

Associative Cognitive Function of Language Found in the Persistence of Motherese

Branton B. Baird^{ID}, John Bryan^{ID}, Patricia D'Urso^{ID}, Carol Pernsteiner^{ID}

Athens State University, Athens, USA

Email: branton.baird@athens.edu, john.bryan@my.gcu.edu, pat.durso@my.gcu.edu, cpernsteiner@my.gcu.edu

How to cite this paper: Baird, B. B., Bryan, J., D'Urso, P., & Pernsteiner, C. (2025). Associative Cognitive Function of Language Found in the Persistence of Motherese.

Psychology, 16, 1115-1128.

<https://doi.org/10.4236/psych.2025.1610064>

Received: August 19, 2025

Accepted: October 8, 2025

Published: October 11, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This study investigates the persistence of maternal vocative expression (VE) effects on adult auditory processing. Using a quasi-experimental, one-way repeated measures ANOVA, the research examines reaction time (RT) differences among adult women responding to their names spoken by their mothers, unfamiliar women, and unfamiliar men. Findings reveal a statistically significant reaction time advantage for maternal voices, supporting Usage-Based Theory (UBT) over the Nativist perspective, which proposes that language capacity (LC) is dissociated from other cognitive systems. Results of a one-way repeated measures ANOVA under three conditions, $F(2, 96) = 52.88$, $p < .001$, $\eta^2 = .52$, suggest that mere exposure to highly frequent, familiar, and emotionally salient stimuli shapes long-term cognitive processing patterns. Implications extend to models of language acquisition, auditory processing, and implicit learning. Learning any variety of motherese, or even the persistence of motherese, is a feature of mere exposure heightened by frequency and salience.

Keywords

Mere Exposure, Motherese, Word Perception, Reaction Time

1. Introduction

Conventional expressions are formulaic sequences used in specific pragmatic contexts, such as in uncertainty, “I don’t know,” hedges, “well, you see, what had happened was...,” and attention grabbing with the use of vocatives, “Hey, John!” or “John” (Bardovi-Harlig & Su, 2018; Culpeper, 2010). The use of a participant’s name to call their attention is called a vocative expression (VE). Proper names have unique implicatures, including them in the category of vocative expressions (Hirata, 2023). The well-known Cocktail Party Effect, first studied by E.C. Cherry in 1953, persists in the field as a well-defined phenomenon, but not in the context

of language learning or development (Jia et al., 2023). Vocative expressions (VEs) provide an appellative category as pragmatic functions (Glušac & Mikić Čolić, 2017). These conventional expressions serve as pragmatic guides to language acquisition in use, as they are not syntactically prominent (Bardovi-Harlig & Su, 2018). In this study, the salience of a mother's voice is evaluated by the reaction times of their adult daughters to establish if familiarity of such cues influences language acquisition at every level, from the role of initial primary caregiver into adulthood (Holmes & Johnsrude, 2023; Mrkva & Van Boven, 2020).

The nature of language acquisition is investigated by cognitive science, contrasting Nativist theories' innate language faculty with Usage-Based Theory (UBT), which emphasizes the role of environmental input (Chomsky, 2017; O'Madagain & Tomasello, 2021; Pearl, 2021). Nativists claim children possess an inborn capacity for language that is activated by minimal input (Yang et al., 2017). In contrast, UBT posits that language is learned through exposure to frequent, salient, and contextually rich linguistic stimuli (Tomasello, 2003). The Nativist approach is necessary but insufficient as a stand-alone theory of language acquisition. Nativists propose that languages are acquired, not learned, and in that respect, language users do not need correction or punishment/reinforcement from language providers such as their own mothers.

Early linguistic exposure, Infant-Directed Speech (IDS) or "motherese," is characterized by exaggerated prosody, simplified syntax, and heightened emotional tone (Fernald, 1989). IDS is known to capture infant attention, enhance language segmentation, and facilitate grammatical learning (Wu et al., 2023). The potential for these exposure effects to persist into adulthood remains underexplored. Some claim that IDS does not exist in every language culture (Pinker, 1995). In many cultures, mothers do not interact with their children in the same way as do many other societies, and this lack of IDS is considered a point in favor of the argument against motherese. However, in that society, children are not held in the swaddled, eye-gazed arms of their mothers as is typical of many cultures of European descent. In the referenced cultures, children are held with their backs to the mother's chest, establishing joint attention through joint intention without eye gaze, but through interaction (Falk, 2009). Tomasello (2021) proposes that joint attention established in the zone of proximal development facilitates the acquisition of linguistic conventions (Vygotsky, 1978). In this respect, infants are not limited to interlocution with only one person but engage the larger community. Motherese in such a scenario would be shared by more than merely the interactions with the mother and the child, but also through the interactions with the child and the rest of the language culture (Falk, 2009). In this scenario, the "mamanèse," that Pinker detracts from, should be considered "culturese" (Pinker, 1995, 2002). The prosodic nuances of language, which provide clues to phonology, morphology, syntax, and pragmatics, are culturally biased, not merely presented by the mother. As early as two days old, infants demonstrate a discrimination between their mother's native language and a nonnative language, especially if it is produced by someone

who is not their mother (Moon et al., 1993).

This current study shows the enduring influence of motherese through adult women's responses to vocative expressions (VEs)—their own names spoken by their mothers, unfamiliar women, and unfamiliar men. We theorized that participants would respond more quickly to maternal VEs, reflecting the influence of mere exposure to frequent, familiar, and emotionally salient stimuli. Consistent and salient linguistic environments should engender language acquisition, establishing motherese as critical input.

2. Literature Review

Motherese and Early Language Acquisition

Motherese, or Infant-Directed Speech (IDS), is a widely observed form of speech characterized by exaggerated intonation, slower tempo, and simplified syntax (Faulk, 2009). IDS enhances infant language learning (Fernald, 1989) and even helps first language learners determine where to place narrowly focused parts of speech, such as articles, nouns, or verbs (Gerken & McIntosh, 1993). IDS has been shown to promote attention, facilitate word segmentation, and support grammatical acquisition (Wu et al., 2023). Neonates are born with a predisposition to parse their native language and their mother's voice when measured as early as two days (Moon et al., 1993). This early exposure represents a cognitive function of mere exposure in utero.

Nativism versus Usage-Based Theory: Implicit Learning and Familiarity Effects

Nativist Universal Grammar theorists propose that linguistic input is insufficient for language learning without a language capacity (LC) (Chomsky, 2017). UBT theorists propose that language is acquired through repeated exposure to meaningful interactions (O'Madagain & Tomasello, 2021). Studies show that frequency, salience, and emotional engagement enhance memory and cognitive processing, supporting UBT (Domingo et al., 2018; Mrkva & Van Boven, 2020).

Research indicates that implicit learning, whereby individuals acquire patterns without conscious awareness, is a foundational mechanism of language acquisition (Arnon, 2019). Familiarity further enhances processing efficiency, with studies showing that familiar voices are recognized and processed faster than unfamiliar ones (Holmes et al., 2021). Considering our mothers are the first to provide spoken stimuli to the young, we considered testing a mother's speech over that of a father's speech, as it is more likely to find a pattern associated with language learning. In studies regarding auditory stimuli, female participants were shown to be more sensitive to speech, especially in pragmatic contexts influenced by intonation, than males (Rao, 2015). Infants tested on electronic simulations of human voices responded more to female-sounding artificial voices than to their artificial male-sounding counterparts (Gerken & McIntosh, 1993). Recent studies in auditory stimuli reveal that women respond faster and more accurately than men when exposed to higher frequency (Hz) stimuli (Aloufi et al., 2023). Given these con-

siderations, women were used in this study to ensure more reliable and valid reaction times. These results may not be as generalizable to the male population, but considering the robust results found, further study should reveal some equal to this study. Furthermore, no cross-cultural samples beyond two bilinguals were included in this study. All other participants were monolinguals from the U.S., and the two bilinguals were mother-daughter dyads. Future studies will isolate men's and women's reactions to their mothers' VEs and include more cross-cultural dyads for the sake of comparison.

Auditory Salience and Memory

Salience, defined as the prominence of a stimulus within a sensory context, is critical in cognitive processing. IDS is inherently salient due to its exaggerated prosody, making it more likely to be encoded in long-term memory (Gaskins et al., 2023). Researchers have found that F0 and vocal tract length provide no significant difference in intelligibility or familiarity responses to auditory stimuli (Holmes & Johnsrude, 2023). An investigation into highly familiar and highly conventional expressions, such as the VE, could yield enlightening results on the nature of language acquisition and speech perception. The more a stimulus is perceived, the more that stimulus is reported as being "liked" (Mrkva & Van Boven, 2020). Therefore, a mother's voice should provide a sensitized perceptual experience to which a highly salient reaction could be found. Since responses to a mother's voice in neonates are well established, the question of whether this salience continues to influence adult cognitive processing is central to the present study. Relative exposure to stimuli has a positive correlative effect on the salience or intensity of a stimulus.

3. Methodology

Participants

A total of 49 adult women aged 18 - 45 ($M = 28.4$, $SD = 6.2$) were recruited. All participants reported normal hearing and native English proficiency, were from the southeastern United States, and maintained regular ongoing contact with their biological mothers to ensure voice familiarity. The mothers of the 49 participants also participated by recording their speech samples in a recording studio in the southeastern United States. Five men of a similar mean age to the mothers recorded themselves speaking the names of all the participants and 30 buffer names.

Materials and Design

The study employed a quasi-experimental, one-way repeated measures design. Participants listened to pre-recorded vocative expressions (a participant's name) (VEs) spoken by three speakers: their own mother, an unfamiliar woman (the mothers of non-active participants), and unfamiliar men, five times each. Participants also heard filler VEs that were not their own names 30 times. Audio recordings were standardized for volume and duration initially through PRAAT (Boersma & Weenink, 2018).

Procedure

Participants completed a reaction time task using a Prime-Probe Reaction Time Instrument (PPRTI) (Mondor & Leboe, 2008). The PPRTI was used without modification of the prime tone of 361 Hz. Participants were instructed to press a response key as quickly as possible upon hearing their name. Each trial was marked by the 361 Hz tone, then a randomized participant's name was presented as specified in previous studies using this PPRTI (Mondor & Leboe, 2008; Schöpper & Frings, 2023). Prior research used two tones to determine a functional difference in the priming effect; these were 361 Hz and 712 Hz (Mondor & Leboe, 2008). No effect for the introduction of this tone was investigated, as it was not contrasted with other tones (Mondor & Leboe, 2008; Schöpper & Frings, 2023). The 361 Hz tone was only used as a benchmark separating trials for the reaction time (RT) participant's benefit. Each participant heard VEs 15 times; 5 times by their mother's voice, 5 times from an unfamiliar woman's voice, and 5 times from an unfamiliar man's voice. Each prompt was buffered by 30 other voices saying VEs that were not of the active participant. Prompts were randomized to account for sequence error by *Inquisit X* (2023). The duration range of the verbal stimuli was unaltered from the phonetic requirements to express the names of the RT participants, except for manually matching the onset of the speech sample with the pitch increase noted by PRAAT (Boersma & Weenink, 2018). Upon hearing the prime tone at 316 Hz for 500 ms before the probe was presented.

The VE probe was presented for a term of 3000 ms for a response. Reactions to auditory stimuli improve up to 2000 ms, allowing more time for the participants to respond without fatigue (Holmes et al., 2018). Loudness was set at 72dB in PRAAT for each prompt (Mondor & Leboe, 2008; Styler, 2023; Zarate et al., 2015). Recent studies in auditory and visual processing established thresholds for fatigue effects as improving while time increases toward 2000 ms (Holmes et al., 2018). Previous auditory studies far exceeded the limits of this study's demand on participants (Domingo et al., 2018; Holmes et al., 2018). Further, the structure of the study yielded highly reliable scores against sequence effects by randomizing trial structure established by *Inquisit X* (2023). Each participant's set of trials was uniquely patterned for their name produced by unknown female speakers, their mother's voice, and the voices of five men. This structure provided for counterbalancing for each participant (Hsu et al., 2023).

Since only one tone was used, a priming effect was not anticipated. Measuring the differences between speakers in the limited samples of only one word was the aim of this study. Spectral matching was not equalized for the sake of maintaining some of the subjective experience of each mother's naturalistic voice. Prompt recordings were made in a soundproof recording studio to counter any possible acoustic confounds (Jia et al., 2023). Therefore, the spectral features were left unaltered to appreciate the differences between the speakers (Feng & Oxenham, 2018). Had the spectral features been equalized, the VE prompts may have been less discriminable. The prime in this study was only used to mark different trials for the participants. No baseline was collected, which could have possibly primed

the participants against the VEs.

If the participant's name was Debra, under the buffer condition, of the 30 buffer prompts, they did not hear Debra spoken. Instead, they heard names from people who were not their mothers not saying their names. Reaction times were recorded in milliseconds by a computer program via *Inquisit X* (2023). Trials were randomized to minimize order effects. Each participant was tested only one time, and their responses were gathered for analysis.

All prompt recording participants and reaction time participants signed informed consent forms and were adults above the age of 18 at the moment of signing. No identifiable data is presented in this manuscript for any participant. Freely given informed consent to participate in the study was obtained from each participant, and to publish the results anonymously after its completion. The Internal Review Board of Grand Canyon University approval was obtained prior to any data collection of any kind. The IRB granted exemption for this study in accordance with the Declaration of Helsinki.

Data Analysis

A one-way repeated measures analysis of variance (ANOVA) was conducted to examine reaction time differences across the three speaker conditions. Post hoc comparisons utilized Tukey Honest Significant Difference adjustments to control for Type I error. One outlier was included in the analysis but did not affect the significance of the results found.

4. Results

Descriptive Statistics

Mean responses for each participant's maternal VE stimuli were faster than those of unfamiliar men and unfamiliar women. Oddly, the results indicate that the unfamiliar men's voices elicited slightly faster, but not significantly faster, reaction times. Mother's voice: $M = 504.47$, $SD = 44$ ms; Unfamiliar woman's voice: $M = 610.96$, $SD = 56$ ms; Unfamiliar man's voice: $M = 589.20$ ms, $SD = 61$ ms.

Inferential Statistics

The one-way repeated measures ANOVA revealed a significant main effect of speaker type on reaction time, $F(2, 96) = 52.88$, $p < .001$, $\eta^2 = .52$, indicating a large effect size.

Post Hoc Comparisons

Significant differences were found in the comparisons with the Mother's VEs compared to the Unfamiliar women's VEs, and the Unfamiliar men's VEs, but not between the unfamiliar women's and unfamiliar men's VE prompts. Mother vs. Unfamiliar Woman: $p < .001$; Mother vs. Unfamiliar Man: $p < .001$; Unfamiliar Woman vs. Unfamiliar Man: $p = .47$ (nonsignificant)

Table 1 presents the ANOVA results of the comparison between the within factors of the mothers vs. the unfamiliar speaker's VEs. The means of the within-subjects factor are presented in **Table 2** and **Figure 1**. Both **Figure 1** and **Table 2** label the participants (MVERTms, UFVERTms, & UMVERTms), which respec-

tively stand for Mother's Vocative Expression Response Times, Unfamiliar Females' Vocative Expression Response Times, and Unfamiliar Males' Vocative Expression Response Times.

Table 1. Repeated measures ANOVA results.

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η^2_p
Within-Subjects						
Within Factor	2	310,224.38	155,112.19	52.88	< .001	0.52
Residuals	96	281,574.29	2,933.07			

Table 2. Means table for within-subject variables.

Variable	<i>M</i>	<i>SD</i>
MVERTms	504.47	100.28
UFVERTms	610.96	101.33
UMVERTms	589.20	89.68

Note. *n* = 49.

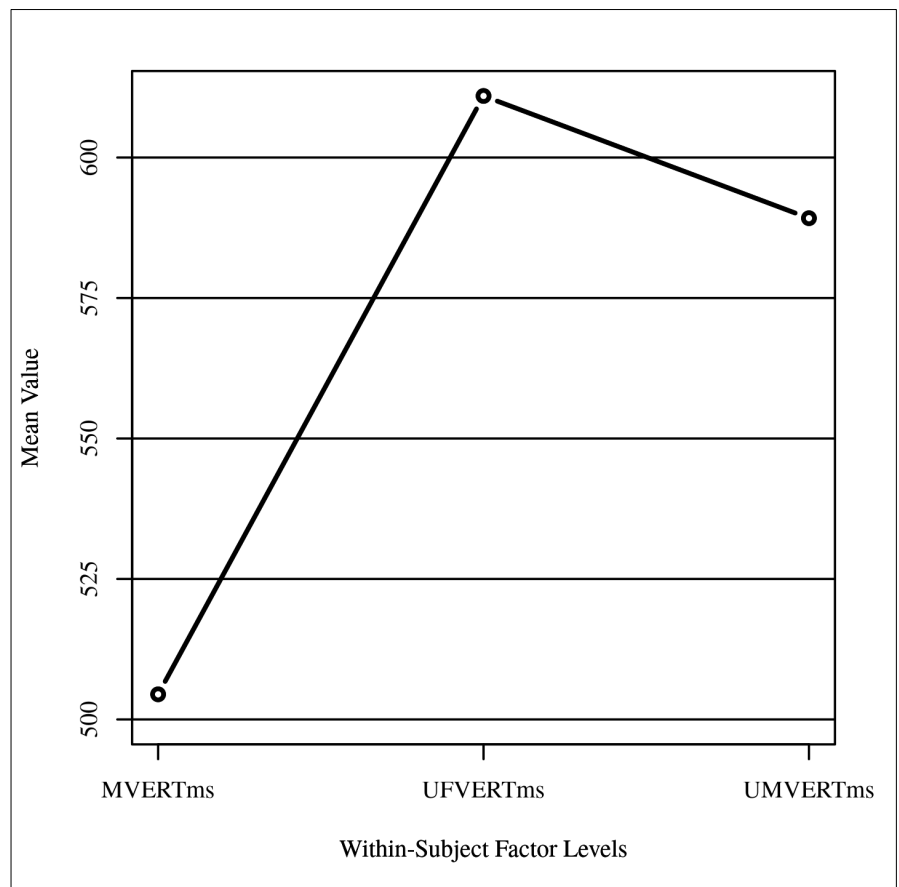


Figure 1. Within-subject variable means.

The pattern in **Figure 1** shows a not statistically significant trend toward unfamiliar male VEs being slightly faster than those of unfamiliar females.

Interpretation

The significant reaction time advantage for maternal voices supports the hypothesis that exposure to frequent, familiar, and emotionally salient stimuli has enduring effects on auditory processing. Effects from a mother's speech, or any speech sample that has about five times more frequency than other stimuli, will result in faster processing of said stimuli.

5. Discussion

Support for Usage-Based Theory

These findings provide strong empirical support for UBT, suggesting that mere exposure to frequent and emotionally salient linguistic stimuli, such as motherese, is sufficient and necessary for language acquisition as a cognitive imprint. When the exposure occurs early in life, as in the case of a mother's voice, the mere exposure effect develops with the linguistic systems. Mother's voice is the primary source of linguistic information by measures of frequency and salience in all of humanity. This effect may be pruned in the case of children adopted at birth, but the mere exposure effect is so prevalent that even the repeated exposure to one stimulus five times provides a significant processing difference compared to stimuli presented only once. A significant familiarity effect between friends and romantic partners has been established for up to a period of 1.5 years. After 1.5 years, cognitive processing is not affected by becoming any easier. This was noted in **Domingo et al. (2018)**, as intelligibility rates increased between familiar speakers for up to 1.5 years. However, in that study, conventional expressions were not standardized for comparison, and therefore this study provides more evidence for language type, which also influences cognitive processing. Another difference between this study and Domingo et al. is that the sentences included a name, but that name was not the VE of the responding participant. It is possible that the significance in this study was influenced not only by the familiarity of the speaker's voice, but also by familiarity with the linguistic signal itself.

Comparisons between Mothers' Voices, Unfamiliar Women's Voices, and Unfamiliar Men's Voices.

It makes more sense that if a mother's voice would elicit faster reaction times, then an unfamiliar woman's voice should elicit faster reaction times than that of unfamiliar men. However, it may be the contrast of the men's voices that elicited slightly faster reaction times from the all-female participant cohort. Questions of sexual preference were not elicited in the recruitment phase of this study and will be considered in future investigations. Also, the subtle comparison between the mother's voices and the voices of the unfamiliar females may show a preference for the mother's voice over that of the unfamiliar women's voices from some unknown gender bias against other women's voices, or even an increased need for inhibition due to the presence of the mother's voices. The men's voices should

have been much easier to discriminate by comparison to those of the unfamiliar women's voices. An argument for novelty could facilitate or affect the processing of stimuli more effectively than frequency or salience in the processing of simple vocative expressions depending on the circumstances of the Umwelt. Further research is required.

Implications for Auditory Processing and Memory

The persistence of faster reaction times to maternal VEs aligns with research on implicit learning and attentional salience, suggesting that familiarity enhances processing efficiency. Intelligibility increases from mere exposure up to a period of 1.5 years in contexts of utterances of the following structure: subject (*Non-participant VE*) + verb + adjective + noun. The non-participant VE is the name of a person who is not associated with the study, or a neutral party's name. However, in the current scenarios, the processing of a mother's voice producing a participant's name was significantly faster, and therefore easier, than any of the other verbal stimuli. If this pattern of learning increases for other conventional expressions, a stronger argument can be made for the influence of the environment on language acquisition instead of relying solely on an undefined language acquisition device and genetic endowment from which language is purported to be generated. Auditory processing and memory are influenced by mere exposure without awareness of the learning. The learning versus acquiring distinction is one of the key discriminatory markers in first and second language acquisition literature. If our speech signals are consistent enough, as consistent as a mother's voice to a 300 billion neuronal infant dependent upon that caregiver for all their needs, then it is possible that language is totally acquired from the environment with only some blueprint DNA structures which create language as an emergent property.

Challenges to Nativism

The significant advantage of maternal voices challenges Nativist assumptions that early exposure merely triggers innate grammatical structures without long-term cognitive effects. The longer the exposure to a mother's speech, the deeper the influence will be toward the child's language acquisition. The sensitivity and habituation of which may be too nuanced for standard statistical models, or for traditional forms of conversation analysis without auditory scene analysis. This is important because current statistical models may not sufficiently predict the character or degree of processing that the brain does globally.

Limitations and Future Directions

The study's sample was limited to adult women, thus also limiting the generalizability to other populations. Future research should explore similar effects in male participants and across diverse cultural contexts. Future research will combine the efforts of [Domingo et al. \(2018\)](#) by adjusting for the familiarity of the speaker and the type of linguistic message delivered by the recording participant. It will also be of interest to investigate the difference between the first reaction times and subsequent reaction times to the vocative expressions or sentences. Further research will investigate the differences in the number of years

the speaker is known compared to a mother's voice, which a neonate has known long before any other speaker. [Domingo et al. \(2018\)](#) pointed out that a period of 1.5 years is the threshold for activation of intelligibility, but their study was not compared against perhaps more salient speakers, such as a reaction time participant's mother.

6. Conclusion

This study demonstrates that mere exposure to maternal vocative expressions produces enduring advantages in adult auditory processing. Since mere exposure to a mother's voice begins in utero, and before language is considered present in the mind of the listener, the mere exposure effect may be so robust that parsing certain kinds of language from one person is facilitated cognitively and neurologically due to the experience. The findings support Usage-Based Theory, highlighting the critical role of frequency, familiarity, and emotional salience in language acquisition. The concrete sounds created by a mother's voice have at least a mere exposure effect during gestation. Because these effects would develop so early in life, the acquisition of language may be biased toward that of the mother's environmental stimuli. Therefore, we can begin to attack the poverty of stimulus argument in that the stimuli have not been adequately measured.

The pattern of reaction times demonstrates that a non-significant difference was found for response times to unfamiliar men's vocative expressions, and response times to unfamiliar women's response times. The difference shows a very slight trend for the unfamiliar men's response times being faster than those of the unfamiliar women. This could be due to a novelty or salience effect of the men's voices over those that were not the participant's mother's voice. The participant's mother's voice should have elicited a mere exposure effect, which is dependent upon frequency and salience ([Mrkva & Van Boven, 2020](#)).

One other failure of measurement in the literature regarding language acquisition is that of linguistic intelligence as a feature of a language capacity separate from other cognitive capacities. [Chomsky \(2017\)](#) has stated that language is a specific cognitive capacity separate from other cognitive processes; it is "dissociated from other cognitive systems." We have tested the language capacity against the cognitive system of mere exposure by eliminating linguistic variables, focusing on vocative expressions. This simple reaction time experiment demonstrates that some kinds of language capacity elicit different responses even when the linguistic structures are the same; therefore, even the linguistic capacity is influenced by mere exposure, frequency of exposure, and Hz frequency at least. Therefore, these two phenomena are not separate cognitive processes, as suggested in the theory of the capacity for language (LC) ([Chomsky, 2017](#)). From the results of this study, an argument can be made that LC is not dissociative from other cognitive systems as has been stated ([Chomsky, 2017](#)). Even genetic roots motivating the emergence of cognitive systems do not stand alone, but rather they overlap hierarchically to create what we conventionally use as communicative language. However, it may be that

language is the natural emergence of the culmination of cognitive processes and their interactions with the sensory-motor system, but not the only method for computation or communication. We suggest that the continua of cognitive processes influence each other, but only language is used in a conventional and systematic fashion between users. This suggests that if other cognitive processes could share joint intention and attention, we could and would use them for computational and communicative purposes as conventionally as we do human language.

Ethics Approval

Research conducted for this study was granted exemption by the internal review board at Grand Canyon University which certifies that this research was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or compatible ethical standards. from review and approved for data collection by federal regulations.

Consent to Participate

All prompt recording participants, and reaction time participants signed informed consent forms and were adults above the age of 18 at the moment of signing. No identifiable data is presented in this manuscript of any participant. Freely given consent to participate in the study was obtained from each participant.

Consent for Publication

We have written consent and permission from each participant to publish from the anonymous data collected during this study.

Availability of Data and Materials

The data is submitted along with the manuscript. All data generated or analyzed during this study are included in this published article and supplementary information files.

Code Availability

The Prime-Probe Reaction Time Instrument was coded by and is available at Inquisit X [Computer software]. Retrieved from <https://www.millisecond.com>.

Author's Contributions

B.B. and J.B. conceived of the presented idea for publication. B.B. and P.D. developed the theory. B.B. and C.P. performed the computations. B.B., J.B. P.D., and C.P. verified the analytical methods. B.B. supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.

Conflicts of Interest/Competing Interests

There were no conflicts of interest or competing interests in the completion of this

manuscript or the submission of this manuscript to the Psych editors of SCIRP.

References

- Aloufi, N., Heinrich, A., Marshall, K., & Kluk, K. (2023). Sex Differences and the Effect of Female Sex Hormones on Auditory Function: A Systematic Review. *Frontiers in Human Neuroscience*, 17, Article 1077409. <https://doi.org/10.3389/fnhum.2023.1077409>
- Arnon, I. (2019). Statistical Learning, Implicit Learning, and First Language Acquisition: A Critical Evaluation of Two Developmental Predictions. *Topics in Cognitive Science*, 11, 504-519. <https://doi.org/10.1111/tops.12428>
- Bardovi-Harlig, K., & Su, Y. (2018). The Acquisition of Conventional Expressions as a Pragmalinguistic Resource in Chinese as a Foreign Language. *The Modern Language Journal*, 102, 732-757. <https://doi.org/10.1111/modl.12517>
- Boersma, P., & Weenink, D. (2018). PRAAT: Doing Phonetics by Computer. [Computer program]. Version 6.0. 37. Retrieved February, 3.
- Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America*, 25, 975-979. <https://doi.org/10.1121/1.1907229>
- Chomsky, N. (2017). The Language Capacity: Architecture and Evolution. *Psychonomic Bulletin & Review*, 24, 200-203. <https://doi.org/10.3758/s13423-016-1078-6>
- Culpeper, J. (2010). Conventionalised Impoliteness Formulae. *Journal of Pragmatics*, 42, 3232-3245. <https://doi.org/10.1016/j.pragma.2010.05.007>
- Domingo, Y., Holmes, E., & Johnsrude, I. S. (2018). The Benefit to Speech Intelligibility of Hearing a Familiar Voice. *Journal of Experimental Psychology: Applied*, 26, 236-247. <https://doi.org/10.1037/xap0000247>
- Falk, D. (2009). *Finding Our Tongues: Mothers, Infants and the Origins of language*. Basic Books.
- Feng, L., & Oxenham, A. J. (2018). Spectral Contrast Effects Produced by Competing Speech Contexts. *Journal of Experimental Psychology: Human Perception and Performance*, 44, 1447-1457. <https://doi.org/10.1037/xhp0000546>
- Fernald, A. (1989). Intonation and Communicative Intent in Mothers' Speech to Infants: Is the Melody the Message? *Child Development*, 60, 1497-1510. <https://doi.org/10.2307/1130938>
- Gaskins, D., Falcone, M., & Rundblad, G. (2023). A Usage-Based Approach to Metaphor Identification and Analysis in Child Speech. *Language and Cognition*, 16, 32-56. <https://doi.org/10.1017/langcog.2023.17>
- Gerken, L., & McIntosh, B. J. (1993). Interplay of Function Morphemes and Prosody in Early Language. *Developmental Psychology*, 29, 448-457. <https://doi.org/10.1037/0012-1649.29.3.448>
- Glušac, M., & Mikić Čolić, A. (2017). Linguistic Functions of the Vocative as a Morphological, Syntactic and Pragmatic-Semantic Category. *Jezikoslovlje*, 18, 447-472.
- Hirata, I. (2023). Implicatures of Proper Name Vocatives in English. *Journal of Pragmatics*, 207, 28-44. <https://doi.org/10.1016/j.pragma.2023.02.001>
- Holmes, E., & Johnsrude, I. S. (2023). Intelligibility Benefit for Familiar Voices Is Not Accompanied by Better Discrimination of Fundamental Frequency or Vocal Tract Length. *Hearing Research*, 429, Article 108704. <https://doi.org/10.1016/j.heares.2023.108704>
- Holmes, E., Kitterick, P. T., & Summerfield, A. Q. (2018). Cueing Listeners to Attend to a Target Talker Progressively Improves Word Report as the Duration of the Cue-Target

- Interval Lengthens to 2,000 ms. *Attention, Perception, & Psychophysics*, 80, 1520-1538. <https://doi.org/10.3758/s13414-018-1531-x>
- Holmes, E., Parr, T., Griffiths, T. D., & Friston, K. J. (2021). Active Inference, Selective Attention, and the Cocktail Party Problem. *Neuroscience & Biobehavioral Reviews*, 131, 1288-1304. <https://doi.org/10.1016/j.neubiorev.2021.09.038>
- Hsu, Y. F., Tu, C. A., Chen, Y., & Liu, H. M. (2023). The Mismatch Negativity to Abstract Relationship of Tone Pairs Is Independent of Attention. *Scientific Reports*, 13, Article No. 9839.
- Inquisit X (2023). *Computer Software*. <https://www.millisecond.com>
- Jia, S., Zhang, T., Zuo, R., & Xu, B. (2023). Explaining Cocktail Party Effect and McGurk Effect with a Spiking Neural Network Improved by Motif-Topology. *Frontiers in Neuroscience*, 17, Article 1132269. <https://doi.org/10.3389/fnins.2023.1132269>
- Mondor, T. A., & Leboe, L. C. (2008). Stimulus and Response Repetition Effects in the Detection of Sounds: Evidence of Obligatory Retrieval and Use of a Prior Event. *Psychological Research*, 72, 183-191. <https://doi.org/10.1007/s00426-006-0095-x>
- Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-Day-Olds Prefer Their Native Language. *Infant Behavior and Development*, 16, 495-500. [https://doi.org/10.1016/0163-6383\(93\)80007-u](https://doi.org/10.1016/0163-6383(93)80007-u)
- Mrkva, K., & Van Boven, L. (2020). Salience Theory of Mere Exposure: Relative Exposure Increases Liking, Extremity, and Emotional Intensity. *Journal of Personality and Social Psychology*, 118, 1118-1145. <https://doi.org/10.1037/pspa0000184>
- O'Madagain, C., & Tomasello, M. (2021). Shared Intentionality, Reason-Giving and the Evolution of Human Culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377, Article 20200320. <https://doi.org/10.1098/rstb.2020.0320>
- Pearl, L. (2021). Poverty of the Stimulus without Tears. *Language Learning and Development*, 18, 415-454. <https://doi.org/10.1080/15475441.2021.1981908>
- Pinker, S. (1995). Language Acquisition. In D. N. Osherson, M. Liberman, & L. R. Gleitman (Eds.), *An Invitation to Cognitive Science* (2nd ed., Vol. 1). MIT Press.
- Pinker, S. (2002). *The Blank Slate: The Denial of Human Nature in Modern Intellectual Life*. Viking Press.
- Rao, R. (2015). Manifestations of /BDG/ in Heritage Speakers of Spanish. *Heritage Language Journal*, 12, 48-74. <https://doi.org/10.46538/hlj.12.1.3>
- Schöpper, L., & Frings, C. (2023). Same, but Different: Binding Effects in Auditory, but Not Visual Detection Performance. *Attention, Perception, & Psychophysics*, 85, 438-451. <https://doi.org/10.3758/s13414-021-02436-5>
- Styler, W. (2023). *Using Praat for Linguistic Research*. University of Colorado at Boulder Phonetics Lab.
- Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press.
- Tomasello, M. (2021). *Becoming Human: A Theory of Ontogeny*. Belknap Press of Harvard University Press.
- Vygotsky, L. S. (1978). *Mind in Society*. Harvard University Press.
- Wu, Y., Taylor, I. M., Chen, H., & Frank, M. C. (2023). Adults Tailor Their Emotional Expressions to Infants through "Emotionese". <https://escholarship.org/uc/item/9vm4j7bp>
- Yang, C., Crain, S., Berwick, R. C., Chomsky, N., & Bolhuis, J. J. (2017). The Growth of Language: Universal Grammar, Experience, and Principles of Computation. *Neurosci-*

ence & Biobehavioral Reviews, 81, 103-119.

<https://doi.org/10.1016/j.neubiorev.2016.12.023>

Zarate, J. M., Tian, X., Woods, K. J., & Poeppel, D. (2015). Multiple Levels of Linguistic and Paralinguistic Features Contribute to Voice Recognition. *Scientific Reports*, 5, Article No. 11475. <https://doi.org/10.1038/srep11475>