DOI: 10.17469/O2104AISV000011

ANTONIO ORIGLIA, ANTONIO RODÀ, CLAUDIO ZMARICH,
PIERO COSI, STEFANIA NIGRIS, BENEDETTA COLAVOLPE,
ILARIA BRAI, CRISTIAN LEORIN

Gamified discrimination tests for speech therapy applications

The integrity of phonetic perception abilities is necessary for a normal functioning future speech development. Since the ability to discriminate linguistic sounds is typically associated to the correct acquisition and production of the same sounds, an alteration of this ability could contribute to the onset of speech and language disorders. Support for presenting discrimination tests to young children (5- and 6-years-old), however, is provided when gamified settings are put in place. Moreover, moving beyond static tests in favour of dynamically generated ones may help personalise the test. In this work, we propose an acoustic discrimination test as the first step for the creation of a renovated *Italian Literacy Tutor*. Presented results show promising indications concerning the application of the proposed approach both from the user experience and from the reporting point of view.

Keywords: speech therapy, gamification, discrimination tests.

1. *Introduction*

1.1 The raising of intersubjectivity and cultural learning in infancy

Typical contact with language, in the first years of life, consists of a playful activity where parents and infants engage protoconversations made of rhythmical and musical content. This manifests the emotional regulation of primary intersubjectivity (Trevarthen, 1979), where interaction with the caregiver, either reciprocally directed or mediating access to objects of interest for the infant, manifests the typical playfulness often observed in mammals. At 9 months, secondary intersubjectivity arises (Trevarthen, 1978) and the baby's interest moves onto sharing the ways companions use objects as she starts to interact with the material world in a more informed way. The caregivers' language also shifts, in this phase, from questions and rhetorical comments to instructions and informative comments to support the baby's interest in participating to a task (Halliday, 1975). This is "[...] the start of cultural information transfer between generations" (Trevarthen, 2009: 74). Playful behaviour adapts to new roles as the child grows older but always stays in the background, motivating access to cultural information, reinforcing memory and supporting the creation of meaning (Trevarthen, Aitken, 2001; Reddy, 2008). Language development strongly depends on intersubjective experiences: from the effective engagement of minds and bodies depends cultural learning (Donald, 2001). The naturalistic and social context is also facilitating phonetic learning, because nine-month-old children can easily learn a new language only if they are involved in a real communicative exchange, but not when they are exposed to the acoustic signal or integrated acoustic-visual signals (like a movie) (Meltzoff, Kuhl, Movellan & Sejnowski, 2009). As a matter of fact, it has been shown that ecological learning is faster, more effective and more lasting than learning from non-naturalistic setting (Kuhl, Conboy, Coffey-Corina, Padden, Rivera-Gaxiola & Nelson, 2008).

In this paper, we will describe a software architecture designed to present discrimination tests in a playful setup depicting a social situation with different kinds of virtual agents. In fact, although humans appear to be born with a natural disposition towards cultural learning (Trevarthen, Aitken, 2003), successful acquisition of cultural skills depends on the interaction quality, especially considering social feedback. Perceived affection and playfulness by the infant towards the parents helps to establish a mutually teasing situation (Reddy, 2008) that focuses attention on rituals that may later become skills (Eckerdal, Merker, 2009). When interaction is insensitive, coercive or qualitatively poor in general, however, it elicits avoidance and protest (Gratier, Trevarthen, 2008), highlighting how "[...] infants are equipped with defensive emotions that repel unsympathetic communication" (Trevarthen, 2009: 80). Digital games, in the modern context, can be a powerful mean to channel literacy contents towards children and modern, engaging technology can be used to design a well-rounded intervention spanning different aspects of the problem at hand. It is therefore necessary to carefully consider what games are, what they are made for, and how they can provide both entertainment and tutoring. The present, ongoing work builds upon the experience of the Colorado Literacy Tutor (Cole, 2003) and of the Italian Literacy Tutor (Cosi, Delmonte, Biscetti, Cole, Pellom & Van Vuren, 2004). We decided to begin the building of such a tutor with a phonetic/phonological module, for two main reasons:

- a correct perception (and reproduction) of the sound system of a language is the *sine qua non* condition to be able to access the other levels of the spoken language;
- speech disorders and language disorders with a phonetic-phonological component are an important, if not the main, portion of the caseloads of speech therapists who deal with voice-speech-language disorders in childhood (Law, Boyle, Harris, Harkness & Nye, 2000).

1.2 Phonological discrimination in childhood

Generally speaking, phonological discrimination means that process of categorical perception through which differences that unfold along a physical continuum (of frequency, intensity, duration) are traced to discrete categories.

Phonological discrimination skills are an essential part of a normal speech perception development, and they are systematically improving up till 10 years of age (Edwards, Fox & Rogers, 2002; Hazan, Barrett, 2000; Nittrouer, 1992), although the cornerstones for a correct discrimination are already laid down by 5 years of age (Weber, Cutler, 2004; Tamashige, Nishizawa, Itoda, Kasai, Igawa & Fukuda, 2009).

Their normal development fortunately can be tested by 4-5 years of age onwards by using the same tests used for adults (Polka, Jusczyk & Rvachew, 1995).

Phonological discrimination tests are an important procedure for assessing proficiency in speech acquisition. In fact, the integrity of phonetic perception abilities is necessary, albeit not sufficient, for a future normal functioning speech development and an alteration of the ability to discriminate "similar" sounds could contribute to the onset of speech and language disorders (Brancalioni, Bertagnolli, Bonini, Gubiani & Keske-Soares, 2012; Freitas, Mezzomo & Vidor, 2015; Nithart, Demont, Majerus, Leybaert, Poncelet, & Metz-Lutz, 2009; Rvachew, Jamieson, 1989; Rvachew, Ohberg, Grawburg & Heyding, 2003; Tallal, 1976). Phonological discrimination tests could vary as to both the form and the content:

- regarding the form, i.e. the procedural paradigm used to test the phonological discrimination skill, the AX or "same/different" paradigm, is to prefer when testing young children, because of less taxing the work-memory in comparison to more sophisticated design (Polka et al., 1995).
- regarding the content, i.e. the verbal material composing the stimuli, the choice is between words and non-words stimuli. While the first are normally easier to administer even to ages earlier than five years, nonwords stimuli are to be preferred because of "the potential usefulness of processing-based measures generally in providing culturally nonbiased assessments of linguistic abilities" (Weismer, Tomblin, Zhang, Buckwalter, Chynoweth & Jones, 2000: 874).

In fact, the non-words discrimination is a task of speech perception less dependent from previous lexical knowledge, thus engaging only the perceptual system and/or the phonological memory, but not the lexical/semantic system. Phonological memory achieves the capacity to face with this task by 4-5 years of age (Polka et al., 1995), and children are successful in comparing short speech sequences out of the context, as it happens in a discrimination task, even because their vocabulary is more than 6000 words great and promotes phonological awareness (Carroll, Snowling, Stevenson & Hulme, 2003). Regarding the perceptual system, a short introduction about the perceptual skills in relationships with phonetic/phonological proficiency of children at the end of pre-school years is needed here. Because of shortage of space, we prefer not even try to resume the huge literature about the development of speech perception abilities from womb up till five years of age, and the interested reader is referred to Choi, Black & Werker (2018), Kuhl et al. (2008), Saffran, Werker & Werner (2006), Walley (2005) for some recent surveys. We focus on the period at the end of the pre-school years because it is the age-range of our sample (the younger age, as written before, at which is possible to apply the same methodologies used with the adult population). We will consider first the ability to process the acoustic dimensions of speech. Jensen, Neff (1993) demonstrated that children tested at four-years-of-age and re-tested 12-18 months later, improved speech discrimination skills beginning with variations in intensity, followed by frequency changes and finally by duration changes, but at the final assessment, for many of them, frequency and duration discrimination were still poorer than adults' discrimination. The delay in sensitivity maturation of the temporal information is due to both the central level of processing and working memory capacity.

Regarding phonological categories, the perception of the consonants is, generally speaking, less categorical and more influenced by the context than adults' perception until 5-6 years-of-age (Walley, 2005). At this age, stops are better recognized than fricatives (Tamashige et al., 2008), whose recognition is still obscured by vocalic transitions. At the same time, vowels' identification is favoured more by their relative durations than by the contextual consonants. According to Walley (2005) all these results are compatible with the hypothesis that the 4-to 5-years-old children are more dependent from a global, syllabic representation than a segmental one (Tamashige et al., 2008; Nittrouer, 1996; Bijeljac-Babic, Bertoncini & Mehler, 1993), and they still need to increase and consolidate their lexicon in order to extract all the relevant phonetic information as adults do (Walley, 2005). In fact, children aged five are less sensitive towards the position of errors within the recently acquired words, than they are towards errors within familiar, earlier-acquired words. Furthermore, the identification of consonants is more disturbed by noise in 5-years old children than in adults, especially for place identification, while, as for sonority, voiced are identified better than voiceless consonants (VOT contrasts are perceived in adults-like manner by 4-to 6 years of age, Tamashige et al. 2008). Regarding the influence of the relative position within the utterance of the consonants to be compared, previous works found that children facing with a non-word discrimination test were found to be more successful for consonants in initial rather than in final position (see McAllister-Byun, 2015). As for manner, nasals, liquids and stops are identified better than fricatives and affricates (Walley, 2005), and the identification is more facilitated if the contextual vowel is [a] rather than [u] or [i]. Generally speaking, the more two consonants share distinctive features, the more they will be confused, especially if they are voiceless, but, according to McAllister-Byun (2015) which compared adults' (Weber, Cutler, 2004) with children's phonological discrimination, the perceived distance between pairs of speech stimuli follows the same trend in the two populations, thus demonstrating that by five years children's discrimination skills are essentially adult-like.

1.3 Distinctive features in infancy

At this point a critical discussion about the concept of "distinctive feature" in relationships to phonological acquisition is needed: the term "distinctive feature" in phonology refers to a particular property of a phone/phoneme; according to the traditional theory, we can imagine the distinctive features as abstract cognitive entities that characterize a certain sound in the mind of the speaker/listener (Chomsky, Halle, 1968). In particular, Cristià, Seidl & Francis (2011) identify two main purposes in using distinctive features:

- distinctive function: they are used to distinguish sounds in contrast with each other (an acoustic difference can lead to a change in meaning, as shown by the minimum pairs, e.g. it. /'pane/ - /'kane/);
- classifying function: they determine the classes of sounds based on common characteristics, which may be subject to the same phonological rule.

It is important to underline that both of the functions described fall within the definition of distinctive features in the adult phonological system, but there is no a priori reason to think that if a child is able to make a distinction between two sounds that we describe as [+FEATURE] and [–FEATURE] (e.g. sound [+ continuous] and sound [–continuous]), then she is also able to group and classify all the sounds belonging to the [+FEATURE] category (e.g. [+continuous]) in opposition to all those with the characteristic [–FEATURE] (e.g. [–continuous]) (Cristià et al., 2011; Menn, Vihman, 2011). This is to say that we need to emphasize the importance of distinguishing the ability to discriminate two sounds (minimum pairs) from the ability to use this contrast in a phonologically relevant way (to learn new sounds), skills that can have different time courses.

Some contrasts are initially difficult to discriminate – for example, /f/ from /θ/ – and errors in the production of these consonants may have their basis in perceptual abilities (Vihman, 1996). Similarly, production errors in older children who have a speech disorder may reflect either motor problems, or an inadequate phonemic representation (Rvachew et al., 2003; Gierut, 1998). At present, the question of how perception and linguistic production are interrelated is still unresolved. According to some hypotheses there would be an integration between the two abilities from the beginning; according to others, instead, they would follow two different development paths, at least at the beginning. However, there is ample evidence that highlights the relationship between production and perception in the child's phonological development, and a large number of studies show that children have a specific difficulty in discriminating the same contrasts that neutralize in their productions (eg, McAllister-Byun, 2012; Vance, Rosen & Coleman, 2009; Whitehill, Francis, & Ching, 2003; Rvachew, Jamieson, 1989; Velleman, 1988; Hoffman, Daniloff, Bengoa & Schuckers, 1985; Locke, 1983).

In recent decades, numerous studies have shown that individual variability in linguistic production is related to individual differences in discrimination and categorical perception of linguistic sounds. According to Perkell, Guenther, Lane, Matthies, Stockmann, Tiede & Zandipour, (2004) adults who exhibited greater sensitivity in discriminating intermediate signals along the continuum /s/-/ʃ/, showed at the same time a greater acoustic contrast in the production of the same consonants. The correlation between perception and production is confirmed by other studies (Newman, 2003; Villacorta, Perkell, & Guenther, 2007; Perkell et al., 2004). The links observed between perceptual acuity and the robustness of contrast in production find a reason in theoretical and computational models such as DIVA - Directions Into Velocities of Articulators (Tourville, Guenther, 2011;

Guenther, 1995). According to this model, speakers who identify a narrower region of auditory space as the target of a certain sound are also more precise in the phonetic realization of that sound in contrast to other phonemes. In line with this model, many studies have shown that children who make mistakes in producing a given contrast also have a lower perception of the same contrast than children who produce it correctly (McAllister-Byun, 2012; Whitehill, Francis & Ching, 2003; Rvachew, Jamieson, 1989; Hoffman et al., 1985; Locke, 1983). In a recent study by Terband, Van Brenk & van Doornik-van der Zee (2014) two groups of Dutch-speaking children were compared: the first was formed by children aged 4 to 8 years with a typical language development, the second by children of the same age with language disorders. The audio recordings of the children's productions of the vowel /e/ in words with CVC structure were played back to them with changes in the height of F1 and F2. The authors observed that children with typical development successfully compensated for changes in both F1 and F2; the children with language disorders, on the other hand, did not compensate in either case, especially for F1, where they even exaggerated the perturbation instead of compensating it. This leads the authors to think that the children of the second group have an adequate auditory-perceptual ability to perceive the perturbation, but lack of the ability to modify their production to compensate for this change. In other words, it seems that children with speech impairment have difficulty integrating auditory feedback with motor planning.

These studies highlight an interesting dissociation. In adults, where the motor and perceptual systems of language are well established, there seems to be a close relationship between the two abilities (perception and production). In children a similar situation is observed, but more complex: in fact, the motor system, as well as the perceptual one, are still in the maturation phase and are developing through experience. During development, it is possible that the child perceives speech at almost the adult level, but does not yet has the motor skills to achieve a certain target (McAllister-Byun, Tiede, 2017). On the other hand, it is possible that the child has adequate motor skills, but still an auditory-perceptual representation too wide of the target, with the consequence of not being able to receive the error feedback that would lead him to modify the motor planning (Shiller, Rochon, 2014). The latter statement suggests that understanding more deeply the relationships between perception and production in the course of development would better clarify the factors that underlie the enormous variability of production capacity observed in children.

Understanding when a dissociation of development between perception and production is likely to occur, and whether one or the other will be a limiting factor, would be useful in the clinical management of language delay/disturbance. However, it is difficult to make good predictions in this regard. In fact, in the perception and in the production of speech the children undergo profound changes in the first stages, followed by a period of time in which the abilities gradually mature and become refined. The development of production compared to perceptual skills can show a variation across the different linguistic targets, speakers and developmental stages (McAllister-Byun & Tiede, 2017). Discrimination tests are usually administered following scripted approaches (as to Italian, see for instance Tressoldi, Vio, Gugliotta, Bisiacchi & Cendron, 2005). These, however, cannot take advantage of information collected during the test itself as a human therapist cannot track the child performance and find the most appropriate stimuli in real time. An Artificial Intelligence provided with fundamental knowledge about language and about the structure of a discrimination test can, instead, continuously track the child's performance and adjust the test dynamically generating new stimuli by estimating their *usefulness*.

2. *System architecture*

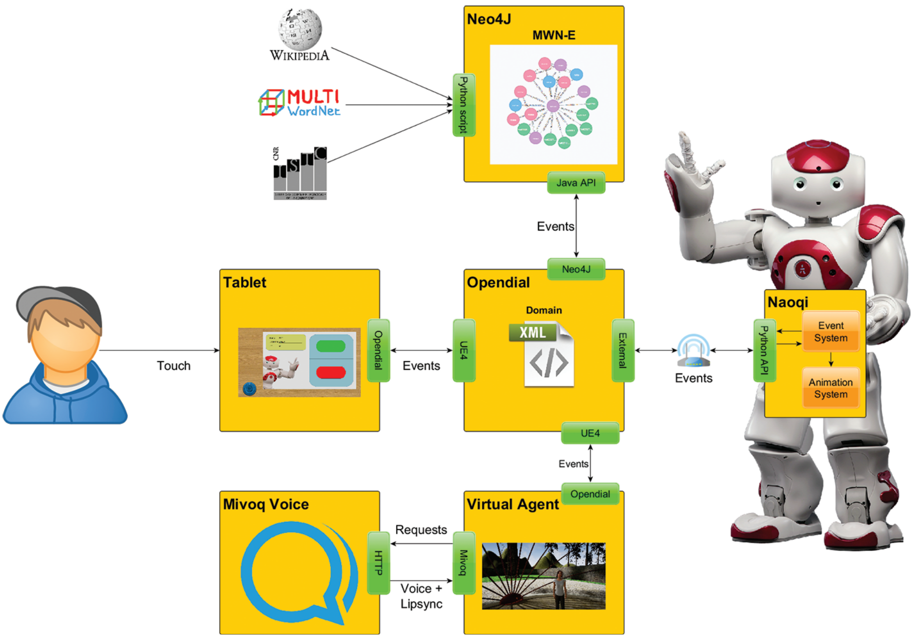
In our work, we represent the discrimination test as a dialogue model where each stimulus, once paired with the child's answer, generates a new stimulus as a system response. This stimulus is selected depending on a utility function taking into account linguistic knowledge and the child's performance. From an architectural point of view, this reflects in a dialogue manager acting as the system's controller and in linguistic knowledge being distributed between the dialogue manager and a database of Italian words. The dialogue manager, implemented using the Opendial framework (Lison, Kennington, 2016) is provided with the capability to establish which kind of information can be obtained by presenting each available stimulus and with a non-words generator which make use of phonotactic rules to avoid structures not belonging to the Italian language. The database contains morpho-syntactic, phonological and frequency data to improve the quality of the selected stimuli. In order to present the discrimination test in a social, gamified, setup, the dialogue manager controls a set of virtual agents with different characteristics. In our case, a virtual avatar is presented on a computer screen and acts as the game's guide while a social robot is used to implement a learning-by-teaching approach. The virtual avatar is controlled using the Unreal Engine 4[1] and its voice is dynamically generated using the Mivoq Voice Synthesis Engine[2], which represents the state of the art of Italian synthesis (Tesser, Sommavilla, Paci & Cosi, 2016). The synthetic voice has a number of advantages: it allows the system to be easily updated as the proposed stimuli are not pre-recorded, it allows the 3D characters to address the child by calling her by name, thus establishing a closer relationship, and it can be adapted to different kinds of characters. In the specific case of Mivoq, personalised voices and specific prosodic styles can also be synthesised, opening to a number of applications for game-like software artefacts. A tablet interface, also controlled using the Unreal Engine 4, is provided to the child to evaluate the proposed stimuli. Since the ability to adequately use a tablet interface appears to be reliable for 5 years old and onwards children (Vatavu, Cramariuc & Schipor, 2015), this is the minimum age recom-

[1] www.unrealengine.com.

[2] www.mivoq.it.

mended to apply this technology. The robot used in our implementation is Nao, which is a well established robotic platform to work with children. An overview of the system is shown in Figure 1. The technical details of the software architecture are presented in Origlia, Cosi, Rodà & Zmarich (2017).

Figure 1 - *System Architecture*



For the task of finding phonological neighbours presenting specific phones in opposition given a syllabic structure, it is possible to exploit the MWN-E database (Origlia, Paci & Cutugno, 2017), implemented as a graph in the Neo4J module (Webber, 2012). Since the phonological transcriptions are included among the properties of words, it is possible to extract phonological neighbours sharing the same syllabic structure to isolate phonological neighbours obtained through a substitution operation. Also, it is possible to specify which phonemes should be involved in the substitution in order to obtain the stimuli needed for the test. Word pairs that include words present in the *Primo Vocabolario del Bambino* (Caselli, Casadio, 1995), the Italian version of the MacArthur-Bates Communicative Developmental Inventories), are given precedence. As an alternative, word pairs with the highest average Wikipedia frequency are selected. The list of consonantal phonemes is shortened for reasons of space. A word pair represents a stimulus in the test. For each possible phoneme opposition, the system checks if a stimulus with the considered syllabic structure exists. If this is the case, the stimulus is a candidate to be presented during the test and its utility is computed, at each turn, using utility functions, computed as described in the next Section. By making the 3D avatar pronounce the target word and the Nao robot pronounce the second word, the system asks the child to decide if Nao repeated the first word correctly or not[3]. This is the same as asking if the two words are the same or not and it has the additional advantage of putting the child in an advantage position with respect to one of the involved agents, thus letting them playing the role of *teachers*.

3. *Utility functions*

In this section, we summarise the principles of the statistical modelling technique used to dynamically choose the best stimulus, given the subject's observed performance among the ones obtained by querying the MWN-E database. First of all, on the basis of the study presented in Zmarich, Bonifacio (2005), we consider the acquisition age of each phoneme. For the sake of simplicity, in this version of the model we assume that the complexity of the phones substitution is the same whether it consists of substituting a phoneme acquired later with a phoneme acquire earlier or vice-versa. For our experiments, we refer to Schmid (1999) for the distinctive features of standard Italian (see Table 1 for the reader's convenience).

Table 1 - *Distinctive features for standard Italian (Schmid, 1999, translated)*

	p	b	t	d	k	g	f	v	s	z	ʃ	ts	dz	tʃ	dʒ	m	n	ɲ	l	ʎ	r	j	w
Consonantal	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-
Sonorant	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+
Continuous	-	-	-	-	-	-	+	+	+	+	+	-	-	-	-	-	-	-	+	+	+	+	+
Del. release	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-	-	-
Voiced	-	+	-	+	-	+	-	+	-	+	-	-	+	-	+	+	+	+	+	+	+	+	+
Nasal	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	-	-	-	-
Lateral	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-
Coronal	-	-	+	+	-	-	-	-	+	+	+	+	+	+	+	-	+	-	+	-	+	-	-
Anterior	+	+	+	+	-	-	+	+	+	+	-	+	+	-	-	+	+	-	+	-	+	-	-
Posterior	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+

In this work, we do not consider the Consonantal feature in our experiments in order to concentrate on more subtle oppositions. On the other hand, we introduce the *length* feature in order to allow the system to distinguish between Italian words that are only differentiated by the phonetic realisation of a phonological geminate, as in palla /'palla/(*ball*) versus pala /'pala/ (*shovel*). With this decision we don't want to take a theoretical position about the phonological status of Italian geminates (Bertinetto, 1981): this choice is dictated by the SAMPA representation of Italian words that is provided by the pronunciations database. The probability of a subject to assign a label to the presented opposition is a binomial distribution (Equal/Different). Therefore, to represent a priori probabilities built using previous feedback, the conjugate prior of the binomial distribution, the Beta distribution, is used. Following the Opendial implementation, a two dimensional Dirichlet

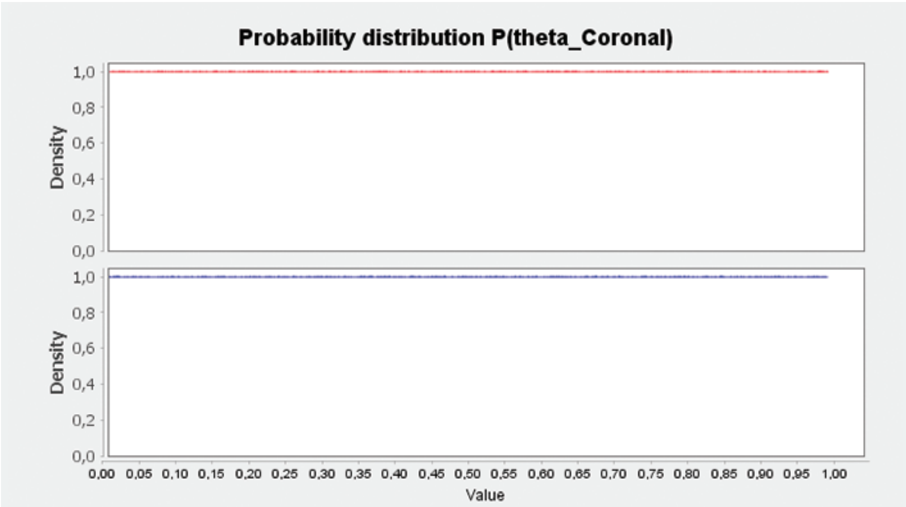
[3] The presented experiments also highlighted technological and theoretical problems due to the highly pioneering nature of the proposed system. First of all, the voice synthesizers have shown, in rare cases, problems in providing non-words. The performance of the Mivoq engine seemed superior to the one provided by the synthesizer built in the Nao robot so this problem has been addressed by integrating the Mivoq engine in the robotic platform.

probability density function with parameters is used to model the conjugate prior. An entropy-based utility function for a given opposition is computed, assigning higher utility values to stimuli presenting the opposing features for which the associated probability density functions present the higher uncertainty. An opposition presenting more than one highly entropic feature is not an optimal choice as it is not possible to evaluate which feature influenced the outcome. This is the reason why, for scripted tests, it is not possible to use phoneme pairs opposing more than one feature, which becomes a problem in tests opposing words as there may not be phonological neighbours with the specified structure opposing exactly the phonemes involved in the feature of interest. In a dynamically generated test, however, the estimated incapability of the subject to perceive oppositions on a given feature may allow to investigate other features that are never opposed in isolation, as in the case of the *posterior* feature always being opposed with the *anterior* feature. For this reason, we compute a utility function based on the mean probability value that each feature has not to be discriminated. This function assigns a higher utility value to oppositions presenting a single, highly entropic, feature together with features that have been found not to be discriminated by the subject. The higher the likelihood of other features not to be discriminated, the higher the utility. Since the task complexity can be influenced by the age acquisition difference in the involved phonemes, we model a substitution-based utility function assigning a higher value to phoneme oppositions that are closer to each other in the acquisition sequence. As this is a relative measure of phoneme-based complexity for the opposition, we also need an absolute measure to prefer phonemes acquired earlier. We therefore define an acquisition-based utility function. Since all utility functions are different measures of the same object (the phoneme opposition) sharing the same range (0, 1), the final utility function for the opposition is computed as the harmonic mean of these four measures. We use the harmonic mean because, in the considered case, different measures sharing the same range (0-1) are performed on the same subject. In this situation, the harmonic mean is the averaging method to be used. This function lets the dialogue manager select the optimal stimulus for the next turn. The algorithm for dialogue management, implemented in Opendial and exploiting the MWN-E data, proposes a stimulus at each step and updates the probability distributions according to the feedback given by the subject using Bayesian inference.

4. *Example*

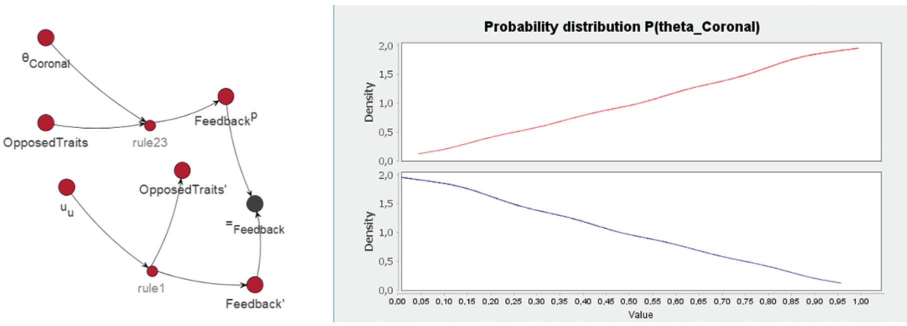
As an exemplification of the way utility functions are included in the dialogue manager, consider the case of an opposition involving the *coronal* feature that is being proposed for the first time. The Bayesian network in Opendial predicts that the probability of the child detecting or missing the opposition, given current information, is maximally entropic as no data are present. In this situation, the probability distribution is uniform on both dimensions and the probability is 50% for both cases. This situation is summarised in Figure 2.

Figure 2 - *The initial distribution of the coronal trait is uniform and maximally entropic for the probability of the child missing/detecting the opposition*



After presenting the stimulus and comparing the actual feedback from the child and the prediction, the dialogue manager uses Bayesian inference to update the probability distributions involved in the previous prediction. In this case, the probability distribution associated with the *coronal* feature is updated: as the child answers correctly, high probabilities are assigned with the first dimension of the distribution while the second is updated symmetrically. This leads to a less entropic distribution, as shown in Figure 3.

Figure 3 - *The Bayesian network compares the actual answer with the previous prediction and updates the appropriate distributions to improve future predictions*



The same strategy used for the considered features is also used to collect feedback on control stimuli, consisting of presenting two times the same stimulus or presenting two very different stimuli. A t-test on the final distributions is used to validate or reject the test.

5. *Experiments*

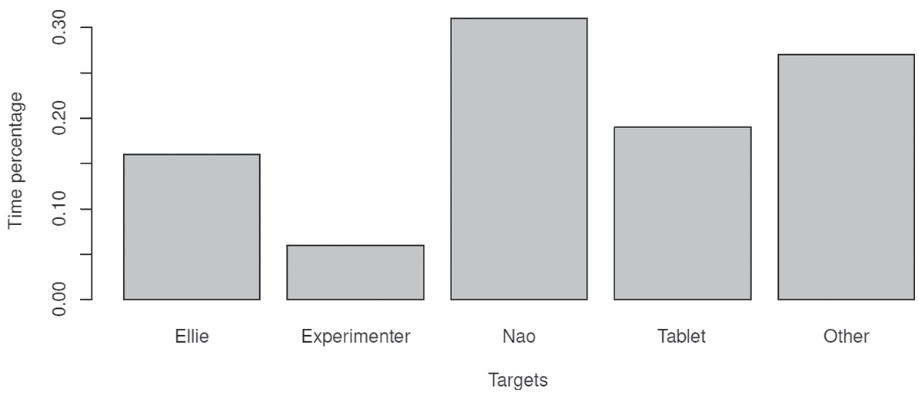
To investigate the validity of the proposed approach, we performed a pilot test by recruiting a group of 5-years old children (3 males, 2 females) to participate in a series of test sessions administered in different days and masked behind the game narrative. For each recruited child, we programmed six sessions distributed over three weeks. In order to establish a baseline for the prototype system, a group of Italian native speaking children with no reported speech and hearing and/or cognitive problems were recruited. As a reference for the children's capabilities, an entry test consisting of a) the non-word phonological discrimination subtest of BVN 5-12 (Tressoldi et al., 2005) and the b) word phonological discrimination and c) word and d) non-word repetition subtests of BVL 4-12 (Marini, Marotta, Bulgheroni & Fabbro, 2015) was also administered.

The first session (Intro) lasted approximately 6 minutes and served as an introduction to let the children familiarise with the experimental setting and with the narrative situation. The experimenter described the problem of the Nao robot, learning to speak, and introduced the virtual character, Ellie, as Nao's teacher and friend. After administering the first battery of standard tests, (BVN), the child was guided through a tutorial session to demonstrate the use of the tablet interface. During the tutorial, Nao performs a small set of funny behaviours following requests from the virtual character and the child was asked to evaluate Nao's performance using the tablet interface. At this stage, in accordance with the narrative, Nao only communicates using a set of non-verbal digital sounds. After the tutorial session, the second battery of standard tests (BVL) is administered and the child is asked to give his consent to continue helping Nao to solve its problem in the following sessions. The second session (NW1) lasts approximately 10 minutes and consists of discrimination and repetition tasks using non-words. Concerning the second phase, the system is not able, at present time, to have Nao repeat the word stated by the child and has to follow the same strategy used for the discrimination test. Currently, the system collects the audio recording to be analysed subsequently by the expert but otherwise keeps using the strategy used during the discrimination test. Future work will consist of integrating a Speech Recognition Engine specialised on children voice to address this current limitation (Cosi, Paci, Sommavilla & Tesser, 2015). The starting situation consists of assigning all features an a-priori probability corresponding to the uniform distribution. This corresponds to an initial situation in which entropy is maximised. The dialogue manager selects the most appropriate stimulus using the utility function described in Section 5.2 and coordinates the two agents so that one presents the first non-word and the second presents the second. The child's feedback, collected through the tablet interface, is used to update the statistical model and select the next stimulus. The discrimination session lasts 3 minutes and is interrupted by a cutscene in which Nao acts sadly and the virtual character explains that it is representing discouragement. This has the effect of providing emotional reactions to Nao so that the child can more easily relate. It also allows the child to release his attention from the task. Nao is ready to start the repetition test after the child, having been instructed by the virtual character, caressed Nao on the head. The repetition test lasts as much as the discrimination test and makes use of Nao's voice activity detection (VAD) system to establish when the child has repeated the non-word pronounced by the virtual character. The system has been set to a slightly higher sensitivity to stimulate the child to repeat the stimuli loud and clear. The NW1 session ends with Nao starting to produce vocalisations together with digital sounds before going to sleep and the virtual character highlights the change. In the proposed approach, Nao's evolution represents the reward for the child's effort as no feedback can be provided during the tests as part of the protocol. The third session (NW2) is identical to the first one. At the end of it, Nao stops producing digital sounds and uses vocalisations only. The virtual character informs the child that, starting from the next sessions, Nao will start to learn real words. The fourth session (W1) is identical to the first two sessions but it makes use of real words instead of non-words. The cutscene is also different, as in this case, Nao will stand up and assume an opposing pose, with its head looking away from the child. The virtual character explains that Nao does not want to study anymore and it has to be scolded. Once again, the VAD system is used to detect the child's voice and have Nao get back to work. At the end of this session, Nao starts mixing real words with non-words. The fifth session (W2) has the dialogue manager initialised with the statistical model obtained at the end of the W1 session, in order to check if the collected information improves by providing more stimuli or if one session is sufficient. The control distribution is the only one reset in this session. The cutscene is also slightly altered as this time Nao will protest to the child scolding him and will need to be scolded again before apologising (by saying sorry instead of a non-word) and going back to work. At the end of this session, Nao only uses isolated real words. The last session (Story) was designed to let us check to what extent the child's attention can be retained by the system and if the same architecture can be used to implement discrimination tests with reaction times. In this session, the virtual character reads, phrase by phrase, the Little Red Riding Hood story and Nao repeats the phrases while trying to introduce funny errors. The AI uses the MWN-E database to substitute words with phonological neighbours having the same grammatical role of the word being substituted. In such cases, for example, *Cappuccetto Rosso* (Little Red Riding Hood) becomes *Cappuccetto Rotto* (Little Broken Riding Hood) and *Nonna debole e malata* (Old and sick granny) becomes *Nonna debole e salata* (Old and salty granny). The child clicks a red button on the tablet increasing a counter to mark Nao's errors. The duration of this session is doubled with respect to the preceding ones in order to check whether it is possible, for the system, to keep the child engaged in a repetitive task for a prolonged period of time. At the end of this session, Nao becomes able to speak correctly and thanks the child for the help while the virtual character congratulates her for completing the task. The child is then administered a set of questions inspired by the USE questionnaire (Lund, 2001) to collect a subjective impression for the experience. The original 7-points Likert scale was substituted with a 3-points scale (Yes, No, So-so) in order to simplify the task for the children. At the end of the recording period, more than 4 hours of video recordings were collected for the subsequent analyses. The sessions logs were also saved to allow offline analyses. In order to objectively evaluate interest and emotional feedback, the collected video material was manually annotated by two human judges using the ELAN software (Wittenburg, Brugman, Russel, Klassmann & Sloetjes, 2006). The judges annotated data on two tiers: on the first one, they marked the children's gaze targets (Ellie, Experimenter, Nao, Tablet) while, on the second one, they marked positive and negative emotional expressions. The annotation directives for the first tier were to mark the frame containing the fixation instant of an object belonging to the experimental setup as the starting instant and the frame preceding the one where the gaze leaves the object as the ending time. Transitions and gazes that were directed away from the experimental setup were automatically marked as Other. Given the strict directives for the first tier, the two annotators produced practically identical results and it was not necessary to merge the two annotations. The annotation directives for the second tier were more subjective. The judges were asked to mark positive and negative expressions. As in this case the annotations were influenced by subjective judgement, the final segments considered for the analysis of the results are obtained by considering the annotations overlaps only. Due to video files corruption, the F1/Intro, M2/Story and M3/W1 sessions could not be analysed for the objective evaluation.

5.1 User experience

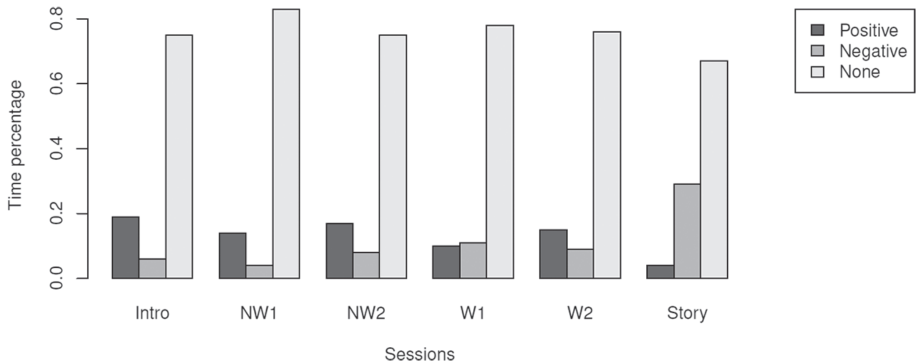
From the manual annotations of eye gaze targets, we obtain an estimate of the degree of attention children gave to the system's actors. From the general view presented in Figure 4, we observe that the Nao robot receives high attention during most of the sessions, particularly during the Intro session, highlighting the novelty effect. Children were looking at an element of the experimental setup (Ellie, Nao, Tablet) 64% of the recording time. The experimenter was not often looked at, indicating that the children had limited need to obtain support during the test and were engaged in performing the given tasks.

Figure 4 - *Overall gaze distribution during the NW1, NW2, W1 and W2 sessions*



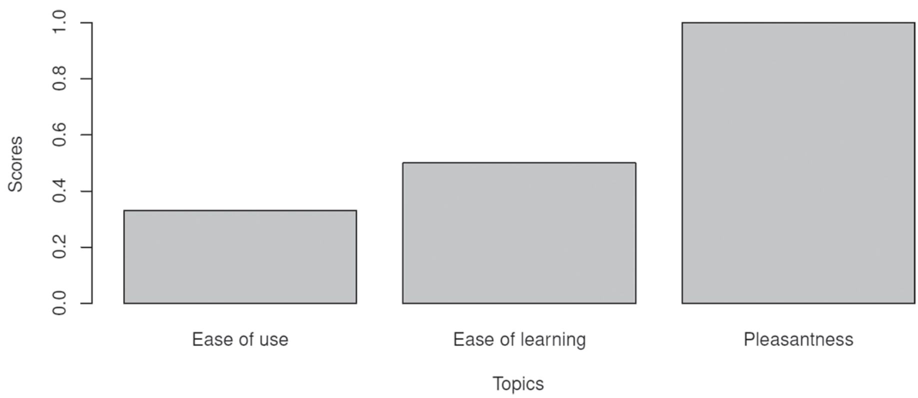
Concerning the display of emotional feedback from the children, obtained results are shown in Figure 5. The amount of positive feedback is generally higher than the amount of negative feedback.

Figure 5 - *Emotional displays during the different sessions*



Lastly, we consider the scores collected by administering the modified USE questionnaire to the children. Given the limited sample and the reduced size of the scale, we consider the median values of the scores, represented in Figure 6. The children appear to have perceived the task as difficult and not so easy to learn but they unanimously considered it to be pleasant and fun. This is consistent with the goal of task gamification: no benefits are to be expected from the point of view of the perceived difficulty of the task, which is cognitively challenging for 5 years old children, but a good disposition of the subjects towards doing it was observed in this pilot study.

Figure 6 - *Median values for the subjective scores*



5.2 Linguistic report

Examples of the final reports obtained by the proposed system are presented in Table 2 (M1/W2) and in Table 3 (F2/W2). The M1 subject was borderline in the entry tests and the system appears to correctly detect the apparent difficulties on this subject. The F2 subject, on the contrary, scored highest in the entry test and the system also identifies her as the best performing subject. The only problem detected on F2 is on the Voiced feature. Given the age of the subject, however, this is consistent with her expected capabilities, thus supporting a positive view of the feedback provided by the proposed system.

Table 2 - *Example summary (M1/W2)*

Feature	$\mu(Ok)$	$\mu(Wrong)$	σ	N	Evaluation
Control	0,865	0,135	0,079	19	Reliable
Sonorant	0,672	0,328	0,245	2	Weak Ok
Continuous	0,418	0,582	0,123	13	Strong problem
Delayed Release	0,273	0,727	0,196	4	Strong problem
Voiced	0,265	0,735	0,166	8	Strong problem
Nasal	0,500	0,500	0,289	0	Unknown
Lateral	0,702	0,298	0,236	2	Weak Ok
Coronal	0,218	0,782	0,140	15	Strong problem
Anterior	0,288	0,712	0,118	17	Strong problem
Posterior	0,626	0,374	0,267	6	Weak Ok
Length	0,227	0,773	0,111	12	Strong problem

A potential theoretical issue coming with the use of distinctive features can be observed in the two Tables: in the model we adopted for our experiments, the considered features are not necessarily opposed in a single pair of phonemes. As a consequence, the AI may not be able to test some features if an opposition on the other features they are opposed with is perceived by the considered subject. As an example, the *posterior* feature is never opposed in isolation in any possible pair of phonemes and, in the best case, a pair presenting an opposition on the *posterior* feature also opposes the *anterior* feature.

Table 3 - *Example summary (F2/W2)*

Feature	$\mu(Ok)$	$\mu(Wrong)$	σ	N	Evaluation
Control	0,735	0,265	0,101	19	Reliable
Sonorant	0,604	0,396	0,266	1	Unknown
Continuous	0,642	0,358	0,142	11	Strong Ok
Delayed Release	0,698	0,302	0,195	3	Weak Ok
Voiced	0,302	0,698	0,113	13	Strong problem
Nasal	0,500	0,500	0,289	0	Unknown
Lateral	0,500	0,500	0,289	0	Unknown
Coronal	0,776	0,224	0,138	7	Strong Ok
Anterior	0,658	0,342	0,142	9	Strong Ok
Posterior	0,500	0,500	0,289	0	Unknown
Length	0,791	0,209	0,166	3	Strong Ok

The *anterior* feature, on the other hand, can be tested in isolation (i.e. by opposing the /s/ and /S/ phones, SAMPA coding). Given the stimuli choice model, the AI therefore tests the *anterior* feature and, if the child answers correctly, estimates that it is not useful to attempt to test the *posterior* feature (i.e. by opposing the /p/ and

/k/ phonemes) as a correct answer by the child can be attributed to the opposition on the *anterior* feature. For this reason, the system checked the *posterior* feature when examining the M1 child after observing that the probability of this particular child to provide a correct answer on a stimulus opposing the *anterior* feature was low. For the F2 child, an opposition on the *anterior* feature was found to be perceived by the child, so the system did not propose stimuli involving the *posterior* feature, which is maximally uncertain (Unknown) in the report. Another problem is represented by the *lateral* feature, which was often not investigated by the system although a pair presenting an opposition on this feature in isolation exists (/l/ and /r/). This is because the system appears to overestimate the importance of the acquisition age in the utility computation (/l/ and /r/ are very distant from each other in the phonological acquisition natural history.

6. *Conclusions*

We have presented a technological system designed to administer discrimination tests to evaluate the linguistic competence of young children using a gamified setup. The system dynamically adapts the test to the children's performance using a Bayesian dialogue manager that combines linguistic knowledge with utility functions to iteratively select the most informative stimuli to be presented. Our experiments, although limited to a small sample, indicate that the chosen gamification style is able to keep the children engaged over multiple sessions distributed in a time window of three weeks, when the novelty effect introduced by the Nao robot, in particular, has worn off. The linguistic competence report obtained at the end of the administered sessions provides a detailed view of the test results, as opposed to standard tests, which provide a more general view. While the clinical validity of the approach cannot be stated at present, the ranking obtained by considering the system's report is compatible with the one obtained by administering standard tests. We consider this result to be very encouraging for our future work. The presented experiments also highlighted technological and theoretical problems due to the highly experimental nature of the proposed system. Once the problems highlighted in this first test will be solved, a larger sample of children with no reported linguistic impairments will be recruited to confirm the indications we obtained. Also, the effectiveness of the system to detect existing problems will be tested by considering children affected from dyslexia (Joanisse, Manis, Keating & Seidenberg, 2000) and/or Phonological Disorder (Brancalioni et al., 2012). The same system can also be extended to support speech treatment.

Acknowledgements

Antonio Origlia's work is supported by Veneto Region and European Social Fund (grant C92C16000250006).

Bibliography

Bertinetto, P.M. (1981). *Strutture prosodiche dell'italiano.* Accademia della Crusca.

Bijeljac-babic, R, Bertoncini, J. & Mehler, J. (1993). How do 4-day-old Infants Categorize Multisyllabic Utterances?. *Developmental Psychology,* 29(4), 711-721.

Brancalioni, A.R., Bertagnolli, A.P., Bonini, J.B., Gubiani, M.B., Keske-Soares, M. (2012). The Relation between Auditory Discrimination and Phonological Disorder. *Jornal da Sociedade Brasileira de Fonoaudiologia* 24(2), 157-161.

Carroll, J.M., Snowling, M.J., Stevenson, J. & Hulme, C. (2003). The Development of Phonological Awareness in Preschool Children. *Developmental Psychology, 39*(5), 913.

Caselli, M.C., Casadio, P. (1995). *Il primo vocabolario del bambino.* Milano: Franco Angeli.

Choi, D., Black, A.K. & Werker, J.F. (2018). Cascading and Multisensory Influences on Speech Perception Development. *Mind, Brain, and Education*, 12, 212-223.

Chomsky, N. & Halle, M. (1968). *The sound pattern of English*. Harper & Row.

Cole, R.A. (2003). Roadmaps, Journeys and Destinations Speculations on the Future of Speech Technology Research, in *Eighth European Conference on Speech Communication and Technology, Eurospeech 2003,* 2905-2908.

Cosi, P., Delmonte, R., Biscetti, S., Cole, R., Pellom, B. & Van Vuren, S. (2004). Italian Literacy Tutor Tools and Technologies for Individuals with Cognitive Disabilities. In *Proceedings Of Instil/Icall2004-NLP And Speech Technologies In Advanced Language Learning Systems-Venice* (Vol. 17), 19-27.

Cosi, P., Paci, G., Sommavilla, G., Tesser, F. (2015) Kaldi: Yet another ASR toolkit? Experiments on Adult and Children Italian Speech. In Vayra, M., Avesani C. & Tamburini F. (Eds.), *Studi Associazione Italiana Scienze della Voce* (Vol. 1), 429-438.

Cristià, A., Seidl, A. & Francis, A.L. (2011). Phonological Features in Infancy, in G.N. Clements & R. Ridouane (Eds.), *Where do Phonological Contrasts come from? Cognitive, Physical and Developmental Bases of Phonological Features,* John Benjamins Publishing Company, 303-326.

Donald, M. (2001). *A mind so rare: The evolution of human consciousness.* WW Norton & Company.

Eckerdal, P., Merker, B. (2009). Music and the 'Action Song' in Infant Development: An Interpretation. In Malloch S., Trevarthen C. (Eds.), *Communicative Musicality: Exploring the Basis of Human Companionship*, Oxford University Press, 241-262.

Edwards, J., Fox, R.A. & Rogers, C.L. (2002). Final Consonant Discrimination in Children: Effects of Phonological Disorder, Vocabulary Size, and Articulatory Accuracy. *Journal of Speech, Language, and Hearing Research,* 45(2), 231-242.

Freitas, C.R., Mezzomo, C.L. & Vidor, D.C.G.M. (2015). Phonemic Discrimination and the Relationship with Other Linguistic Levels in Children with Typical Phonological Development and Phonological Disorder. In *CoDAS 2015,* 27(3), 236-241.

Gierut, J.A. (1998). Treatment Efficacy: Functional Phonological Disorders in Children. *Journal of Speech, Language, and Hearing Research,* 41, 85-100.

Gratier, M., Trevarthen, C. (2008). Musical Narrative and Motives for Culture in Mother-Infant Vocal Interaction. *Journal of Consciousness Studies,* 15(10), 122.

Guenther, F.H. (1995). Speech Sound Acquisition, Coarticulation, and Rate Effects in a Neural Network Model of Speech Production. *Psychological Review,* 102(3), 594-621.

Halliday, M. (1975). *Learning how to mean*. London: Edwin Arnold.

Hazan, V., Barrett, S. (2000). The Development of Phonemic Categorization in Children Aged 6-12. *Journal of Phonetics,* 28(4), 377-396.

Hoffman, P.R., Daniloff, R.G., Bengoa, D. & Schuckers, G.H. (1985). Misarticulating and Normally Articulating Children's Identification and Discrimination of Synthetic [r] and [w]. *Journal of Speech and Hearing Disorders,* 50(1), 46-53.

Jensen, J.K., Neff, D.L. (1993). Development of Basic Auditory Discrimination in Preschool Children. *Psychological Science, 4*(2), 104-107.

Joanisse, M.F., Manis, F.R., Keating, P. & Seidenberg, M.S. (2000). Language Deficits in Dyslexic Children: Speech Perception, Phonology, and Morphology. *Journal Of Experimental Child Psychology 77*(1), 30-60.

Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M. & Nelson, T. (2008). Phonetic Learning as a Pathway to Language: New Data and Native Language Magnet Theory Expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences,* 363(1493), 979-1000.

Law, J., Boyle, J., Harris, F., Harkness, A. & Nye, C. (2000). Prevalence and Natural History of Primary Speech and Language Delay: Findings from a Systematic Review of the Literature. *International Journal of Language and Communication Disorders, 35*, 165-188

Lison, P., Kennington, C. (2016). Opendial: A Toolkit for Developing Spoken Dialogue Systems with Probabilistic Rules, *Proceedings of ACL-2016 System denonstrations*, 67-72.

Locke, J.L. (1983). Clinical Phonology: The Explanation and Treatment of Speech Sound Disorders. *Journal of Speech and Hearing Disorders,* 48, 339-341.

Lund, A.M. (2001). Measuring usability with the USE questionnaire*, Usability Interface* 8(2), 3-6.

Marini, A., Marotta, L., Bulgheroni, S. & Fabbro, F. (2015). BVL_4-12. *Batteria per la valutazione del linguaggio in bambini dai 4 ai 12 anni*. Firenze: Giunti OS.

MacAllister - Byun, T. (2012). Bidirectional Perception-Production relations in Phonological Development: Evidence from Positional Neutralization. *Clinical Linguistics & Phonetics*, 26(5), 397-413.

Mcallister - Byun, T. (2015). Perceptual Discrimination across Contexts and Contrasts in Preschool-Aged Children. *Lingua 160,* 38-53

Mcallister - Byun, T., Tiede, M. (2017). Perception-Production Relations in Later Development of American English Rhotics. *PloS one, 12*(2).

Meltzoff, A.N., Kuhl, P.K., Movellan, J. & Sejnowski, T.J. (2009). Foundations for a New Science of Learning. *Science, 325* (5938), 284-288.

Menn, L. & Vihman, M. (2011). Features in Child Phonology: Inherent, Emergent or Artefacts of Analysis? In Clements G.N. & Ridouane R. (Eds.), *Where do Phonological Contrasts Come From? Cognitive, Physical And Developmental Bases of Phonological Features.* John Benjamins Publishing Company, 261-301.

Newman, R.S. (2003). Using Links between Speech Perception and Speech Production to Evaluate Different Acoustic Metrics: A Preliminary Report. *The Journal of the Acoustical Society of America*, 113(5), 2850-2860.

Nithart, C., Demont, E., Majerus, S., Leybaert, J., Poncelet, M., Metz-Lutz, M. (2009). Reading Disabilities in SLI and Dyslexia Result from Distinct Phonological Impairments. *Developmental Neuropsychology*, *34*(3), 296-311

Nittrouer, S. (1992). Age-Related Differences in Perceptual Effects of Formant Transitions within Syllables and Across Syllable Boundaries. *Journal of Phonetics.*

Origlia, A., Cosi, P., Rodà, A. & Zmarich, C. (2017). A Dialogue-Based Software Architecture For Gamified Discrimination Tests. In *Proc. of the 1st Workshop on Games-Human Interaction (online)*.

Origlia, A., Paci, G. & Cutugno, F. (2017). MWN-E: A Graph Database to Merge Morpho-Syntactic and Phonological Data For Italian. In *Proc. of Subsidia 2017*.

Perkell, J.S., Guenther, F.H., Lane, H., Matthies, M.L., Stockmann, E., Tiede, M. & Zandipour, M. (2004). The Distinctness of Speakers' Productions of Vowel Contrasts is Related to Their Discrimination of the Contrasts. *The Journal of the Acoustical Society of America,* 116(4), 2338-2344.

Polka, L., Jusczyk, P.W., Rvachew, S. (1995). Methods for Studying Speech Perception in Infants and Children. In Strange, W. (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, New York: Press Timonium, 49-89.

Reddy, V. (2008). *How infants know minds.* Harvard University Press.

Rvachew, S. E, Jamieson, D.G. (1989), Perception of Voiceless Fricatives by Children with a Functional Articulation Disorder. *Journal of Speech and Hearing Disorders,* 54, 193-208.

Rvachew, S., Ohberg, A., Grawburg, M. & Heyding, J. (2003). Phonological Awareness and Phonemic Perception in 4-year-old Children with Delayed Expressive Phonology Skills. *American Journal of Speech-Language Pathology, 12,* 463-471.

Saffran, J.R., Werker, J.F. & Werner, L.A. (2006). The Infant's Auditory World: Hearing, Speech, and the Beginnings of Language. In Kuhn, D., Siegler, R. (Eds.), *Handbook of Child Psychology,* 1308-1315.

Schmid, S. (1999). *Fonetica e fonologia dell'italiano.* Paravia scriptorium.

Shiller, D.M. & Rochon, M.L. (2014). Auditory-perceptual learning improves speech motor adaptation in children. *Journal of Experimental Psychology: Human Perception and Performance, 40*(4), 1308.

Tallal, P. (1976). Rapid Auditory Processing in Normal and Disordered Language Development. *Journal of Speech and Hearing Research,* 19, 561-571.

Tamashige, E., Nishizawa, N., Itoda, H., Kasai S., Igawa, H.H. & Fukuda S. (2009). Development of Phonemic Distinction in Japanese Preschool Children, *Folia Phoniatrica et Logopaedica,* 60, 318-322

Terband, H., Van Brenk, F. & Van Doornik-Van Der Zee, A. (2014). Auditory Feedback Perturbation in Children with Developmental Speech Sound Disorders. J*ournal of Communication Disorders,* 51, 64-77.

Tesser, F., Sommavilla, G., Paci, G. & Cosi, P. (2016). Automatic Creation of TTS Intelligibility Tests, in Savy R., Alfano I. (Eds.), *La fonetica sperimentale nell'insegnamento e nell'apprendimento delle lingue straniere*, Studi AISV, Vol. 2, 289-303.

Tourville, J.A., Guenther, F.H. (2011). The DIVA Model: A Neural Theory of Speech Acquisition and Production. *Language and Cognitive Processes,* 26(7), 952-981.

Tressoldi, P., Vio, M., Gugliotta, M., Bisiacchi, P. & Cendron, M. (2005). BVN 5-11. *Batteria di valutazione neuropsicologica per l'età evolutiva*. Trento: Erickson.

Trevarthen, C. (1978). Secondary Intersubjectivity: Confidence, Confiding and Acts of Meaning in the First Year. *Action, Gesture, And Symbol: The Emergence of Language.*

Trevarthen, C. (1979). Communication and Cooperation in Early Infancy: A Description of Primary Intersubjectivity. *Before Speech: The Beginning of Interpersonal Communication,* 321-347.

Trevarthen, C. (2009). The functions of emotion in infancy. In: Fosha D., Siegel, S.M.F. & Solomon, M. (Eds.), *The healing power of emotion: Affective neuroscience, development & clinical practice* (Norton Series on Interpersonal Neurobiology), WW Norton & Company, 55-85.

Trevarthen, C., Aitken, K.J. (2001). Infant intersubjectivity: Research, theory, and clinical applications *Journal of child psychology and psychiatry 42*(1), 3-48.

Trevarthen, C., Aitken, K. (2003). Regulation of Brain Development and Age-Related Changes in Infants, Motives: The Developmental Function of "Regressive" Periods. In M. Heimann (Ed.), *Regression Periods in Human Infancy.* Mahwah, NJ: Erlbaum, 107-184.

Vance, M., Rosen, S. & Coleman, M. (2009). Assessing Speech Perception in Young Children and Relationships with Language Skills. I*nternational Journal of Audiology,* 48(10), 708-717.

Vatavu, R.D., Cramariuc, G. & Schipor, D.M. (2015). Touch Interaction for Children Aged 3 to 6 Years: Experimental Findings and Relationship to Motor Skills. *International Journal of Human-Computer Studies*, 74*,* 54-76.

Velleman, S.L. (1988). The Role Of Linguistic Perception in Later Phonological Development. *Applied Psycholinguistics*, *9*(3), 221-236.

Vihman, M.M. (1996). *Phonological Development: The Origins of Language in the Child.* Blackwell Publishing.

Villacorta, V.M., Perkell, J.S. & Guenther, F.H. (2007). Sensorimotor Adaptation to Feedback Perturbations of Vowel Acoustics and Its Relation to Perception. *The Journal of the Acoustical Society of America*, *122*(4), 2306-2319.

Walley, A.C. (2005). Speech Perception in Childhood. In Pisoni D.B., Remez R.E. (Eds.), *The Handbook of Speech Perception.* Blackwell publishing, UK, 449-468.

Webber, J. (2012). A programmatic introduction to neo4j. In *PROCEEDINGS OF THE 3RD ANNUAL CONFERENCE ON SYSTEMS, PROGRAMMING, AND APPLICATIONS: software for humanity*. ACM, 217-218.

Weber, A., Cutler, A. (2004). Lexical Competition in Non-Native Spoken-Word Recognition. *J. Mem. Lang. 50*, 1-25.

Weismer, S.E., Tomblin, J.B., Zhang, X., Buckwalter, P., Chynoweth, J.G. & Jones, M. (2000). Nonword Repetition Performance in School-Age Children with and without Language Impairment. *Journal of Speech, Language, and Hearing Research,* 43(4), 865-878.

Whitehill, T.L., Francis, A.L. & Ching, C.K. (2003). Perception of Place of Articulation by Children with Cleft Palate and Posterior Placement. *Journal of Speech, Language, and Hearing Research,* 46(2), 451-461.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A. & Sloetjes, H. (2006). ELAN: A Professional Framework for Multimodality Research. In *5th International Conference on Language Resources and Evaluation (LREC 2006)*, 1556-1559.

Zmarich, C., Bonifacio, S. (2005). Phonetic Inventories in Italian Children Aged 18-27 Months: A Longitudinal Study. In *Proceedings of Interspeech*, Lisboa, 757-760.