



Evaluation de la saillance d'annonces vocales par un paradigme de double-tâche

T. Bouchara et G. Mahé

LIPADE/Université Paris Descartes, 45 rue des saint Pères, 75006 Paris, France
tifanie.bouchara@gmail.com

L'étude présentée ici concerne la **saillance sonore d'annonces vocales**, c'est-à-dire leur capacité à attirer l'attention. Cette étude s'inscrit dans le projet I'City For All du programme européen Ambient Assisted Living qui vise à améliorer les annonces vocales diffusées dans les gares et aéroports pour guider les voyageurs. Nous envisageons la saillance sonore comme un moyen d'améliorer l'intelligibilité des annonces qui émergent alors mieux du bruit de fond environnant. Nous proposons ici une **nouvelle méthode expérimentale** pour évaluer subjectivement la saillance sonore. Cette méthode repose sur un paradigme de double tâche où les participants doivent effectuer deux tâches en parallèle. La première tâche est une tâche d'attention soutenue appelée Paced Auditory Serial Addition Test (PASAT) dans laquelle les participants doivent additionner les deux derniers chiffres d'une séquence numérique. La seconde tâche est une tâche d'identification de parties d'annonces vocales. La saillance des annonces est alors évaluée en comparant les performances de chacune des tâches effectuées seules ou en parallèle. Sachant qu'un son est plus saillant si son niveau sonore est plus élevé, une expérience perceptive incluant 8 participants a permis d'étudier la validité du protocole expérimental, en comparant des annonces vocales de différents niveaux.

1 Introduction

Cette étude fait partie du projet I'City For All qui vise à améliorer l'intelligibilité des annonces de gare et d'aéroport, en particulier pour les personnes âgées souffrant de presbyacousie¹[1]. Une des difficultés provoquées par ce genre de troubles auditifs est qu'elle diminue notre capacité d'attention sélective (aussi nommé *effet cocktail party*) qui permet, dans un environnement bruyant comme les gares, de se concentrer sur une seule source sonore en ignorant le reste des sources présentes. Il est donc plus difficile pour les malentendants et les personnes âgées de focaliser leur attention sur les annonces vocales, ce qui réduit l'intelligibilité perçue de ces annonces. De plus, même si une annonce vocale est très intelligible, c'est-à-dire que l'ensemble du message contenu dans l'annonce est compréhensible, cela ne signifie pas que les voyageurs vont écouter l'annonce. En effet, pendant la diffusion de l'annonce, les usagers peuvent être engagés dans une tâche parallèle (lecture, discussion téléphonique...). Nous cherchons donc à augmenter la saillance des annonces, c'est-à-dire leur capacité à attirer l'attention des voyageurs concernés.

Plusieurs paramètres acoustiques comme le niveau sonore, la fréquence ou le timbre peuvent augmenter la saillance d'une source sonore et la faire ressortir du fond sonore [2, p.35]. Ainsi, d'après [3], des tons modulés temporellement émergent parmi des sons stationnaires tandis que des sons de durée longue ressortent parmi des sons courts (et inversement). Fondés sur cette hypothèse, des modèles perceptifs ont été proposés pour estimer la saillance d'un signal acoustique [4, 5, 6, 7, 8, 9, 2]. Cependant leur validation par une évaluation perceptive reste insuffisante [10] car les protocoles proposés reposent sur des stimuli courts, de laboratoire et aucun n'a cherché à mesurer la saillance d'une phrase entière.

Notre objectif est alors de concevoir un protocole expérimental original pour évaluer le niveau de saillance auditive d'une annonce et déterminer les paramètres acoustiques responsables de la saillance sonore. Nous proposons d'exploiter le paradigme de double tâche, habituellement utilisé pour mesurer l'effort d'écoute [11], et par lequel les participants doivent effectuer deux tâches en parallèle, pour l'adapter à la mesure de l'attraction d'un stimulus sonore. Les parties 3 et 4 de cet article présentent la mise en place et les résultats d'une expérience pilote reprenant ce protocole expérimental. Sachant que le niveau

sonore est un facteur acoustique reconnu pour influencer la saillance auditive, nous avons évalué l'efficacité de notre protocole expérimental en comparant le pouvoir d'attraction d'annonces vocales de différents niveaux.

2 Revue de littérature

2.1 Mesures subjectives de la saillance sonore

Alors que les études sur la vision bénéficient du suivi oculaire pour identifier directement ce qui attire l'oeil, aucun outil équivalent n'est disponible en audition. Ainsi pour évaluer les modèles de saillance sonore, les valeurs de saillance obtenue par mesures objectives sont comparées à des notations subjectives ou à des mesures comportementales obtenues sur des tâches plus ou moins adaptées.

Une des solutions consiste à demander directement aux participants d'évaluer le niveau de saillance de différents sons. Par exemple, dans [12] et [4], les participants devaient écouter deux stimuli sonores, présentés l'un après l'autre, composés chacun d'une ambiance sonore et d'un son spécifique, puis décider dans une tâche à choix alternatif, lequel des deux stimuli était le plus saillant. Cette méthodologie oblige les expérimentateurs à expliciter la notion de saillance aux participants puis à considérer que chaque participant interprète cette définition de la même manière. Pour éviter de définir la saillance, Tsuchida [8] propose lui aussi une expérience de type comparaison par paire, mais demande aux participants d'indiquer « le son le plus intéressant » au lieu du son le plus saillant. Cependant, cela ne résoud pas le problème lié à l'interprétation commune entre les participants.

Comme un son plus saillant émerge du fond sonore, il peut être détecté à des niveaux sonores moindres. La mesure du seuil de détection d'un son est d'ailleurs la mesure la plus fréquemment utilisée pour évaluer la saillance perçue de ce son [10]. Par exemple, dans [4], des sons d'environnements étaient plongés, à différents niveaux sonores, dans des ambiances sonores (réalistes ou de bruit blanc) de niveau sonore fixe. Les sons pour lesquels le seuil de détection était le plus bas étaient considérés comme étant les plus saillants. Cette expérimentation ne tient cependant pas compte du fait qu'un son saillant attire l'attention au détriment des autres sources sonores présentes. C'est ce qu'ont récemment proposé Tordini et ses collègues [9]. Leur protocole est fondé sur l'hypothèse que, dans le cas d'une écoute concurrente entre deux flux audio simultanés, l'auditeur écouterait plus souvent le flux le plus saillant. Malheureusement leur étude

1. perte progressive de l'audition, liée à l'âge, surtout dans les fréquences aiguës.

ne concerne que des stimuli non réalistes constitués de sons très courts.

2.2 Mesure de l'effort d'écoute et paradigme de double tâche

Nous nous intéressons ici principalement au phénomène d'attention divisée pour évaluer si une annonce plus saillante peut attirer l'auditeur, c'est-à-dire le distraire d'une autre tâche, ou encore si son traitement perceptif est suffisamment facilité pour qu'on puisse réaliser les deux tâches en parallèle.

Une méthode commune en psychologie expérimentale pour mesurer l'effort mental provoqué par une tâche est appelé paradigme de double tâche [13]. Cette méthode suppose que nos ressources cognitives sont limitées si bien que lorsqu'elles sont déjà engagées sur une tâche dite primaire, moins de ressources sont disponibles pour résoudre une seconde tâche. On demande alors aux participants d'exercer les deux tâches en parallèle. L'effort mental demandé par la tâche primaire est alors mesuré en comparant les performances obtenue pour cette tâche effectuée seule et celles obtenues lorsque les deux tâches sont effectuées en même temps.

Ce paradigme est particulièrement intéressant dans le cadre du projet I'City For all car il est souvent utilisé pour évaluer l'effet de l'âge et des problèmes auditifs sur l'intelligibilité perçue [14], et de manière plus générale, pour mesurer l'effort d'écoute ([11, p.47-48]). En effet, alors que de nombreux tests d'intelligibilité ne permettent pas toujours de mettre en évidence des différences entre normo-entendants et mal-entendants sur des tâches de reconnaissance de mots, les expériences à double tâche montrent que pour obtenir les mêmes performances, les personnes mal-entendantes allouent juste plus de ressources à cette tâche. Cela les conduit à une fatigue auditive inévitable et à de moindres performances sur tout autre tâche à réaliser de façon concurrente. Les protocoles utilisés reposent sur des procédures similaires d'une étude à l'autre : la tâche primaire concerne l'activité d'écoute (par exemple des tests de reconnaissance de parole dans le silence puis pour différents rapports signal-sur-bruit) ; la tâche secondaire peut alors être dans la même modalité (tâche de mémoire auditive/rappel) ou dans une autre (réaction au changement de la couleur d'une fenêtre par exemple). La priorité peut être donnée à l'une, l'autre ou aucune des deux tâches.

Ce paradigme de double tâche est également très intéressant pour évaluer le niveau de saillance auditive. En effet, les phénomènes perceptifs régis par la saillance ne requièrent pas d'attention volontaire, c'est-à-dire pas d'effort pour être traités. Ainsi, une tâche isolée, impliquant des stimuli sonores saillants, devrait toujours être bien réussie. En revanche, mettre en concurrence cette tâche avec une autre devrait mettre en évidence des variations de saillance via une dégradation des performances sur l'une des deux tâches. Une expérience utilisant un paradigme de double tâche pour évaluer la saillance sonore a d'ailleurs été récemment proposé dans [7]. Les participants devaient réaliser deux tâches auditives en parallèle. La première, à forte charge cognitive, consistait à compter le nombre de notes graves (100 Hz) dans une séquence de 25 notes entre 100 Hz et 200 Hz. La seconde tâche consistait à détecter si un ton modulé était présent ou non parmi quatre tons. La

modulation étant un des facteurs acoustiques de la saillance auditive, cette seconde tâche impliquait moins de ressources cognitives que la tâche primaire. L'étude a confirmé qu'en augmentant l'amplitude des modulations et donc la saillance des sons dans la seconde tâche, les participants étaient capables d'obtenir de meilleures performances sur cette tâche, tout en maintenant le niveau de performances sur la tâche primaire. Cependant, la principale limitation de cette étude reste l'utilisation de stimuli de laboratoires, non écologiques, pour les deux tâches. Pour évaluer l'efficacité des modèles de saillance dans un contexte appliqué à notre projet d'annonces vocales, nous proposons donc de définir un protocole expérimental plus complexe, faisant intervenir des stimuli de paroles.

3 Etude pilote : influence du volume

3.1 Tâches

Conformément au paradigme de double tâche, deux tâches ont été proposées aux participants. Elles devaient être effectuées soit séparément (condition contrôle) soit en même temps (double tâche).

Nous avons choisi une tâche primaire difficile, demandant de nombreuses ressources attentionnelles, de sorte qu'il soit impossible de la réaliser sans porter volontairement son attention sur elle. Sinon, il aurait été possible d'être attentif aussi à la seconde tâche sans que cela ne nuise aux performances de chacune des tâches, ne permettant pas de dégager d'information sur la saillance des stimuli de la seconde tâche.

Ainsi, la Tâche 1 est une tâche dérivée du test des additions en série (*Paced Auditory Serial Addition Test* ou PASAT [15], [16], [17]) couramment utilisé en neuropsychologie pour détecter des troubles cognitifs. Pour cette première tâche, les participants devaient écouter au casque une série de chiffres de 1 à 9, diffusés l'un après l'autre, dans un ordre aléatoire. L'intervalle de temps séparant deux stimuli était fixé à 3 secondes. Le participant avait pour consigne d'indiquer le plus rapidement possible (avant le prochain chiffre), à l'aide du clavier numérique, la somme des deux derniers chiffres entendus, sans tenir compte des chiffres précédents ou du précédent résultat.

La Tâche 2 est une tâche de discrimination parmi deux possibilités. Une série d'annonces du type « Le train a destination de < *DESTINATION* > partira à < *HORAIRE* > quai < *LETTRE* > » était diffusée au casque, dans un ordre aléatoire, toutes les 15s. Avant chaque série d'annonces, une destination cible était spécifiée. Il était demandé au participant de réagir aux annonces concernant la destination cible (et seulement celle-ci) en indiquant au clavier, le plus rapidement possible, la lettre du quai mentionné dans cette annonce. En résumé, il s'agissait de répondre à la question "quai A ou quai B?". Des étiquettes étaient collées sur les touches F et H d'un clavier pour faciliter les réponses avec une seule main.

Les variables dépendantes que nous avons mesurées sont :

- pour la Tâche 1, les temps de réponse, le pourcentage de réponses correctes, la plus longue série de réponses correctes ;

- pour la Tâche 2, les temps de réponse, le pourcentage de réponses correctes, de fausses alarmes (réponse à une annonce ne correspondant pas à la destination cible) et de faux négatifs (pas de réponse à une annonce correspondant à la destination cible).

En plus de comparer les performances obtenues pour la combinaison Tâche 1 + Tâche 2 (T1+T2) avec les performances de la tâche 1 prise seule (T1 seule), nous avons également testé les performances de cette tâche en présence des annonces vocales sans que les participants aient à effectuer la tâche de discrimination du quai (T1+annonces). Cette condition permet de distinguer si la diminution des performances, dans le cas de la double tâche par rapport à la tâche primaire seule, est due simplement à un masquage sonore, qu'il soit énergétique ou informationnel (la présence des annonces masque les chiffres donc il est plus difficile de réaliser la tâche d'addition), ou si il s'agit d'une limite de ressources cognitives que l'on doit partager entre deux tâches.

3.2 Hypothèses

Hypothèse 1 : les ressources cognitives disponibles pour l'écoute sont limitées et doivent être réparties entre les différentes tâches à réaliser. A ce titre, exécuter les deux tâches simultanément devrait conduire à des performances moindres que dans les cas où seule la Tâche 1 d'addition est exécutée.

Hypothèse 2 : augmenter le volume sonore des annonces devrait diminuer la gêne auditive exercées par d'autres sources sonores simultanées. Les performances de la tâche 2 devraient donc être améliorées.

Hypothèse 3 : une annonce de niveau sonore plus élevé est plus saillante. Elle devrait attirer involontairement l'attention. Plus de ressources attentionnelles devraient donc être allouées à la Tâche 2 plutôt qu'à la Tâche 1, conduisant à l'observation de performances diminuées pour la Tâche 1.

3.3 Stimuli

Les stimuli ont été prononcés en français par la même locutrice. Ils ont été enregistrés en 16 bits à une fréquence d'échantillonnage de 44.1 kHz, dans une salle isolée acoustiquement.

Les enregistrements de chiffres entre 1 et 9 ont été extraits d'une collection de stimuli validée dans une précédente expérience sur l'attention auditive [18]. La durée moyenne des stimuli de chiffres est de 456 ms (min 204 ms - max 685 ms). Ils ont été normalisés en niveau RMS.

Les annonces sont construites par concaténation de morceaux pré-enregistrés pour suivre le schéma : « Le train a destination de < DESTINATION > partira à < HORAIRE > quai < LETTRE > ». Nous avons choisi 10 destinations (+ une pour la phase d'apprentissage), 6 horaires (tels qu'aucune valeur ne puisse être confondue avec la série de chiffres de l'autre tâche) ainsi que deux lettres possibles, comme indiqué Tableau 1. Le niveau sonore de ces annonces a été décliné en trois versions : même niveau RMS que les chiffres de la Tâche 1, +6 dB par rapport aux chiffres ou -6 dB par rapport aux chiffres. La normalisation a été effectuée sur chacun des morceaux

d'annonce séparément, avant la concaténation. Au final, les annonces complètes durent entre 5.55 s et 6.59 s (moyenne 6.17 s).

TABLEAU 1 – Listes des destinations, horaires et quais choisis pour former les annonces de la Tâche 2.

*Cette destination n'apparaît que dans la phase d'apprentissage.

Destinations	Horaires	Quais
Paris Saint-Lazare	11h20	A
Marseille Saint-Charles	12h30	B
Angers Saint-Laud	13h40	
Montpellier Saint-Roch	14h50	
Rouen Rive Droite	15h20	
Toulouse Matabiaud	16h30	
Lyon Part-Dieu		
Lille Flandres		
Strasbourg-ville		
Bordeaux Saint-Jean		
Massy TGV *		

3.4 Participants

Huit volontaires (dont 5 femmes ; âge moyen 32 ans ; tous francophones) ont participé à cette étude. Aucun ne présentait de problème auditif connu. Ils ont été testés individuellement, dans une pièce isolée acoustiquement.

3.5 Procédure

Le plan d'expérience suit un design factoriel croisé inter-participants avec pour variables indépendantes la tâche à réaliser (Tâche 1, Tâche 1 bruitée par la présence d'annonces, Tâches 1+2) et le niveau sonore des annonces par rapport aux chiffres (-6 dB, 0 dB, +6 dB). Nous n'avons pas fait passer la tâche 2 seule car nous savions, d'après des tests préliminaires, que cette tâche était extrêmement facile ($\approx 100\%$ de réponses correctes dans toutes les conditions de niveau sonore testées). De plus, pour diminuer le nombre de conditions et ainsi réduire la durée de l'expérience et la fatigue auditive, nous n'avons fait passer qu'une seule série pour la Tâche 1 seule, le niveau sonore des annonces n'intervenant pas dans ce cas. Au final, l'expérience était divisée en 7 séries, de sorte que ni la condition de niveau sur les annonces ni la consigne ne changent au sein d'une série. Les 7 séries étaient donc :

- Tâche 1 seule,
- Tâche 1 + annonces à -6 dB,
- Tâche 1 + annonces à 0 dB,
- Tâche 1 + annonces à +6 dB,
- Tâche 1 + Tâche 2 (annonces à -6 dB),
- Tâche 1 + Tâche 2 (annonces à 0 dB),
- Tâche 1 + Tâche 2 (annonces à +6 dB).

L'ordre des séries était aléatoire. Chaque série durait 3 minutes et était constituée de 61 chiffres séparés de 3 s, tirés aléatoirement de sorte que chaque chiffre apparaisse environ le même nombre de fois. Les séries d'annonces étaient

composées de 12 annonces (6 annonces avec destination cible et 6 annonces avec destination non-cible), séparées de 15 s.

Les participants étaient invités à prendre une pause entre deux séries successives pour diminuer la fatigue auditive.

Afin de vérifier qu'ils avaient bien compris chacune des deux tâches, les participants étaient invités en début d'expérience à réaliser six séries plus courtes (une minute seulement soit 20 chiffres ou 4 annonces), reprenant une série avec la Tâche 1 seule, une série avec la Tâche 2 seule, une série avec la Tâche 1 et les annonces au même niveau (0 dB) et trois séries avec la Tâche 1 et la Tâche 2 en parallèle (-6 dB, 0 dB, +6 dB).

Au total, l'expérience durait 45 minutes par participant.

4 Analyses des résultats

Nous avons analysé les données séparément pour chacune des deux tâches.

4.1 Analyse des résultats de la tâche 1 (PASAT)

Les trois variables dépendantes mesurées pour cette tâche sont le temps de réponse (TR), le taux de réponses correctes ($\%_{Correct}$) et la longueur de la plus longue série de bonnes réponses. Elles sont reportées en figure 1. Puisque le niveau sonore des annonces n'intervenait pas dans la tâche 1 seule, nous avons dupliqué les performances obtenues dans cette condition de façon à analyser les résultats par une analyse de variances (ANOVA) à mesures répétées sur les deux facteurs : présence de la tâche 2 (T1 seul, T1+annonces non traitées, T1+T2) et le niveau des annonces par rapport à celui des chiffres (-6 dB, 0 dB, +6 dB).

4.1.1 Taux de réponses correctes

Le $\%_{Correct}$ est défini comme le nombre de bonnes réponses par participant et par condition. Comme on peut le voir sur la figure 1, le taux de réponses correctes diminue lorsque les deux tâches sont effectuées en parallèle. L'analyse ANOVA confirme que la présence ou non de la tâche 2 affecte significativement les résultats ($F_{2,14} = 27.63, p < 0.001$). En revanche, le niveau sonore n'affecte pas le taux de bonnes réponses ($F_{2,14} = 2.49, p = 0.08$) et il n'y a pas d'interaction entre les deux facteurs ($F_{4,28} = 1.27, p = 0.28$).

Un test post-hoc avec corrections de Bonferroni confirme que le $\%_{Correct}$ diminue dans la condition T1+T2 par rapport aux deux autres conditions où seule la tâche 1 était effectuée ($p's < 0.001$). L'absence de différence significative entre T1 seule et T1+annonces ($p = 0.45$) confirme que le simple fait de bruyé les chiffres par l'ajout d'annonces n'a pas d'influence sur le taux de bonnes réponses.

4.1.2 Longueur des séries de bonnes réponses.

La longueur de la plus longue série de bonnes réponses correspond au plus grand nombre de bonne réponses que peut enchaîner un participant sans faire d'erreur. Nous avons calculé cette valeur pour chaque condition et pour chaque participant. L'analyse de cette variable a révélé un effet significatif de la présence de la tâche 2 ($F_{2,14} = 3.9, p = 0.02$). Aucune influence du niveau

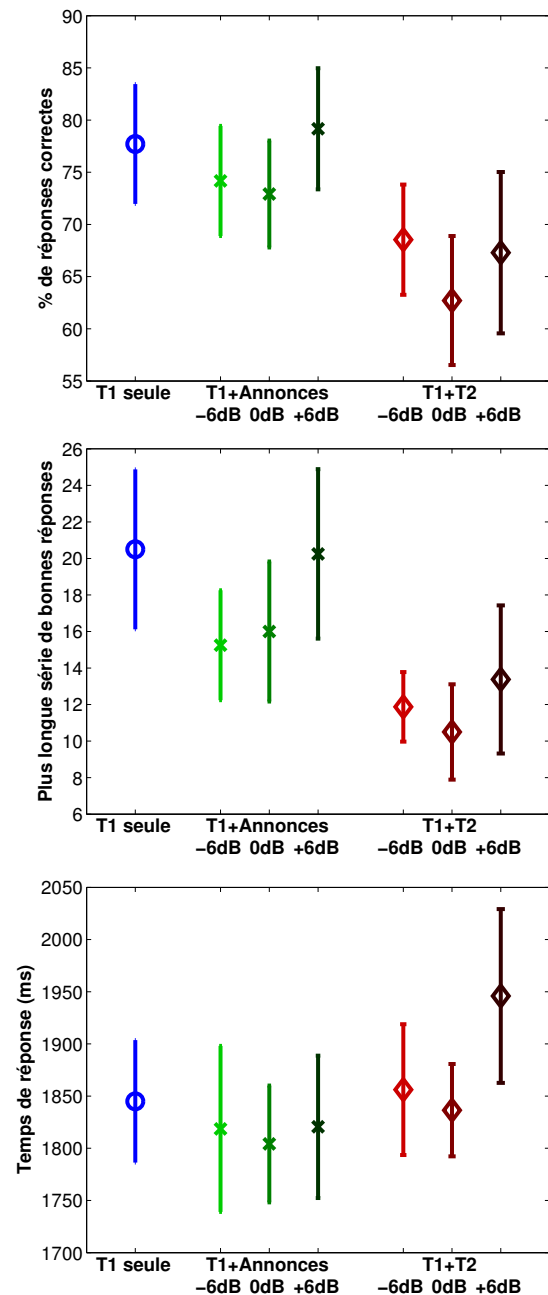


FIGURE 1 – Moyennes des performances obtenues par l'ensemble des participants sur la tâche 1 en fonction des conditions sur la tâche 2 (présence et rapport annonces-sur-chiffres). Les temps de réponse ne concernent que les essais réussis. Les barres d'erreur représentent l'erreur standard.

sonore ni d'interaction entre les deux facteurs n'a pu être relevé sur cette variable (resp. $F_{2,14} = 0.36, p = 0.70$ et $F_{4,28} = 0.14, p = 0.97$).

En cohérence avec l'hypothèse 1, un test post-hoc sur le facteur Tâche, avec corrections de Bonferroni, montre que les enchaînements de réponses correctes sont réduits lorsque les deux tâches sont effectués en parallèle par rapport à la condition contrôle sans annonces ($p = 0.015$). Cela confirme que les participants ont du mal à rester concentrer aussi longtemps que dans la condition contrôle sans annonces. Aucune différence significative ne peut être observée entre les autres conditions ($p's > 0.2$).

4.1.3 Temps de réponses

Les TR sont calculés entre l'instant où débute la lecture du signal correspondant au deuxième chiffre à ajouter et l'instant où le participant saisit au clavier le premier chiffre du résultat à saisir. Nous avons analysé les TR en ne considérant que les bonnes réponses. Les analyses ne montrent qu'un effet quasi-significatif de la présence des annonces sur le temps de réponse ($F_{2,14} = 2.74, p = 0.06$), les TR étant plus légèrement plus longs lorsque les deux tâches sont à réaliser en même temps. Le niveau sonore et l'interaction présence annonces \times niveau sonore n'ont pas d'influence sur les TR (resp. $F_{2,14} = 1.22, p = 0.29$ et $F_{4,28} = 0.89, p = 0.46$).

4.2 Analyse des résultats de la tâche 2 (discrimination de quai)

Nous considérons dans cette partie les performances obtenues pour la tâche 2 sur les 3 séries T1+T2 impliquant cette tâche.

Sur les 288 essais enregistrés (12 *essais* \times 3 *conditions* \times 8 *participants*), nous n'avons relevé que 4 erreurs : une mauvaise interprétation du quai (pour la condition de volume 0 dB) et trois non-détection de la cible (deux pour la condition 0 dB, une pour la condition +6 dB). Nous avons ensuite analysé les TR obtenus lorsqu'une réaction du participant était attendue (cas des annonces cibles). Comme la durée de prononciation des destinations et horaires pouvaient varier d'une annonce à l'autre, les mesures de TR correspondent en fait à la durée entre le début du morceau d'annonce qui donne le quai et l'instant où le participant donne sa réponse. Bien qu'en moyenne les TR soient plus courts pour les annonces à +6 dB que pour les annonces à 0 dB ou -6 dB (Table 2), ce qui tend à confirmer notre hypothèse 2 selon laquelle la tâche de discrimination est plus facile pour des annonces plus saillantes, l'analyse de variance ne révèle aucune influence significative du niveau sonore des annonces par rapport aux chiffres sur le temps de réponse ($F_{2,137} = 1.45, p = 0.23$).

TABLEAU 2 – Moyenne en millisecondes des temps de réponse obtenue sur la tâche 2 pour l'ensemble des participants (\pm erreur standard)

	Rapport de niveau sonore annonces/chiffres		
	-6 dB	0 dB	+6 dB
moyenne	1837 (± 99)	1903 (± 147)	1643 (± 86)

4.3 Bilan de l'expérience pilote

Conformément à la littérature, les résultats de cette expérience ont confirmé notre Hypothèse 1 selon laquelle les ressources cognitives disponibles pour l'écoute sont limitées. Ainsi la réalisation des deux tâches en parallèle a conduit les participants à faire plus d'erreurs, plus souvent et en répondant plus lentement sur la tâche d'addition. Cela ne peut être dû à un simple effet de masquage sonore des annonces sur les chiffres car aucune différence n'est observable entre la condition contrôle T1 seule et les conditions T1 bruitées par les annonces.

En revanche, aucune influence significative du niveau sonore des annonces n'a pu être relevée, ni pour les variables dépendantes de la tâche 1 ni pour celles de la tâche 2. Ainsi nos hypothèses 2 et 3, selon lesquelles une annonce plus forte en volume serait, d'une part, plus facile à traiter qu'une annonce moins saillante et, d'autre part, qu'elle attirerait plus l'attention, ne sont pas validées ici.

5 Discussion et perspectives

Dans sa version actuelle, le protocole expérimental que nous avons proposé ne suffit pas pour évaluer la saillance auditive d'annonces vocales. Bien que le principal facteur que nous ayons fait varier dans cette étude soit le niveau sonore, un paramètre acoustique ayant déjà été validé dans la littérature comme facteur d'influence, nous n'avons pu le redémontrer par ce protocole. Plusieurs raisons peuvent expliquer cela.

Tout d'abord, le manque d'impact du niveau sonore peut être dû à une trop faible variation de niveau pour les annonces (de -6 dB à +6 dB) là où d'autres expériences comme [4] se permettaient d'aller jusqu'à une différence de -20 dB.

De plus, pour augmenter le nombre d'observations sur la tâche 2, nous avons arbitrairement décidé que 50 % des annonces seraient des annonces cibles. Diminuer la fréquence d'apparition des annonces cibles, pour qu'en majorité les annonces n'aient pas à être traitées, permettrait d'éviter que les participants ne se concentrent volontairement sur cette tâche. L'effet d'attraction involontaire vers une annonce saillante serait donc renforcé.

Dans cette étude nous avons explicitement demandé aux participants de ne pas donner de priorité à l'une ou l'autre des deux tâches, mais implicitement la tâche continue d'addition (PASAT) était considérée comme la tâche principale car, contrairement à l'autre tâche, elle apparaissait dans toutes les séries. Cela nous permettait d'évaluer le pouvoir d'attraction, c'est-à-dire le pouvoir à désengager l'attention d'une tâche initiale. Pour mettre en évidence une facilitation du traitement des informations contenues dans les annonces (meilleure intelligibilité, diminution de l'effort d'écoute), nous pouvons cependant modifier la consigne pour rendre la tâche 2 prioritaire. Les performances sur la tâche 1 ne dépendraient alors que des ressources attentionnelles libérées lorsque la tâche de discrimination est plus facile.

La tâche d'addition (PASAT), que nous avons choisie parce qu'elle requiert une attention soutenue et de nombreuses ressources cognitives, présente aussi un énorme avantage dans le cadre du projet I'City For All. En effet, il est facile d'adapter cette tâche pour des personnes âgées : d'une part, il est possible d'afficher les chiffres au centre de l'écran, un par un, de manière synchrone à la diffusion de sons, pour rendre la tâche plus facile comme cela a déjà été proposé [16] ; d'autre part, il est possible de faire varier l'intervalle de temps qui sépare deux chiffres (ISI pour *interstimulus interval*) de façon à laisser plus de temps aux participants pour réaliser les additions. L'ISI est d'ailleurs la variable principale dans la version originale du PASAT [15]. Dans notre expérience, l'ISI choisi était de 3 s ce qui correspond à une vitesse plutôt lente du test du PASAT dont la version rapide correspond à un ISI de 2 s ou moins. Nous avons choisi cette valeur car, contrairement au

test original, nous allons combiner cette tâche à une tâche parallèle. Il est cependant possible que forcer le participant à augmenter sa concentration sur la tâche primaire, en diminuant l'ISI, permette de mieux mettre en évidence le pouvoir d'attraction/distraction d'un son extérieur tel qu'une annonce.

Une des autres raisons envisagées pour laquelle aucune influence du volume sonore n'a été observée est la durée des annonces. Dans les précédents travaux sur l'étude de la saillance sonore, les stimuli utilisés étaient toujours de courte durée. Même si le début des annonces fortes a pu attirer plus l'attention des participants que le début des annonces de niveau faible, les participants devaient attendre la fin de l'annonce pour obtenir l'information de quai, temps pendant lequel leur attention devait retourner à la tâche d'addition. Dans un premier temps, nous allons donc réaliser des analyses complémentaires pour évaluer si il y a une correspondance entre les débuts d'annonces et les essais de la tâche d'addition pour lesquels les calculs mentaux ont pris plus de temps ou ont conduit à des erreurs.

Finalement, faute de capteurs permettant de mesurer directement notre attention auditive, les propositions que nous avons faites jusqu'à présent ne concernent que des mesures comportementales. Or les études de Coutrot et al. [19] montrent qu'il y'a une corrélation directe entre la saillance d'un évènement sonore et les saccades oculaires de l'auditeur/observateur. Une mesure physiologique, par suivi oculaire, pourrait ainsi permettre d'identifier si un son est saillant ou non.

Remerciements

Ce projet a été financé par le programme européen Ambient Assisted Living. Nous tenons à remercier tous les volontaires qui ont participé à cette expérience, ainsi que le Centre d'Études de la Sensorimotricité (Cesem) et le Centre de Neurophysique, Physiologie, Pathologies (CNPP) pour le prêt de leur chambre sourde.

Références

- [1] K. Pichora-Fuller et G. Singh, Effects of age on auditory and cognitive processing : implications for hearing aid fitting and audiologic rehabilitation, *Trends in Amplification*, **10**(1), 29-59 (2006)
- [2] S. H. Chon, Timbre Saliency, The Attention-Capturing Quality of Timbre, Thèse de doctorat, McGill University (2013)
- [3] R. Cusack et R. P. Carlyon, Perceptual asymmetries in audition, *Journal of experimental psychology : Human Performances* **29**(3), 713-725 (2003)
- [4] C. Kayser, C. I. Petkov, M. Lippert et N. K. Logothetis, Mechanisms for allocating auditory attention : An Auditory Saliency Map, *Current Biology*, **15**, 1943-1947 (2005)
- [5] O. Kalinli et S. Narayanan, *A saliency-based auditory attention model with applications to unsupervised prominent syllable detection in speech*, Proceedings of InterSpeech, 1941-1944, Antwerp, Belgique (2007)
- [6] B. De Coensel et D. Botteldooren, *A model of saliency-based auditory attention to environmental sound*, 20th International Congress on Acoustics, 1-8, Sydney, Australie (2010).
- [7] V. Duangudom, *Computational auditory saliency*, Thèse de doctorat, Georgia Institute of Technology (2012)
- [8] T. Tsuchida et G. W. Cottrell, *Auditory Saliency Using Natural Statistics*, Annual meeting of the cognitive science society, 1048-1053, Sapporo, Japon (2012)
- [9] F. Tordini, A. S. Bregman, A. Ankolekar, T. E. Sandholm, J. R. Cooperstock, *Towards an improved model of auditory saliency*, 19ème International Conference on Auditory Display, 189-196, Lodz, Pologne (2013)
- [10] Discussion Comparing Auditory Saliency Models, Telluride Neuromorphic Cognition Engineering Workshop, (2011) disponible à : <http://neuromorphs.net/nm/wiki/2011/att11/AuditorySaliency>
- [11] P. A. Gosselin, et J.-P. Gagné, Use of a Dual-Task Paradigm to Measure Listening Effort, *Revue canadienne d'orthophonie et d'audiologie*, **34**(1) 43-51 (2010)
- [12] V. Duangudom et D. V. Anderson, *Using auditory saliency to understand complex auditory scenes*, European Signal Processing Conference, 1206-1210, Poznan, Pologne (2007)
- [13] H. Pashler, Dual-task interference in simple tasks : Data and theory, *Psychological Bulletin*, **116**(2), 220-244 (1994)
- [14] M. Crossley et M. Hiscock, Age-related differences in concurrent-task performance of normal adults : Evidence for a decline in processing resources, *Psychology and Aging* **7**, 499-506 (1992)
- [15] D. M. Gronwall, Paced Auditory Serial-Addition Task : a measure of recovery from concussion, *Perceptual and motor skills* **44**, 367-373 (1977)
- [16] T. N. Tombaugh, A comprehensive review of the Paced Auditory Serial Addition Test (PASAT). *Archives of clinical neuropsychology* **21**(1), 53-76 (2006)
- [17] F. Reuter, K. Baumstarck-Barrau, A. Loundou, J. Pelletier, et P. Auquier, Paced auditory serial addition test : normative data in a French population. *Revue Neurologique* **166**(11), 944-947 (2010)
- [18] T. Bouchara, C. Jacquemin et B. F.G. Katz, Cueing multimedia search with audio-visual blur, *ACM Transactions on Applied Perception* **10**(2), 1-13 (2013)
- [19] A. Coutrot, N. Guyader, G. Ionescu et A. Caplier, Video viewing: do auditory salient events capture visual attention ?, *Annals of Telecommunications* **69** (1-2), 89-97 (2014)