Chiara Celata

Partial phonological contrasts in native and non-native speech perception

(extended version of the oral communication given at the workshop "From phonetics/phonology to linguistic modeling", Pisa, April 24, 2008)

1 Introduction

In a recent model of the interplay of speech perception and phonology proposed by E. Hume and K. Johnson, phonology is conceived of as the abstract cognitive representation of a sound system, where symbols are grouped in an inventory and manipulated through a set of procedures. Several external factors, such as articulatory ease, socio-phonetic variability, and perception, are said to interact with the abstract phonological module in a bi-directional way: in other words, phonology both shapes and is shaped by external factors. Fig. 1 illustrates the model, based on the diagram contained in Hume / Johnson (2003:4) but modified to the extent that it focuses on the relation between phonology and speech perception.

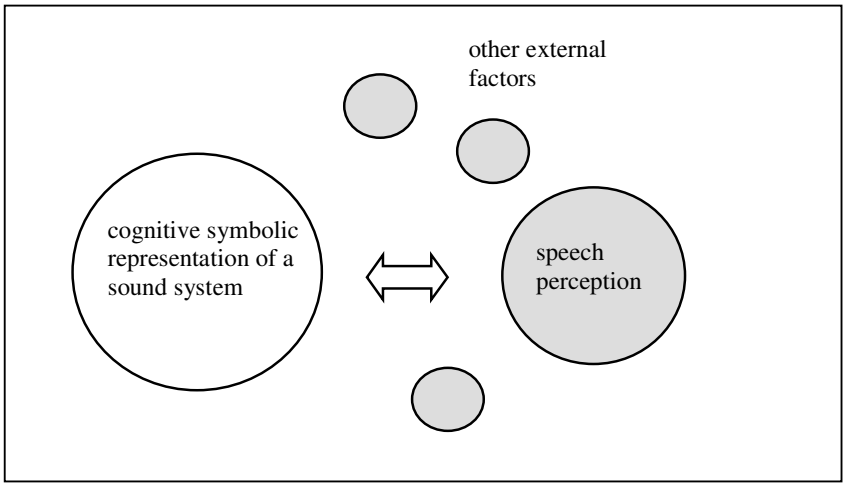

Fig. 1 The interplay of speech perception and phonology. Adapted from Hume /Johnson (2003)

There is much evidence in the phonological literature for both cases, where cognitive language sound patterns are shaped by perceptual factors (consider for example the assimilation rules), and where speech perception is influenced by phonological structure. As for the latter, it has been widely demonstrated that listeners' perceptual sensitivity is greater for sounds of one's native language than for sounds of a language acquired later in life. Japanese learners of English have difficulty discriminating the English pair /l/ vs. /r/, which does not exist in their own mother language, as Goto (1971) and many others have shown. This circumstance is explained by admitting that listeners are able to distinguish between contrastive native sounds, but definitely less successful when they have to distinguish between non-native sounds that do not serve a contrastive function in their own language.

However, as suggested by Hume / Johnson (2003), non-contrastiveness should not be conceived of as a monolithic condition, opposed to phonological contrastiveness, but rather as a range of at least three possibilities going from total non-contrastiveness (due to the lack, in a given language, of one of the two members in an opposition) to a condition of 'context-dependent' contrastiveness; the intermediate step is that of context-dependent allophony. Let us spend some more words to explain these concepts, summarized and exemplified in (1), which is also inspired by Hume / Jonhson (2003).

(1)

I. Total non-contrastiveness: non-occurrence
E.g. French /a/ - /ɑ/ for Italian listeners

II. Context-dependent allophony
E.g. French [χ] - [ʁ]

III. Context-dependent contrastiveness
E.g. Tuscan Italian N/s/ → N[ts]

IV. Full contrastiveness

For Italian listeners, the French opposition between /a/ and /ɑ/ is non-contrastive, simply because /ɑ/ does not exist in Italian. This is what we refer to as total non-contrastiveness. Two sounds can also be both present in a given language, but non-contrastive anyway, if they are involved in an allophonic relation. There are at least two different types of allophony. The first is a context-dependent allophony, i.e. when the two variants stand in complementary distribution and are automatically selected by the context. In French, for example, either [χ] or [ʁ] is obligatorily present if the following consonant is either voiced or voiceless, respectively; no other possibility is admitted. Allophones of this sort have been demonstrated to be more difficult in perception for native listeners, at least when they are inserted in their mandatory context, with respect to contrastive pairs of sounds (e.g. Peperkamp / Pettinato / Dupoux 2003, Whalen / Best / Irwin 1997, Pegg / Werker 1997, Boomershine / Currie Hall / Hume / Johnson 2008, Shea / Curtin 2005). Such an 'allophonic effect' is generally explained considering that two allophones, behaving as two surface variants of one and the same phonological object, are cast on a single phonological category; this is assumed to invalidate their perceptibility.

The second type of allophonic relation is that of contrast neutralizations. This is a case of allophonic non-contrastive relation again, but a step further toward full contrastiveness; for that reason, the label of context-dependent contrastiveness can be proposed (beside that of partial phonological contrast proposed by Hume / Johnson 2003). When a neutralization occurs, two contrastive sounds merge in a specific context, where one segment surfaces as the exclusively (or preferred) variant, the other being ungrammatical (or at least, less grammatical than the first). The reason why one

variant, and not the other, is selected in that particular context connected to the phonetic characteristics of the surrounding context. However, different from context-dependent allophony, the selection of the variant is not mandatory, and can be influenced by factors such as those regulating speech production (for example, in clearly uttered speech the alternative variant is not unlikely to surface).

The perceptual status of this class of sound objects is unclear. The possible predictions in this respect are conflicting. If the above mentioned allophonic effect is grounded on category distinction, two allophones stemming from a neutralization rule should behave like fully contrastive sounds, since they pertain to two different phonological categories. On the other hand, if their reduced contrastiveness in one specific context has an impact on the perceptual distinctiveness of the two sounds, we should expect them to pattern with other non-contrastive sound pairs, and in particular with allophones in a complementary distribution.

There is an additional problem. If the functional load of a sound pair is relevant to shape its perceptual distinctiveness, one also wonder which level of phonological processing is the locus of this effect. Listeners indeed are likely to be sensitive to different degrees of sound contrastiveness when they are processing the phonological content of word-sized units; on the other hand, they presumably are not, when asked to process the low-level acoustic information carried by a linguistic stimulus, and access to the lexical level is not requested. Previous studies on the perception of non-contrastive sound pairs, such as the ones cited above, did not directly address this question. This is a large-scale problem, going beyond the scope of this study; it can only be addressed by investigating listeners' performance on different task types. In order to address the question, then, the task variable itself has to be considered as an experimental factor. In the following paragraphs, a first attempt is made to state the problem in its concrete terms.

To test the perceptual status of context-dependent contrastive sound pairs, the Tuscan Italian allophonic rule of post-sonorant affrication was chosen. In some Western Tuscan dialects, the two phonemes /s/ and /ts/ contrast in all positions, with the exception of post-sonorant contexts, where a neutralization occurs. Here, the affricate /ts/ is the only possible outcome, as sketched in (2):

(2) /s/ → [ts] / [r, l, n] ___

The process occurs in both word-internal and sandhi position (e.g. /orso/ → [ortso] 'bear'; /il sole/ → [il tsole] 'the sun'). As is often the case for neutralizations, the process is more likely to apply, the more the speech is casual, uncontrolled, informal, fast and so on. Thanks to distributional gradience, etymological /s/ is in principle recoverable by the listener and, when ever produced, totally grammatical.

With respect to previous studies, where segments involved in the phonemic contrast were different from those involved in the allophonic contrast (e.g. the discrimination of [χ] - [ʁ] was compared with the discrimination of [m] – [n] by native French speakers in Peperkamp et al. 2003), in the present study the discrimination of the *same* consonant pair is tested in different contextual conditions (intervocalic vs. post-sonorant /s/ - /ts/). As a consequence, the possibility that acoustic dissimilarity is the real reason for perceptual differences between intervocalic and post-sonorant /s/ and /ts/ can be ruled out.

2 First experiment

The first experiment was a classical forward gating task, adapted for highly skilled early bilinguals (Sebastián-Gallés / Soto-Faraco 1999, Grosjean 1996).

The choice of this experimental design was motivated by the three following circumstances. First of all, the adapted version of Sebastián-Gallés / Soto-Faraco (1999) makes this paradigm sensitive enough to be used for testing populations which are at the upper limit of non-native phonemic processing (highly skilled early bilinguals, as in Sebastián-Gallés / Soto-Faraco 1999, or diglots, like the subjects of the present study). Second, it addresses the issue of the amount of acoustic-phonetic information needed to identify a stimulus. Third, in addition to that, it is a measure of online processing in mental activation of word-sized forms (stimuli are identified as word-like entities).

2.1 Participants

Participants were 24 Western Tuscan speakers and 24 Northern Italian speakers. The second group was chosen because in its native variety the /s/ - /ts/ opposition in post-sonorant context does not undergo neutralization.

2.2 Materials

12 minimal pairs of disyllabic non-words stressed on the first syllable were created. They were divided on two subsets: experimental, and control items. The experimental items were 3 pairs of non-words containing a post-sonorant /s/ or /ts/ (e.g. *ansu – anzu*) and 3 pairs of non-words containing an intervocalic /s/ or /ts/ (e.g. *assu – azzu*). Control items were exactly the same, with the only exception of the target consonant: 3 pairs of non-words contained a post-sonorant /t/ or /d/ (e.g. *antu – andu*), and 3 pairs of non-words contained an intervocalic /t/ or /d/ (e.g. *attu – addu*). The /t/ - /d/ contrast was chosen because of its contrastive status in any context and for any variety of regional Italian.

Following the standard procedure, the *alignment* point for each stimulus pair was determined by visual inspection (acoustic analysis). The alignment point is defined as the point where the members of each pair diverged. This point was assigned with gate 3. From gate 3 on, successive gates were created by adding or subtracting 20 ms. For all stimuli, gate 10 corresponded to the whole non-word.

2.3 Procedure

As a first stage, subjects had to familiarize with the experimental stimuli. The two whole non-word stimuli in the first pair were auditorily presented on headphones in an alternating sequence, and with simultaneous visual feedback on a computer screen. 12 repetitions were allowed as a maximum. Next, the first gate of one of the two stimuli was auditorily presented, with simultaneous visual display of both. Participants had then to identify the stimulus, and press either keys ‘Alt’ or ‘AltGr’ on the computer keyboard to indicate that the fragment corresponded to the one written on the left or on the right of the screen, respectively. The elicitation of a confidence judgment using a 9-point scale followed (1 for totally unconfident, 9 for totally confident responses), again pressing the corresponding button of the computer keyboard.

When this procedure was completed, the next gate was presented. When gate 10 was reached, familiarization with the following stimulus pair began.

There was no time constraint for responses. The presentation of the stimuli was fully balanced across participants (each participant only listened to one member of each pair for each gate) and each participant was presented with a different random order of the pairs. Participants performed the experiment individually.

2.4 Analysis

The *isolation* and *recognition* points were determined for each participant and each pair. The isolation point (IP) is defined as the gate at which a participant correctly identifies the target word, without any change in response thereafter. The recognition point (RP) is defined as the gate at which the participant has not only correctly identified the stimulus, but shows a confidence rating of 8 or more thereafter.

Moreover, accuracy (error rate) was analysed.

The main factors in the analysis were STATUS (experimental vs. control), GROUP (Northern vs. Tuscan speakers), and CONTRAST (intervocalic vs. post-sonorant).

2.5 Results

The isolation and recognition points were determined first.

When a participant failed to correctly identifies even the tenth stimulus, we adopted the conservative procedure proposed by Sebastián-Gallés / Soto-Faraco (1999) and scored the responses incorrectly identified at gate 10 as 11 (assuming therefore that a participant who did not correctly identify the stimulus 10 'almost' did so).

The mean values showed that the two populations did not differ overall (Tuscan speakers: IP 4.90, RP 8.39; Northern speakers: IP 4.62, RP 8.43). The same was true for the percentage correct answer at gate 10 (Tuscan speaker 98.3%, Northern speakers 98.8%).

The interaction STATUS x GROUP x CONTRAST was calculated for IP and RP separately. In the first case, the interaction was tendentially significant ($F(7, 616) = 3.194$, $P = .062$), while in the second case the interaction was non significant ($F(7, 616) = 0.142$, $P > .50$). The marginal significance for the IP was then further investigated, as far as the difference between experimental and control stimuli was concerned. While in

the control condition the interaction GROUP x CONTRAST resulted to be non significant in both analyses (by subjects: F1 (3, 28) = 0.443, P > .05; by items: F2 (3, 95) = 0.109, P > .05), in the experimental condition the interaction appeared to be significant, at least in the analysis by items (F1 (3, 28) = 1.385, P > .05; F2 (3, 95) = 8.777, P < .05). Then the behaviour of Tuscan vs. Northern subjects was investigated. The first group showed a significant difference between the post-sonorant and the intervocalic discrimination, with the first more impaired than the second (average IP for post-sonorant condition 5.76, for the intervocalic condition 4,60). In the case of Northern speakers, on the contrary, there was no difference between the two conditions (average IP for post-sonorant condition 4.59, for the intervocalic condition 4,83). We can conclude from this that, as far as IP is concerned, clear evidence for the existence of an allophonic effect for Tuscan subjects' discrimination was found. On the contrary, RP was totally uninformative at this regard (average RP for Tuscan speakers on experimental items: 8.18 for the post-sonorant condition, 8.45 for the intervocalic condition).

As for accuracy, the pattern of results appeared to be similar to that found for IP. It was only for Tuscan speakers, in the experimental condition, that a significant difference was found, for both analyses, between intervocalic and post-sonorant items' discrimination (error percentage on intervocalic stimuli 9.3, on post-sonorant stimuli 13.1; $\chi 2$ (1) = 6.722, P < .01; $\chi 2$ (2) = 10.940, *P* < .005).

In conclusion, IP and accuracy showed than Tuscan subjects needed significantly more information than Northern subjects to make a correct choice on exactly the same materials. Tuscan speakers, who were able to perceive the phonemic contrast in intervocalic context, needed longer portions of information to correctly label the two sounds in the contrast neutralization context.

Therefore, it appears from this experiment that two sounds in a relation of context-dependent contrastiveness tend to pattern as if they were non contrastive.

One could legitimately ask why no effect on RP was found. In order to answer this question, we should recall the above-mentioned distinction between IP and RP, insofar proponents of the gating technique have elaborated it. The gating technique has been developed within the framework of the classical *Cohort Theory* for the recognition of words in isolation (see e.g. Marslen-Wilson / Tyler 1980). Following this theory, IP should correspond to the point at which a candidate is activated, while RP should

correspond to the point at which the word becomes uniquely distinguishable from all other candidates. It follows from this explanation that, in our experiment, Tuscan and Northern speakers seem to differ in the size of the segment needed to identify the stimulus, while do not differ in the lexical representation of it.

3 Second experiment

The second experiment (Celata 2008 in press) aimed at testing whether the experience with a neutralization rule in the L1 affected the perception of a corresponding consonant contrast in an L2. This is equivalent, on one hand, to verifying the robustness of the allophonic effect (is it robust enough to influence L2 speech perception?). On the other hand, we need to provide subjects with sufficient contextual information in order to elicit different processing strategies for native vs. non-native stimuli; word-level recognition units must be the target of the discrimination. The gating technique did not appear to be suitable for that. A two-alternative forced choice identification task (2AFC) mixed with a procedure of word identification in noise was rather chosen (Gerrits 2001; Garcia Lecumberri / Cook 2006). In this technique, stimuli encapsulated in full sentences provide subjects with a specific linguistic background (either native or non-native); moreover, the paradigm is supposed to be a measure of online processing in mental activation of word-sized forms (since stimuli are identified as word-like entities). An additional disadvantage for non-native speech perception, with respect to the native one, is expected due to noise.

3.1 Participants, materials and procedure

For details on materials and method, see the extended version in Celata (2008 in press). Participants were the same of the previous experiment. It is worth stressing here that they were totally inexperienced in Russian.

Six minimal pairs of disyllabic Italian pseudo-words were created and embedded in nonsense frame sentences. Similar to the previous experiment, there were both experimental (e.g. /ansa/-/antsa/, /issa/-/ittsa/) and control items (e.g. /anta/-/anda/, /etta/-/edda/). Moreover, six minimal pairs of Russian words, embedded in frame sentences, were also created, which matched the Italian ones as far as the vocalic

neighborhood was concerned (e.g. experimental items *romansa - livantsa*, *melissa - te litsa*; control items *dva banta - ta banda*, *obed dam -banket tam*). As an example of the frame reference, take *rapo triparsa la tufi* for Italian and the corresponding *Ya tebe' povtorja'ju tri barsa čëtko* for Russian. The sentence stimuli could be uttered either in quiet, or in noise. As for noise, a fragment of babble noise was selected from SPIB database (http://spib.rice.edu/) and added to cover and exceed the duration of the sentence stimulus. The SNR was a rather adverse one (0 dB). Total duration was 4 sec, intensity was set at 70 dB, RMS amplitude at 0.063 Pa.

Initially, the two alternatives were presented visually on a computer screen, written in capital letters (e.g. ANSA ANZA). After, the sentence stimulus containing one of the two (non)word stimuli was auditorily presented on headphones. Then subjects had to identify the stimulus, pressing either keys 'Alt' or 'AltGr' on the computer keyboard, to indicate that the stimulus contained in the sentence was the one written on the left or on the right of the screen, respectively. Subjects were also requested to give a confidence judgment elicitation, using a 9-point scale. As mentioned before, for each subject half of the stimuli were presented in quiet, half in noise (random order).

3.2 Results

First of all, the effect of noise was checked. Since it resulted to be significant in both accuracy and confidence rating analyses, we are allowed to conclude that the addition of noise was an altering factor for the discrimination of native and non-native speech, as expected. As for accuracy, the percentages of correct and incorrect answers were 41.1 and 8.9, respectively, for the noise condition, and 44.3 and 5.7 for the quiet condition, calculated on the very total ($\chi 2 = 10.292$, $P < .01$). Similarly, the mean confidence rating was significantly higher for the quiet condition than for noise (7.76 vs. 6.16, $\chi 2 = 135.140.292$, $P < .01$).

Given these premises, we were allowed to consider the noise condition alone as the relevant subset for further analysis.

Within the experimental items, the number of errors made by Tuscan subjects resulted to be significantly higher for post-sonorant items than for intervocalic ones (error percentage for post-sonorant context 18.1, for intervocalic context 1.9, calculated within the factor CONTRAST for Tuscan speakers only; $\chi 2 = 10.438$, $P < .01$). No other

comparisons, for Northern speakers or within the control condition, resulted to be significant.

The analysis focused then on the performance of Tuscan speakers alone, in order to clarify whether the effect found above was due to either the native or the non-native stimuli identification, or even both. The number of errors resulted to be significantly higher in post-sonorant condition than in intervocalic condition, for both the native and the non-native Russian stimuli (error percentages for the native stimuli 17.3 vs. 1, for the non-native 19.2 vs. 3.8, calculated within the factor CONTRAST for Tuscan speakers; $\chi2 = 9.102$, $P < .05$ and $\chi2 = 5.275$, $P = .062$, respectively).

As for the confidence ratings, a similar but not identical picture emerged. The only significant comparison between intervocalic and post-sonorant stimuli was found for the native subset of stimuli in Tuscan speakers' perception (average confidence rating 8.12 vs. 7.15, respectively; $\chi2 = 5.837$, $P = .054$); no effect was found for non-native stimuli identification.

We are thus allowed to conclude that Tuscan speakers, who were able to perceive the phonemic contrast in intervocalic context, were less successful and less accurate than Northern speakers at identifying the two sounds in the contrast neutralization context. Moreover, the experience with a neutralization rule in the L1 plays some role in the perception of a corresponding consonant contrast in an L2, at least as far as accuracy was concerned.

4 Final remarks

Despite its status as contrastive in the language, the /s/-/ts/ pair displayed a perceptual merging effect for Tuscan listeners similar to what we might expect for non-contrastive elements.

Therefore, we are allowed to conclude that neutralizations behave like allophonic processes with variants in complementary distribution, at least when a categorical mode of perception is involved.

As others have also suggested, a fully predictive model of the influence of phonology on speech perception needs to take into account the different types of phonological relations that hold between sounds in a given language.

5 Bibliographical references

Boomershine A., K. Currie Hall, E. Hume & K. Johnson (2008), *The Impact of Allophony vs Contrast on Speech Perception.* In P. Avery, B.E. Dresher & K. Rice (eds.), *Contrast in Phonology: Perception and acquisition*. New York: Mouton de Gruyter.

Celata C. (2008 in press), "The impct of allophonic variation on L2 speech perception". In M.A. Watkins, A.S. Rauber & B.O. Baptista (eds.), *Recent Research in Second Language Phonetics/Phonology: Perception and Production*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Garcia Lecumberri M.L. & M.P. Cooke (2006), "Effect of masker type on native and non-native consonant perception in noise", *Journal of the Acoustical Society of America* 119, 2445-54.

Gerrits E. (2001), *The categorisation of speech sounds by adults and children*. Utrecht, LOT.

Goto H. (1971), "Auditory perception by normal Japanese adults of the sounds "l" and "r"", *Neuropsychologia* 9, 317-323.

Grosjean F. (1996), "Gating", *Language and Cognitive Processes* 11(6), 597-604.

Hume E. & K. Johnson (2003), "The impact of partial phonological contrast on speech perception", in *Proceedings of the 15th International Congress of Phonetic Sciences*.

Marslen-Wilson W.D. & L.K. Tyler (1980), "The temporal structure of spoken language understanding". *Cognition* 8, 1-71.

Pegg J.E. & J.F. Werker (1997), "Adult and infant perception of two English phones**",** *Journal of the Acoustical Society of America* 102, 3742-3753.

Peperkamp S., F. Pettinato & E. Dupoux (2003), "Allophonic Variation and the Acquisition of Phoneme Categories". In B. Beachley, A. Brown & F. Conlin (eds.), *Proceedings of the 27th Annual Boston University Conference on Language Development*, vol. II, Sommerville, Cascadilla Press, 650-661.

Sebastián-Gallés N. & S. Soto-Faraco (1999), "Online processing of native and non-native phonemic contrasts in early bilinguals", Cognition 72, 111-123.

Shea C. & S. Curtin (2005). "Learning Allophones from the Input". Poster presented at the 30th Annual Boston University Conference on Language Development. [also available at http://128.197.86.186/posters/30/SheaBUCLD2005.pdf]

Whalen D.H., C.T. Best & J.R. Irwin (1997), "Lexical effects in the perception and production of American English /p/ allophones", *Journal of Phonetics* 25, 501-528.