multilingual listeners, they seem to show the most interestingly divergent pattern, namely relying mainly or exclusively on a cue which is almost discarded or seldomly used by the children and by the Dutch adults. This pattern of reliance seems to be compatible with the attested availability of durational cues in the acquisition of new sound categories, as shown in Flege, Jang & Bohn (1997), Escudero & Boersma (2004), and others. These authors showed that Spanish learners of English use durational differences to classify English vowels which differ in spectral and durational properties. It seems that many multilingual listeners can also make use of duration when the spectral information is not available to them. As shown in the table, the multilinguals differed depending on their general Dutch proficiency because some multilingual listeners with high Dutch proficiency had the same performance as the children and Dutch adults, which means that native-like use of fine-grained perceptual cues can be attained in the case of adult third language learners. However, general language proficiency does not always correlate with the multilinguals' perceptual cue weighting, as shown by the exclusive use of duration even in some multilinguals with high proficiency in Dutch.

*Discussion***.** Following the Second Language Linguistic Model (L2LP, Escudero 2005), it can be said that some

Portuguese-English-Dutch multilinguals and some monolingual Portuguese listeners start by being able to perceive the quality differences between the Dutch vowels and therefore can perform the task in a native-like fashion. As for the multilinguals who use duration, the model proposes that there are two factors involved in their performance, namely prior experience with languages other than Dutch and their level of Dutch proficiency. With respect to experience with other languages, these listeners may transfer their knowledge of English vowels, which has vowel duration differences, to the task of learning Dutch, given that Portuguese does not have vowel duration differences, a possibility which is being explored in a new project (Escudero 2006). What is important to remember is that these multilinguals will need to be able to use duration and spectral information to the same extent as the monolingual adults and the monolingual and bilingual children in order to achieve the same level of accuracy in sound perception. In the next section, we will see that the level of accuracy in sound perception has an effect in the identification of words, an ability which in turn influences language comprehension in general.

STUDY II: Recognition of similar sounding words in proficient bilinguals
In the previous study, we saw that the mismatches between the vowels of the languages of multilingual listeners lead to problems in differentiating between sounds in a native-like fashion. Here I review a different study, namely Escudero, Hayes-Harb & Mitterer (submitted), which tested the effects of vowel perception difficulties on the comprehension of words containing auditorily confusable L2 sounds. This study aimed at further exploring previous research which had investigated the interrelation between the perceptual and the lexical problems faced by highly proficient bilinguals.

    Background. Weber & Cutler (2004) found that Dutch speakers with a very high proficiency in English had problems recognizing similar sounding English words which differ in sounds that they are likely to confuse in auditory perception. The authors used the *eyetracking paradigm* which provides a measure of the recognition process as it unfolds over time. In this paradigm, listeners wear a head-mounted camera which tracks the movements of their eyes as they carry out simple spoken instructions. In Weber & Cutler's study, listeners wore this eye-tracker while having to match an auditorily presented word to its correct picture and were instructed to click on the pictures they had on the screen. The eye-tracking technology then recorded their eye movements and gaze fxations while they perfom the word identification task. Using this design, Weber & Cutler found that listeners fixated their visual attention longer and more frequently on pictures of words which are auditorily confusable than pictures with no auditorily confusable words. The authors used words such as *cattle* and *kettle*, which contain a vowel contrast which has been shown to be difficult for Dutch speakers of English (Cutler, Weber, Smits & Cooper 2004). Weber & Cutler found that the picture of a *kettle* attracted more looks than pictures of less auditorily confusable words (*e.g.*, *beetle* and *bottle*) when the target word was *cattle*.

    These authors also showed that the listeners' pattern of inappropriate English word recognition was asymmetric with targets such as *cattle* yielding the recognition of *kettle* but not the reverse. Therefore, the authors concluded that Dutch listeners have established the English /ɛ/-/æ/ contrast in their lexical representations despite the fact that they cannot perceive it. To explain this unusual pattern, Cutler, Weber & Otake (2006) speculate that explicit instruction in the language, either in the form of an awareness of the contrast or through the knowledge that the words are

spelled differently, may allow these listeners to realize that a pair of words containing /æ/ and /ɛ/ is contrastive, even when they cannot reliably perceive the vowel contrast in listening tasks.

Research questions. To further explore the issues raised in Weber & Cutler (2004) and Cutler et al. (2006), Escudero et al. (submitted) had the following research questions: 1) can proficient bilinguals learn new L2 words containing auditorily confusable L2 sounds in a short period of time?, and 2) if so, can they learn a difference between these similarly sounding words if they are explicitly made aware of such a difference, e.g. through the availability of a different spelling for these words?

Participants. Fifty highly proficient Dutch-English bilinguals participated in the study. They were selected from the same pool of subjects used by Weber & Cutler (2004).

Methodology. Instead of asking the participants to recognize known English words, Escudero et al. asked them to learn the meanings of 20 different new British English words, e.g. [tEnz↔], [tænd↔k], [tunz↔r], by linking them to their pictures. The pictures were line drawings randomly selected from the same data base used by Shatzman & McQueen (2006). Figure 3 shows examples of the three drawings and their corresponding auditory forms. This design is similar to that used by Shatzman & McQueen (2006) with Dutch listeners learning novel Dutch words. In Escudero et al (submitted), participants were given instructions and addressed in English only, just like in Weber & Cutler (2004). In addition, they were told that the words they were going to learn were invented English words. That is, the authors tried as much as possible to place the participants in an English language mode (Grosjean 2000).

<Figure 3 about here>

In Escudero et al. (submitted), ten of the words had a first syllable containing either /æ/ and /ɛ/, while the other 10 words served as control fillers and contained a vowel sound which was very different from the target vowels, e.g. [tunz↔], [tund↔k]. These control fillers were included to encourage subjects to pay attention to the quality of the vowels in the first syllables of the words and not to the consonant only. The participants were randomly assigned to one of two conditions. In Condition 1, subjects learned to link the auditory form of a word with a picture. Specifically, they were presented with the auditory word and had to choose from two pictures the one that matched the auditory form. Immediately after they made their choice, they were given auditory feedback on their accuracy followed by the target word presented auditorily and matched with its correct pictured meaning. In Condition 2, the first part was identical but the feedback also contained a clue to differentiate the similarly sounding words, i.e. a difference in the spelling associated to them (e.g.<tenzer> for the word [tEnz↔], <tandek> for the word [tænd↔k]. The listeners heard five blocked repetitions of the 20 words with two picture options to choose from and five blocked repetitions with four options. This word learning phase lasted approximately 30 minutes. After learning the words, the participants in both conditions performed the same word recognition task used in Weber & Cutler (2004), which employs eyetracking technology to record the participants' eye gazes and fixations for the items presented on a computer screen.

Results and discussion. Escudero et al.'s word learning results give a positive answer to their first research question, i.e., whether bilinguals could learn similarly sounding words in a rather short period of time (30 minutes). That is, 48 of the participants achieved 90% or higher accuracy in the last block of the word learning phase.

As for the word recognition test, the results also give a positive answer to the authors second question, i.e. whether explicit awareness of a vowel contrast aids word recognition. This conclusion can be made because the authors found that only the group of subjects who were provided spelling differences during the word learning phase were able to differentiate between the words, as illustrated in Table 2. As can be seen in the table, the listeners who learned the words with their spelling differences only looked at the picture of a word containing /ɛ/ when they heard an auditorily presented word containing such a vowel, which is a pattern that replicates Weber & Cutler's finding. In contrast, the listeners who did not receive spelling diffences looked at pictures containing /ɛ/ and /æ/ when presented with a word containing /ɛ/.

<Table 2 about here>

In the table, we also observe that both groups look at the pictures of the two vowels when presented with auditory instances of words containing the vowel /æ/. It can be argued that the group that did not receive extra information to differentiate the words simply learned words which sounded the same. These results show that without explicit instruction these proficient bilinguals cannot differentiate between words containing these vowels. In contrast, the group who received a difference in spelling during word learning gained knowledge of an abstract difference between words containing the two vowels. However, they could not reliably use such new knowledge in the online perception of words containing /æ/ probably because they sometimes heard them as /ɛ/, as suggested by Cutler et al. (2006). This pattern also replicates Weber & Cutler's (2004) findings.

SUMMARY AND CONCLUSIONS

In this paper, I have reviewed an article which shows that multilinguals have problems perceiving the sounds of a third language in a native-like fashion. In addition, the second study that I reviewed shows that problems with the perception of sounds result in problems with the understanding of words even in highly proficient bilinguals. In fact, Cutler (2005) has shown that the problems with similar sounding words found in the language comprehension of bilinguals can significantly slow down their speed in understanding the speech of native speakers. In this respect, the word recognition study reviewed here shows that learners benefit from the abstract knowledge of the existence of new contrasts, which in this case was provided to them through a difference in spelling. As for the relation between sound perception and word recognition, further research should show if specific training in the perception of auditorily confusable sounds could have a positive effect on the accuracy and speed of word recognition and general language comprehension in bilinguals and multilinguals.

REFERENCES

Adank, P., van Hout, R., & Smits, Roel (2004). An acoustic description of the vowels of Northern and Southern standard Dutch. *Journal of the Acoustical Society of America, 116*, 1729-1738.

Boersma, P., & Escudero, P. (2005). Measuring relative cue weighting: a reply to Morrison. *Studies in Second Language Acquisition* 27, 4, 607-617.

Brasileiro, I. (2006). Language Mode and bilingual's speech perception. *UiLOTS Yearbook 2006*, Utrecht University.

Brasileiro, I., & Escudero, P. (in preparation). A three-way comparison in perceptual development: monolingual children vs. simultaneous bilingual children vs. adult L2 learners.

Cheshire, J. (ed.). (1991). *English around the world: Sociolinguistic perspectives.* Cambridge: Cambridge University Press.

Cutler, A. (2005). The lexical statistics of word recognition problems caused by L2 phonetic confusion. In *Proceedings of the 9th European Conference on Speech Communication and Technology (EUROSPEECH2005)* (pp. 413-416). Lisbon.

Cutler, A., Weber, A., Smits, R. & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. Journal of the Acoustical Society of America, 116. 3668-3678.

Cutler, A., Weber, A. & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics* 34, 269-284.

Escudero (2005). *Linguistic Perception and Second Language Acquisition.* LOT Dissertation series 113, Utrecht University.

Escudero, P. (2006). "Did you say *sheet* or *sh*t, beach* or *b*tch, fax* or *f*cks*?" A longitudinal study of how vowel sounds can either facilitate or impede the acquisition of a third language by immigrant communities. Veni grant proposal, awarded in July 2006, NWO personal grant No. 275.75.005.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition, 26*, 551–585.

Escudero, P., Boersma, Bion, R., Rauber, A., Nobre, D., & Wempe, T. (2007): A cross-dialectal acoustic description of vowels: Brazilian versus European Portuguese.

Escudero, P., Hayes-Harb, R., & Mitterer, H. (submitted). Novel second language words and asymmetric lexical access.

Flege, J. E., Bohn, O.-S., Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. J. Phonetics 25, 437–470.

Grimes, B. F. (1992). *Ethnologue: Languages of the world.* Dallas: Summer Institute of Linguistics.

Grosjean, F. (2000). The bilingual's language modes. In: Nicol, J. (Ed.). *One Mind, Two Languages: Bilingual Language Processing.* Oxford: Blackwell.

Shatzman, K.B. & McQueen, J.M. (2006). Prosodic knowledge affects the recognition of newly acquired words. *Psychological Science* 17, 5, 372-377.

Tucker, G. R. (2002). A global perspective on bilingualism and bilingual education. *Journal of Iberian and Latin American literary and cultural studies*, volume 2, number 2.

Weber, A. & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language* 50, 1-25.

FIGURE CAPTIONS

Figure 1. Mean F1 and vowel duration values of acoustically close Dutch and Portuguese (circled) vowels.
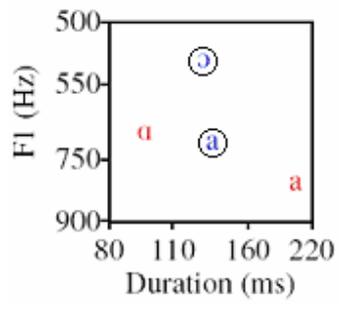
Figure 2. Stimuli used for the listening task including the Portuguese (circled) and Dutch vowels shown in Figure 1.
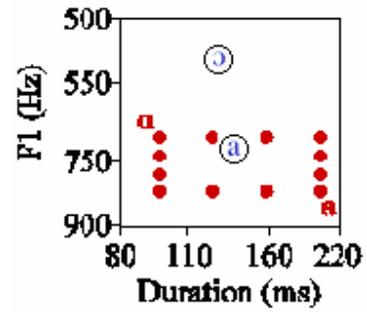
Figure 3. Three of the novel words learned by the Dutch-English bilinguals during the study's word learning phase.

TABLE CAPTIONS

Table 1. Individual relative cue reliance for the categorization of /ɑ/ and /a/ in native Dutch adults and multilingual Portuguese-English-Dutch adult listeners. R = Reliance ratio
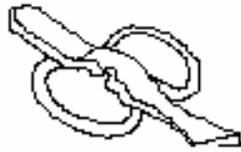
Table 2. Summary of looks to the target picture and its competitor for the listeners in the two word learning conditions, i.e. without spelling differences and with spelling differences.

[tɛnzə]   [tunzə]   [tɛndək]

Type of reliance:	exclusive duration R > 4	mainly duration R: 2-4	duration & spectrum R: 1-2	spectrum & duration R: 0.5-1	mainly spectrum R: 0.25-0.5	exclusive spectrum R< 0.25
Dutch adults	0	0	1	2	**4**	**7**
Multilinguals low proficiency.	**1**	**4**	1	4	0	0
Multilinguals high proficiency	**5**	**3**	1	0	2	1

	Learning **without** spelling differences		Learning **with** spelling differences	
	Looks to picture		Looks to picture	
	Picture: "tenzer"	Picture: "tandik"	Picture: "tenzer"	Picture: "tandik"
Auditory input: [tEnz↔]	YES	YES	**YES**	**NO**
Auditory input: [tænd↔k]	YES	YES	YES	YES