

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Subjective Confidence of Acoustic and Phonemic Representations During Speech Perception

Permalink

<https://escholarship.org/uc/item/44d31767>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 34(34)

ISSN

1069-7977

Authors

Schoenherr, Jordan

Logan, John

Winchester, Amy

Publication Date

2012

Peer reviewed

Subjective Confidence of Acoustic and Phonemic Representations During Speech Perception

Jordan Schoenherr (psychophysics.lab@gmail.com)

John Logan (john_logan@carleton.ca)

Amy Winchester (awinches@connect.carleton.ca)

Department of Psychology, Carleton University
1125 Colonel By Drive, Ottawa, ON K1S5B6 Canada

Abstract

Although many speech perception studies have suggested that long-term memory representations of phonemes induce categorical perception along an acoustic continuum (e.g., voice-onset time; VOT) when identifying speech sounds, other studies have suggested that acoustic information is preserved and that graded responses can be observed in within-category comparisons. Using subjective confidence reports, we present findings that support the use of both acoustic and phonemic cues during speech perception. Replicating earlier findings, we observed evidence for two well-defined phoneme categories along the voice-onset time continuum. Additionally, we also observed overconfidence in responses suggesting that the explicit representation of phonemes differs from the representations used to make identification and discrimination responses. Taken together with results from other studies, our findings support the claim that listeners can access both phonemic and acoustic representations, with explicit knowledge of the former but not the latter.

Keywords: speech perception, category boundaries, confidence processing

Introduction

Listeners presented stimuli from a continuum of sounds varying in an acoustic cue such as voice-onset time respond as though they perceive sharp discontinuities or category boundaries when tested in identification and discrimination tasks. On their own, however, measures of identification (ID) and discrimination (e.g., AX) do not indicate the extent to which participants are aware of these categories or their boundaries. In order to examine subjective awareness of categories and category boundaries, we used post-decisional confidence reports after the ID and discrimination responses to examine how response certainty varies across the continuum.

Categorical Perception of Speech Sounds

Results from speech perception studies were originally interpreted as indicating speech sounds are perceived as members of discrete phonemic categories (e.g., Liberman, Harris, Hoffman, Griffith, 1957). In such studies an acoustic cue such as VOT is varied to generate a continuum of stimuli (e.g., Pisoni, 1973; for a review, see Raphael, 2005). Participants then identify items using two categories such as /ba/ and /pa/. Importantly, the resulting ID functions show a well-defined category boundary that partitions the continuum into two phonemic categories. Moreover, in the accompanying discrimination task, participants exhibit greater accuracy when discriminating between speech

sounds from two phonemic categories relative to those from the same phonemic category even when the absolute acoustic differences are equivalent. Collectively, these ID and discrimination results constitute the phenomenon of categorical perception.

Categorical perception is a robust phenomenon. However listeners also retain some with-category acoustic information (Pisoni & Tash, 1974) and can show graded responses within a phonemic category (e.g., McMurray, Tanenhaus, Aslin, & Spivey, 2003; Miller, J. L., & Volaitis, 1989). Using an AX task, Pisoni (1973) found that when two speech sounds were presented with a long inter-stimulus interval (ISI), participants perceived could not discriminate between stimuli. When stimuli were presented with shorter ISIs, however, participants' accuracy at making within-category comparisons improved (cf. Werker & Logan, 1985). Pisoni (1973) used these findings to suggest a two-stage model for speech perception with an initial stage that processes acoustic information and a second stage that retrieves categorical cues from long-term memory (see also Pisoni & Tash, 1974). Further support for multiple representations also comes from studies of so-called sine-wave speech wherein a participant will identify the stimuli as either noise or speech depending on their prior expectations (e.g., Remez, Rubin, Pisoni, & Carrell, 1981; Davis & Johnsrude, 2007; for other top-down effects, see also Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005).

One possible account of these findings is that participants only have explicit access to phonemic information and are insensitive to acoustic properties of the stimuli that could be used to parse the continuum in an alternative manner. If an effective measure of explicit knowledge about category structure can be obtained that can be contrasted with performance on perceptual tasks, we should be capable of determining the extent to which listeners are aware of stimulus properties beyond explicit phonemic categories.

Subjective Awareness and Confidence Reports

Whether participants can maintain a representation and yet have little or no awareness of it is a controversial issue. For instance, in a typical experiment assessing such awareness, participants perform a task and indicate how certain they are in their response on a subjective probability scale (e.g., with 50% representing a guess and 100% representing complete certainty). In these tasks participants'

confidence reports typically deviate from their actual performance, i.e., they are miscalibrated. *Calibration* assesses the difference between the subjective probability of an event (confidence level) and the observed probability of a correct response (proportion correct; for formulae, see Baranski & Petrusic, 1994). In this way, calibration represents the extent to which participants are aware of their primary decision on a trial-to-trial basis. When assessing average confidence and accuracy in a given condition, participants are typically either *overconfident* [confidence > p(cor)] or *underconfident* [confidence < p(cor)]. These systematic deviations have been argued to be evidence for both implicit and explicit representations of knowledge (e.g., Dienes & Berry, 1997). For instance, if overconfidence is observed, this suggests an explicit representation that is less accurate than the implicit representation used to classify stimuli. Confirming this, recent studies do find overconfidence bias in perceptual tasks and no overconfidence bias in conceptual tasks thought to involve information stored in long-term memory (e.g., Kvavilashvili & Koutstaal, 2008). Such effects suggest two representations, with moderate calibration suggesting at least some explicit awareness but with overconfidence suggesting a well-defined explicit representation that does not reflect the actual representation used to discriminate and classify stimuli. If a multiple-representation account of speech perception is correct (e.g., Pisoni, 1973) then a confidence report methodology might be capable of assessing these different representations.

An important concern related to the existence of multiple sources of information is how confidence reports are generated. Traditional approaches to confidence processing have been agnostic about the nature of the representations used to make the primary decision and report confidence. Early models of confidence assumed a decisional-locus (e.g., Ferrel & McGooney, 1980; Gigerenzer, Hoffrage, & Kleinbolting, 1991; Pleskac & Busemeyer, 2010) wherein confidence reports are based solely on information used by the primary decision process thereby requiring no additional processing, a post-decisional locus wherein confidence is computed following the primary decision (e.g., Audley, 1960; Vickers & Packer, 1980), or an alterable locus wherein confidence processing can occur during or after the primary decision depending on speed or accuracy stress (Baranski & Petrusic, 1998). For instance, in a study conducted by Baranski and Petrusic (2001) participants were given blocks of trials wherein they were required to simply make a decision or make a decision followed by a post-decisional confidence report. They found that response latencies for the primary decision were significantly longer when confidence was required relative to a no confidence condition indicating an additional set of operations was required to compute confidence. More recently, Schoenherr, Leth-Steensen, and Petrusic (2010) found that information that is nondiagnostic of the primary decision can create variations in confidence reports independently of accuracy. Applied to phonemic

categorization, if acoustic information is available from a perceptual process and phonemic representations are available from the activation of long-term memory representations, then both sources of information should influence confidence reports. Substantial differences in the patterns observed between accuracy and confidence would suggest the existence of acoustic and phonemic representations.

Present Study

In order to assess whether participants have explicit knowledge of acoustic information and phonemic category boundaries, the present study compares confidence reports to performance in ID and discrimination (AX) tasks. In the ID task, awareness of acoustic cues would be evidenced if certainty increases as a function of the distance from the category boundary. This would suggest that the ID function is merely an artifact of the requirement that participants use only two labels to categorize stimuli when they have in fact encoded and stored (temporarily) acoustic information. If, on the other hand, there is no systematic deviation of confidence and ID performance, then it seems reasonable to conclude that participants are only using phonemic information to identify stimuli. Alone, however, ID performance might not be capable of differentiating.

In order to determine whether participants have access to both acoustic and phonemic representations, we replicated Pisoni and Tash's (1974) paradigm wherein response times in the AX task were used to suggest different levels of processing. In addition, the present study also used postdecisional confidence reports. Again, deviations between performance and confidence reports would suggest that two representations are used to classify and discriminate speech sounds. If participants only perceive speech sounds as exemplars of discrete phonemic categories, then they should be reasonably well-calibrated on a trial-to-trial basis and exhibit little or no over-/underconfidence bias. If participants exhibit poor calibration and overconfidence in the AX task, then this suggests that despite the availability of acoustic properties within the implicit system the explicit representation is phonemic. More specifically, such a finding suggests that participants have an explicit representation of the phonemic category but also maintain graded acoustic information from stimuli along a continuum. Given the intuitive saliency of the phoneme, we assume that category boundaries are an explicit representation but that some acoustic information must remain available (e.g., Pisoni, 1973).

Experiment

The goals of this experiment were threefold. First, we sought to validate the use of subjective confidence reports in a speech perception task by comparing a confidence and no confidence condition. Second, using subjective calibration measures, we examined whether participants had explicit awareness of the phonemic category boundary. Third, we examined whether this awareness was task-dependent by using both ID and AX tasks.

Participants

Listeners were 15 Carleton University students who received course credit for their participation. All participants reported normal hearing and no speech pathologies.

Materials

Using the paradigm developed by Pisoni and Tash (1974) participants were presented with /b/ and /p/ stimuli that varied along the VOT continuum. Seven speech stimuli corresponding to 0 to 60 ms VOT, originally synthesized by Lisker and Abramson (1967), were obtained from the Haskins Laboratories website (HL, 2011). The sounds were originally recorded on reel-to-reel tape and later converted into AIFF format. Stimuli were pre-processed using a DC offset correction to eliminate high frequency noises present in the AIFF versions and converted into WAV files. These stimuli were used in both the ID and AX tasks.

Procedure

Trials in the ID task had one or two components depending upon block. In both blocks of trials participants reported whether the stimulus was a /b/ or /p/ using the 'V' or 'N' key, respectively. For one block participants also rated the confidence they had in their ID responses using a 6-point scale, with 50% representing a guess and 100% representing certainty. Participants completed a total of 180 trials in each block of the ID task.

Trials in the AX task also had one or two components depending on block. In both blocks of trials participants decided whether two stimuli separated by a 250 ms ISI were the same or different, using the 'D' and 'K' key, respectively. Replicating Pisoni and Tash (1974), stimulus pairs differed in either 0-, 2-, 4- or 6-steps and were either selected from the same phonemic category or different phonemic categories. Three replications of the eight within-category comparisons and four replications of the six between-category comparisons were presented in a block of 48 trials. For one block of AX trials participants also provided a confidence rating of their AX decision using the same scale described above.

Half of the participants performed the ID task first whereas the other half performed the AX task first. Half of the blocks of trials required participants to provide confidence reports whereas the other half only required participants to complete the ID and AX tasks alone. Presentation of confidence and no confidence blocks was counterbalanced as were the responses keys for the AX task. The experiment required approximately 30 minutes to complete. Stimuli were presented via headphones using PsychoPy software (Peirce, 2007).

Results

In the following analyses we use two sets of assumptions. Following Pisoni and Tash (1974), we assume that stimuli 1-3 are assigned to the /ba/ category whereas stimuli 4-7 are assigned to the /pa/ category. From this we derive measures of accuracy. In the AX task we additionally assume that accuracy is determined by acoustic properties when making paired comparisons.

Calibration was computed by obtaining the average differences between proportion correct for each confidence category with calibration scores range from 0.0 (perfect calibration) and 1.0 (perfect miscalibration). Notably, calibration scores above 0.10 are rare. Under-/Overconfidence was computed by taking the difference between mean confidence and mean accuracy for each condition, with positive values representing overconfidence and negative values representing underconfidence (for further details see, Baranski & Petrusic, 1994).

For all ID and AX analyses, repeated measures ANOVAs were conducted using dependent variables for the primary decision and confidence reports. Greenhouse Geisser adjusted values are reported with unadjusted degrees of freedom. Bonferroni pairwise comparisons were also performed as a post-hoc test.

Identification Task

Proportion Identification. Figure 1 shows the mean response frequency for each category label in the ID task in the confidence report condition. Participants clearly identified two discrete categories for /ba/ and /pa/, respectively, with a category boundary situated between +20 and +30 ms VOT. This pattern replicates the findings obtained by Pisoni and Tash (1974) as well as other studies (e.g., Experiment 1 in McMurray et al., 2003).

The proportion of correct ID responses was analyzed for each VOT stimulus and whether a confidence report was provided or not. The only significant finding observed was the location of the stimuli along the VOT continuum, $F(6,84) = 6.394$, $MSE = .019$, $p = .02$, $\eta^2 = .314$. The absence of a main effect or interaction of confidence reports is important as it suggests that the addition of confidence reports did not significantly affect ID performance thereby permitting a straightforward interpretation of the remaining results.

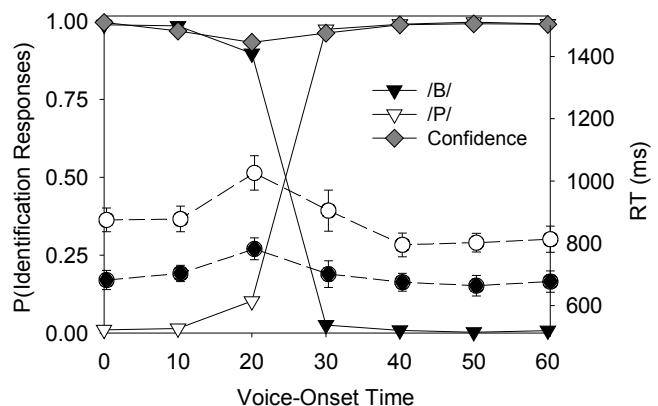


Figure 1. Mean identification functions, response times for confidence (unfilled circles) and no confidence (filled circles) conditions and mean confidence across VOT continuum. Identification function uses performance in confidence condition to allow comparison with mean confidence.

Mean Confidence. Figure 1 also demonstrates the effect of confidence measures. Like ID accuracy, we found that subjective confidence varied along the VOT continuum, $F(1,14) = 6.55$, $MSE = 44.11$, $p = .008$, $\eta^2 = .319$. Pairwise comparisons revealed that this effect arose from the difference in confidence between stimuli located at 20 and 30 ms VOT ($p = .035$), which corresponds to the stimuli adjacent to the category boundary.

Calibration Indices. An analysis of subjective calibration revealed only a marginally significant difference across the VOT continuum, $F(6,84) = 3.401$, $MSE = .013$, $p = .085$, $\eta^2 = .195$. This suggests that the greatest difference between subjective awareness and performance occurs for the 20 ms VOT stimulus. Our comparison of over/underconfidence bias did not reveal any significant effects, $F(6,84) = 1.948$, $MSE = .035$, $p = .183$, $\eta^2 = .122$. Together, these findings suggest that participants are only explicitly aware of the phonemic representation.

Table 1. Mean dependent measures for “Same” responses to Within-Category Paired Comparisons

| | /ba/ | | /pa/ | |
|---------|--------|--------|--------|--------|
| | A-A | A-a | A-A | A-a |
| Mean RT | 756 ms | 818 ms | 693 ms | 737 ms |
| % Error | 4.4 | 11.1 | 1.1 | 1.1 |
| Conf | 97.44 | 98.67 | 99.67 | 99.56 |

Identification Response Time. Analysis of our RT data also replicated Pisoni and Tash’s (1974) findings (see Figure 1). Specifically, we observed significant changes in RT across the VOT continuum, $F(6,84) = 8.323$, $MSE = .030$, $p < .001$, $\eta^2 = .373$. We found that responses to the stimulus located at 20 ms VOT were longer than those for stimuli at 0, 10, 50, and 60 ms VOT. Moreover, replicating Baranski and Petrusic (2001), we also observed a significant main effect of confidence condition, $F(6,66) = 4.701$, $MSE = .041$, $p = .021$, $\eta^2 = .572$. Participants took longer to identify stimuli when confidence reports were required ($M = 871$ ms) relative to the no confidence condition ($M = 698$ ms).

Discrimination Task

Proportion “Same” Responses. AX responses were assessed based on a category criterion such that ‘same’ responses were considered correct for within category comparisons and incorrect for between category comparisons. Table 1 provides a point of comparison with Pisoni and Tash (1974), wherein AA and Aa represent, acoustically and phonemically identical stimuli, respectively. Table 2 provides mean dependent measures for within (AA, Aa) and between (AB through AB”) for 2-step through 6-step, respectively) category comparisons.

Using a category criterion, we observed an interaction between phoneme category (/ba/ v. /pa/) and the comparison type (within v. between), $F(1,13) = 13.421$, $MSE = .004$, $p = .003$, $\eta^2 = .508$. Participants were far more accurate in making within-category comparisons from the /pa/ category relative to all others (see Table 2).

Table 2. Mean dependent measures for correct and incorrect responses for within- and between-category comparisons collapsed across confidence condition.

| | /ba/ | | /pa/ | |
|---------|--------|--------|--------|--------|
| | AA/Aa | All AB | AA/Aa | All AB |
| Mean RT | 753 ms | 809 ms | 693 ms | 734 ms |
| % Error | 7.6 | 5.6 | 1.1 | 7.2 |
| % Conf | 96.6 | 96.4 | 99.6 | 97.3 |

Mean Confidence. Our analysis of mean confidence revealed a pattern similar to that of the accuracy analysis. We found an a marginally significant interaction of phoneme category and comparison type, $F(1,13) = 4.589$, $MSE = 3.8393$, $p = .052$, $\eta^2 = .261$. The only significant effect was for phonemic category, $F(1,13) = 5.895$, $MSE = 9.627$, $p = .030$, $\eta^2 = .312$. Participants expressed more confidence when responding to stimuli from the /pa/ category ($M = 98.456$) relative to those from the /ba/ category ($M = 96.507$). Taken together with the results of the ID task, this suggests that the representation of the /pa/ category is more well-defined than the /ba/ category for these VOT stimuli.

Calibration Indices. As with the mean confidence analysis, the interaction of phoneme category and comparison type was only marginally significant when using category coding, $F(1,13) = 3.613$, $MSE = .0004$, $p = .080$, $\eta^2 = .217$. However, when responses are scored with acoustic coding we find significant miscalibration, $F(4,52) = 776.8$, $MSE = .019$, $p < .001$, $\eta^2 = .984$. Like the ID task, we did not observe any overconfidence bias in the AX task when category coding of accuracy was used, all $F_s < 2.5$. Again this suggests that participants access a phonemic representation to make their confidence decision. Confirming this, significant overconfidence was observed when acoustic coding of accuracy was used to compute overconfidence, $F(4,52) = 1709$, $MSE = .007$, $p < .001$, $\eta^2 = .992$. Post-hoc paired comparisons on these means (see Table 3) revealed that AA pairs differed from all other pairs ($p < .001$) but no other differences were observed.

Table 3. Mean calibration and overconfidence for comparison type using acoustic coding for response accuracy.

| | Comparison Type | | | | |
|-----|-----------------|------|------|------|------|
| | AA | Aa | AB | AB’ | AB” |
| OU | .013 | .976 | .950 | .962 | .994 |
| CAL | .008 | .959 | .913 | .936 | .988 |

Discrimination Response Time. Following Pisoni and Tash (1974), response latencies across comparison pairs were also analyzed. Only correct response latencies were analyzed (i.e., “Same” responses for within-category comparisons and “Different” responses for between-category comparisons). Replicating their findings, the type of comparison affected response latencies, $F(4,52) = 5.976$, $MSE = .088$, $p = .007$, $\eta^2 = .315$. Importantly, an interaction was also observed between the confidence condition and comparison type, $F(1,13) = 9.072$, $MSE = .029$, $p = .01$, $\eta^2 = .414$.

= .315. As Figure 2 indicates, participants were fastest when responding to acoustically similar stimuli and slowest to compare stimuli between categories separated by small steps along the VOT continuum. The additional requirement of confidence increased response latencies for acoustically dissimilar pairs.

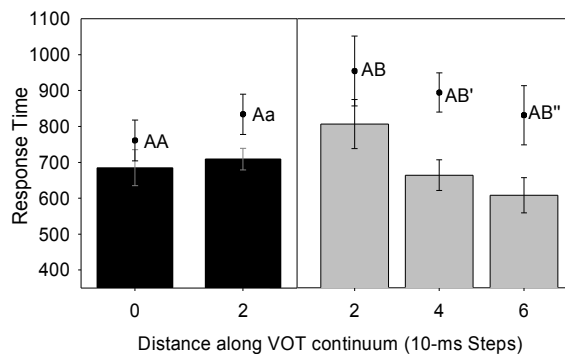


Figure 2. Discrimination response time for paired-comparisons within and between phonemic categories for no confidence (bars) and confidence (points) responses. Error bars represent standard error of the mean.

As in the ID task, we also observed a main effect of confidence condition, $F(1,13) = 4.92$, $MSE = .352$, $p = .045$, $\eta^2 = .275$. More time was required to discriminate stimuli when confidence was required ($M = 861$ ms) relative to the no confidence condition ($M = 696$ ms)¹.

Discussion

Results generally replicated those observed by Pisoni and Tash (1974): participants perceived two distinct phonemic categories along the VOT continuum. In addition, we also found that stimuli could be discriminated with greater speed and accuracy when they were selected from two contrasting phoneme categories rather than within a category. However, although our results suggest participants represented stimuli as phonemic categories, the /ba/ category was not as well defined as the /pa/ category. This was evidenced in a shallower portion of slope between stimuli at 10 and 20 ms VOT in the ID function and the reduced accuracy and confidence when comparing acoustically dissimilar stimuli within that category.

Of equal importance, the experiment also replicated findings in the confidence processing literature: participants were faster when they made ID and AX decisions alone compared to when they were additionally required to report confidence post-decisionally (e.g., Baranski & Petrusic, 2001). Such a finding suggests that confidence processing requires a secondary set of operations to generate a confidence report. Importantly, however, the requirement of

a confidence report did not appear to adversely affect ID or AX performance.

Confidence performance complimented ID and AX performance: participants expressed the most uncertainty in the /ba/ category, and more specifically in the stimulus at the category boundary (20 ms VOT). An absence of overconfidence bias and excellent calibration in the category analysis of both ID and AX tasks suggests that participants' explicit knowledge of phoneme categories guides their classification. Supporting this interpretation, when we reanalyze discrimination accuracy in terms of acoustic properties we found that mean indices of overconfidence and miscalibration were at ceiling for all stimuli other than AA pairs, indicating an inability to discriminate acoustically dissimilar pairs within the same phonemic category.

Conclusions

The present study replicated findings in both the speech perception and confidence processing literatures. Participants' responses indicated categorical perception of acoustically dissimilar stimuli along a voicing continuum and confidence reports required additional time to process. Moreover, confidence reports revealed that participants do not appear to be aware of acoustic information used to activate phoneme representations under the presentation conditions. Although the present study used only one ISI, follow-up studies will vary ISI in an AX task to further differentiate subjective awareness from performance. Moreover, phoneme categories that participants might have less familiarity with (e.g., Pisoni et al., 1982) should demonstrate larger differences in overconfidence and calibration. In short, our findings suggest that confidence reports can be used along with other measures (see also McMurray et al., 2003; Miller & Volaitis, 1989) to assess metalinguistic awareness in the context of speech perception.

Several caveats remain. First, VOT represents one among many physical cues that have been implicated in speech perception. In as much as the processing of speech in a natural environment requires multiple cues, the findings of the present study might be limited to this continuum. One possibility is that with a greater number of cues subjective certainty might increase. This concern about the limits of using synthesized speech has been a recurrent theme in speech perception research (Raphael, 2005). Second, space limits prevent inclusion of an analysis of individual differences, such as working memory capacity and individual ID functions. Finally, studies will need to assess whether these findings generalize to non-speech sounds that share similar properties such as the relative onset of two tones (Pisoni, 1977) or whether overconfidence is limited to only speech sounds. Despite these caveats, the results of the current work suggest that the application of a confidence report methodology holds considerable promise in clarifying the nature of the representations used for speech perception and how these representations are accessed. Calibration can be used to assess whether one or multiple acoustic cues are

¹ An analysis of all (correct and incorrect) responses also revealed the effect of comparison type, $F(4,52) = 7.729$, $MSE = .037$, $p < .001$, $\eta^2 = .373$, and the requirement of confidence, $F(1,13) = 5.07$, $MSE = .247$, $p = .042$, $\eta^2 = .281$. Post hoc paired comparisons revealed that AA differed from both Aa and AB (all $ps < .035$).

used whereas under-/overconfidence suggests the extent to which phonemic and acoustic information is available to listeners.

References

- Baranski, J. V., & Petrusic, W. M. (1994). The calibration and resolution of confidence in perceptual judgements. *Perception & Psychophysics*, *55*, 412-428.
- Baranski, J. V., & Petrusic, W. M. (1998). Probing the locus of confidence judgments: Experiments on the time to determine confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 929-945.
- Baranski, J. V., & Petrusic, W. M. (2001). Testing the architectures of the decision-confidence relation. *Canadian Journal of Experimental Psychology*, *55*, 195-206.
- Davis, M.H., Johnsruide, I.S., Hervais-Adelman, A., Taylor, K. & McGettigan, C.M. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, *134*, 222-241.
- Dienes, Z., & Berry, D. (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bulletin & Review*, *4*, 3-23.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *98*, 506-528.
- Haskins Laboratories (2011). Abramson/Lisker VOT Stimuli. Retrieved 01/12/2011. From <http://www.haskins.yale.edu/featured/demo-liskabram/index.html> /.
- Kvidera, S., & Koustaal, W. (2008). Confidence and decision type under matched stimulus conditions: overconfidence in perceptual but not conceptual decisions. *Journal of Behavioral Decision Making*, *21*, 253-281.
- Liberman, A.M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*, 358-368.
- Lisker, L., & Abramson, A. S. (1967). The voicing dimension: Some experiments in comparative phonetics. *Proceedings of the 6th International Congress of Phonetic Sciences*. Prague: Academia.
- McMurray, B., Tanenhaus, M. K., Aslin, R. N., & Spivey, M. J. (2003). Probabilistic constraint satisfaction at the lexical/phonetic interface: Evidence for gradient effects of within-category VOT on lexical access. *Journal of Psycholinguistic Research*, *32*, 77-97.
- Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, *46*, 505-512.
- Peirce, J. W. (2007) PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*, 8-13.
- Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, *13*, 253-260.
- Pisoni, D. B. (1977) Identification and discrimination of the relative onset of two component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America*, *67*, 1352-1361.
- Pisoni, D. B., & Tash, J. B. (1974) Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, *15*, 285-290.
- Raphael, L. J. (2005). Acoustic cues to the perception of segmental phonemes. In D.B. Pisoni & R.E. Remez (Eds.), *The handbook of speech perception* (pp. 182-206). Oxford: Blackwell.
- Remez, R.E., Rubin, P.E., Pisoni, D.B., Carrell, T.D. (1981). Speech perception without traditional speech cues. *Science*, *212*, 947-949.
- Vickers, D., & Packer, J. S. (1982). Effects of alternating set for speed or accuracy on response time, accuracy, and confidence in a unidimensional discrimination task. *Acta Psychologica*, *50*, 179-197.
- Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three-factors in speech perception. *Perception & Psychophysics*, *37*, 35-44.